International Journal of Bioelectromagnetism
Vol. 14, No. 4, pp. 198 - 203, 2012

www.ijbem.org

A Comparison of Performances of Different Feature Selection Methods applied to Biomedical Data

Samanta Rosati[a], Gabriella Balestra[a]

[a]*Departement of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy*

Correspondence: S Rosati, Department of Electronics and Telecommunications, Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy. E-mail: samanta.rosati@polito.it, phone +39 011 090 4207

Abstract. Migraine is a debilitating disease whose causes are not yet completely explained. Near-InfraRed Spectroscopy (NIRS) is a non-invasive technology commonly used for the assessment of the cerebral autoregulation during active stimuli.

Feature Selection (FS) allows dimensionality reduction of multivariate datasets, highlighting the most informative variables and deleting redundant and irrelevant information. Rough Set Theory (RST) is one of the most used tool for FS, enables to manage incomplete and imperfect knowledge without any assumption about data model.

This study involved a total of 80 subjects, divided in 3 groups: 15 healthy subjects taken as controls, 14 women suffered from migraine without aura and 51 women from migraine with aura. We apply three different methods of FS based on RST to a set of 26 parameters extracted from NIRS signals recorded in the subjects during breath-holding (BH) and hyperventilation (HYP). We compare the extracted subsets of features in the subjects' classification by means of Artificial Neural Networks. The results show good performance for all subsets, with a percentage of correct classification above the 90%.

Keywords: Artificial Neural Networks; Feature Selection; Migraine; Rough Set Theory; Time-Frequency Distributions

1. Introduction

Migraine is a neurological disorder associated with many factors, even if in the last years it is more often considered as a pathology linked to neurovascular impairments [Tietjen, 2009]. Particularly, the association between migraine and impaired cerebral autoregulation or vasomotor tone has widely been investigated [Liboni et al., 2007; Nowak and Kacinski, 2009; Vernieri et al., 2008]. The assessment of the cerebral autoregulation is usually performed during active stimuli, like breath-holding (BH) or hyperventilation (HYP), and basing on signals derived from transcranial Doppler sonography [Molinari et al., 2006] or Near-InfraRed Spectroscopy (NIRS) [Liboni et al., 2007]. Specifically, NIRS is a real-time and non-invasive tool for the monitoring of the concentration of oxygenated (O2Hb) and reduced hemoglobin (HHb) in brain cortex.

Feature Selection (FS) is a procedure allowing dimensional reduction of multivariate data, deleting the redundant attributes, in order to extract from a high-dimensional dataset the features with most significant information. Moreover, a too large number of features does not necessarily allow increasing the classification accuracy: several attributes may be irrelevant or, even worse, may introduce some kind of noise which decreases the classifier performances [Jensen and Shen, 2008].

An exhaustive search of the best feature subset could be performed exploring the whole space of possible subsets (brute-force approach). However, such a procedure results inapplicable when the number of initial variables is relatively medium-high, that is for most real applications. These considerations led to the development, during the past years, of several methods for FS based on an heuristic search [Saeys et al., 2007; Somol et al., 2007; He and Yu, 2010]. Heuristics identify a wide class of algorithms used in order to solve optimization problems, bypassing the complexity induced by real world applications. Time and computational costs reduction is achieved by addressing the solution search toward a high-quality space of admissible solutions. They are based on the evolution of an heuristic function that provides an estimate of the current solution goodness. Their adaptability to different fields with relatively few modifications makes the heuristic algorithm a suitable tool also for FS.

The main idea of the study presented here is to compare the performances of three different FS algorithms based on heuristic search in selecting the most important features among those extracted from NIRS signals recorded in migraine sufferers during BH and HYP, so that we can have a full description of the pathology.

2. Methods

The dataset used here was constructed processing the NIRS signals recorded during a study about the cerebral hemodynamics in subjects affected by migraine with (MwA) and without (MwoA) aura. Aura is a specific disturbance associated with migraine that can cause visual, speech, or perceptional impairments.

It is the result of a study that was conducted at Gradenigo Hospital of Turin (Italy) and involved a total of 80 subjects, divided in 3 groups, based on pathology: 15 healthy subjects were healthy controls (age: 29.2 ± 8.5), 14 women suffered from MwoA (age: 44.4 ± 9.7) and 51 women from MwA (age: 38.0 ± 12.1). All the subjects were instructed about the purposes of the study and signed an informed consent prior to undergoing tests. Migraine with and without aura was diagnosed according to the criteria of the International Headache Society [Headache Classification Committee of the International Headache Society, 1988]. Migraine subjects were tested in their interictal periods (i.e. when they were free of pain).

