



Special issue SOCO 2017: AI and ML applied to Health Sciences (MLHS)

Francisco Javier de Cos¹ · Michał Woźniak² · Juan Albino Méndez³ · José R. Villar Flecha¹

Published online: 21 January 2019

© Springer-Verlag London Ltd., part of Springer Nature 2019

The seven papers included in this special issue represent a selection of extended contributions presented at the 12th International Conference on Soft Computing Models in Industrial and Environmental Applications, SOCO 2017 held in Leon, Spain, September 6th–8th, 2017, and organized by the Department of Mechanical, Computer and Aerospace Engineering of the University of Leon.

This special issue is aimed at practitioners, researchers and postgraduate students who are engaged in developing and applying the latest theoretical advances and real-world applications in Artificial Intelligence and Machine Learning applied to Health Sciences.

The contribution by Artime et al. focuses on a classification problem to predict computer vision syndrome occurrence in health personnel. An algorithm combining support vector machines and genetic algorithms is developed to identify those characteristics of the subjects that have the strongest influence on the syndrome. The iterations of the genetic algorithm are focused on maximizing the value of the area under the ROC curve of each of the SVM models calculated. The Youden's index is also used to corroborate the high performance of the model developed.

The work of Casteleiro-Roca et al. introduced a new model to predict the AntiNociception Index (ANI) signal in the near future. After testing that there is possibility to model the signal, the new approach predicts the variation

of the signal with more than 30 s in advance. Artificial Neural Networks and Support Vector Regression have been used in the hybrid model to increase the performance of the model. This hybrid model achieves very good results in the validation phase with real cases (that are not used for training).

The contribution of Sánchez Lasheras, J. E., et al. presents two different approaches for detecting prostate cancer regarding genetic influences by means of the study of single nucleotide polymorphisms. The first technique used, Multivariate Adaptive Regression Splines, allows them not only to model the problem and develop a classification algorithm, also to specify the most relevant genes, regarding the results obtained from these models. The other approach is Multi-Layer Perceptron trained as classifiers for determining the predisposition of the patient to suffer prostate cancer. The conclusion that the authors present is that the Multi-Layer Perceptron trained as classifier results to be the model that reaches the best performance in terms of AUC.

The work presented by Fernandez-Lequerica et al. applies Multivariate Adaptive Regression Splines (MARS) and Self-Organizing Maps (SOM) to determine the prognosis of cancer of salivary glands using clinical and histological variables as well as molecular markers. The results obtained outperform previous researches in terms of specificity and sensitivity. Please note that as far as we know, this research is the first one that introduces immunohistochemically studied molecular prognostic factors together with relevant clinical variables in prognostic indices of parotid cancers.

The paper submitted by Tomasz Andrysiak presents the use of sparse representation of ECG signal and overcomplete dictionary learning method for detecting anomalies in the analyzed electrocardiographic record. The two sorts of dictionaries were used herein: The first one uses analytic Gabor functions, while the second sort of dictionaries employs the modified Method of Optimal Directions. The dictionaries became a basis for the analyzed sparse representation of electrocardiographic record. During the anomaly detection process, which was based on

✉ Francisco Javier de Cos
fjcos@uniovi.es

Michał Woźniak
michal.wozniak@pwr.edu.pl

Juan Albino Méndez
jamendez@ull.edu.es

José R. Villar Flecha
villarjose@uniovi.es

¹ University of Oviedo, Oviedo, Spain

² Wrocław University of Science and Technology, Wrocław, Poland

³ University of La Laguna, La Laguna, Spain

decomposition of the analyzed signal into correct values and outliers, a modified Alternating Minimization Algorithm was used. Performance of the proposed solution was tested by means of a widely available database of ECG signals, i.e., the MIT-BIH Arrhythmia Database, and the obtained experimental results confirmed its effectiveness for anomaly detection (simple distortions and four arrhythmia types) in the analyzed ECG signals.

Alvarez-Arenal et al. introduced in their publication a novel artificial neural network model able to classify individuals suffering from bruxism in clenching and grinding patients according to the value of certain occlusal variables and other parameters. Patients suspected of bruxism represent a very heterogeneous group. Some require immediate treatment, while others, with only minor disorders, may not need treatment at all. The models were analyzed using Receiver Operating Characteristics (ROC) curve analysis, calibration assessments, inter- and intra-method variations. The artificial neural network model obtained can distinguish between clenching and grinding patients requiring only the analysis of a few variables and with a high rate of success.

In the last contribution, De La Cal et al. propose a new method to balance multivariate Time Series (TS) datasets.

Multiple sources TS datasets are typical in Internet of Things and e-Health problems where TS are analyzed to classify the current status or to detect certain events, among others. Unfortunately, the events to detect are rare or highly infrequent; therefore, the TS datasets need balancing. Besides, the design of methods in the literature to balance TS datasets has been scarcely studied, representing a new challenge in the next few years.

This study copes with this problem, presenting an extension of the well-known SMOTE (Synthetic Minority Oversampling TEchnique) algorithm for multivariate TS datasets. While the existing solutions are only suitable for a univariate and single-source TS dataset, generating new synthetic datasets or ghost points in the distance space, this new proposal is able to balance multi-sources TS dataset while preserving the shape of the new synthetic multivariate TS using the DTW distance function. Finally, a complete evaluation of the algorithm is performed dealing with real-world Time Series datasets coming from the e-Health domain.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.