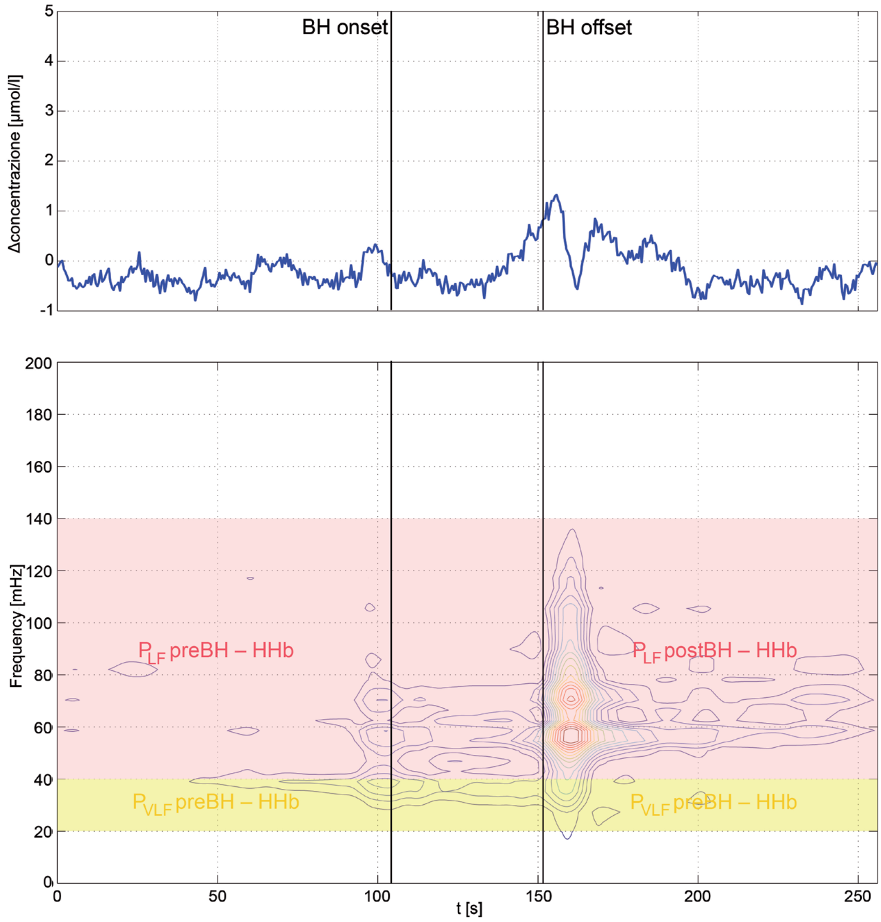

Figure 1. *HHb concentration signal (upper panel) recorded on a healthy subject lasting 256 seconds with the BH in the middle of the analysis window. The onset and the offset of the event are marked by vertical lines. In the lower panel is showed the Choi-Williams distribution of the signal (σ=0.05). The yellow zone represents the VLF band (20-40 mHz) while the pink one indicates the LF band (40-140 mHz).*

2.1. Feature Extraction

Because of the non-stationarity of the NIRS signals acquired during vaso-active maneuvers, a time-frequency processing approach is required in order to extract information from them. Specifically, the Choi-Williams transform was used in this study to analyze the two NIRS signals. Moreover the time-frequency Squared Coherence Function (SCF), between the concentration signals of O2Hb and HHb, was computed on the basis of the Choi-Williams representations.

We considered signals lasting 256 seconds, with the active maneuvers (BH and HYP) in the middle of the analysis window. Hence, spectral resolution was better than 4 mHz.

The time–frequency distributions, concerning both Choi-Williams transforms of the two signals and the SCF, were analyzed in two specific bands, Very Low Frequencies (VLF, between 20 and 40 mHz) and Low Frequencies (LF, between 40 and 140 mHz), before and after BH and HYP. The percentage of signal power in the two bands (referred to the total power of the signal) was calculated before and after each event. A Choi-Williams representation of the HHb NIRS signal during BH for a healthy subject is reported in Fig. 1.

Moreover, two additional variables, derived from the analysis in the time domain, were measured: the BHI indexes for HHb and O2Hb signals calculated as the percent variation of the concentration as effect of BH, normalized with respect to the BH duration. Such a procedure allowed to extract 26 variables listed in the first column of Table 1.

2.2. Feature Selection

Three FS methods were applied to our dataset, all involving the Rough-Set Theory (RST) concepts. RST was introduced by Pawlak [Pawlak, 1982] in order to manage imperfect and incomplete knowledge, without any assumption about data model. The basic principle of RST says that if two objects are indiscernible with respect to a certain variable, then they should be classified in the same class.

As heuristic methods require an appropriate function to measure the relevance of the chosen solution, we decide to evaluate the feature subset by means of the dependency degree $\gamma_C(D)$ measured between a decision attribute D and the subset of conditional features C. If all values from D are uniquely determined by values of attributes C, $\gamma_C(D)$ results equal to 1 and the dataset is defined as consistent. Real dataset are usually not consistent so the maximum value is less than 1, and in our case the maximum value is 0.975.

The Genetic Rough-Set Attribute Reduction (GenRSAR) [Jensen and Shen, 2003] is an heuristic algorithm employing a genetic search strategy in order to determine the optimal reduced subset. It uses a standard genetic algorithm structure in which the fitness function considers both the size of subset R and its suitability in terms of dependency degree.

The Ant Rough-Set Attribute Reduction (AntRSAR) [Jensen and Shen, 2003] is another methodology based on heuristic search that employs the same strategy used by ants in order to find the best path in the direction of food. In this case the ant colony optimization is obtained using a number of ants equal to the number of features and setting the dependency degree as stopping criterion.

The last method tested was the QuickReduct Algorithm (QRA) [Jensen and Shen, 2003]. It is a standard Rough Set algorithm allowing to resolve reduct search problems without generating all the possible subsets. The main idea is to add to the current reduct subset those attributes producing a larger increase in the dependency degree $\gamma_C(D)$.

MATLAB environment was used in order to implement all procedures for FS. As for the GenRSAR and AntRSAR algorithm, the initial parameters were set as suggested in [Jensen and Shen, 2003]. To overcome the stochastic nature of the two methods we performed twenty runs lasting one hundreds epochs for each of them. The best solutions was defined as the one with the higher dependency degree.

Artificial neural networks (ANNs) were used in order to compare the classification performances of the different subsets. The basic idea was that a good procedure of FS allows removing redundant features so that the reduct provides the same quality of classification of the original set [Chen et al., 2011] or even improve it.

In this study we built three networks, one for each subset, with similar structures. Particularly, the number of input neurons was equal to the number of selected features and in the output layer there was only one neuron. Moreover, we chose to use one hidden layer with a number of neurons approximately equal to 1/2 of the input neurons. As for the neuron activation functions, we used a logarithmic sigmoid function for the hidden layers and a linear function for the output layer. Back-propagation was chosen as the learning algorithm and the mean squared error was used as performance function. The initial values of interconnection weights were set randomly.

3. Results and Discussion

The FS results are reported in Table 1 in terms of selected features, number of features and dependency degree. As for GenRSAR and AntRSAR only the best result of twenty runs is described. The three methods select from 7 to 10 features. All reducts allow obtaining a very high dependency degree, even if GenRSAR has a lower performance while for QRA and AntRSAR the dependency is equal to the maximum obtainable value.

Fig. 2, 3 and 4 present the results of applying the ANNs based on the different features subsets in terms of percentage of correct classification considering the whole population of patients and each subjects' class individually.

Table 1. *Results of the three FS procedures. First column contains the 26variables used as input for the features selection strategies. From the second to the last column are reported the results of QRA, GenRSAR and AntRSAR (1: feature selected). The selected parameters are highlighted in light grey. The last rows contain the number of features selected in each subset and the subset dependency degree.*

Features	*QRA*	*GenRSAR*	*AntRSAR*
PVLF preBH O2Hb	0	1	1
PVLF postBH O2Hb	0	0	0
PLF preBH O2Hb	0	0	0
PLF postBH O2Hb	0	0	0
PVLF preBH HHb	0	0	0
PVLF postBH HHb	1	0	0
PLF preBH HHb	0	1	0
PLF postBH HHb	0	0	0
PVLF preHYP O2Hb	1	1	1
PVLF postHYP O2Hb	0	0	0
PLF preHYP O2Hb	0	0	1
PLF postHYP O2Hb	0	1	1
PVLF preHYP HHb	0	0	0
PVLF postHYP HHb	0	0	0
PLF preHYP HHb	1	0	0
PLF postHYP HHb	0	0	1
SVLF preBH	0	0	0
SVLF postBH	1	0	0
SLF preBH	0	0	0
SLF postBH	0	0	0
SVLF preHYP	1	0	0
SVLF postHYP	1	1	1
SLF preHYP	1	1	0
SLF postHYP	1	1	1
BHIO2	1	1	1
BHICO2	0	0	0
# of features	9	8	8
dependency degree	0.975	0.950	0.975

In Fig. 2 the classification results related to GenRSAR for the twenty runs are reported. In all the runs the overall classification is always below 100%.

In Fig. 3 the classification results related to all the runs of AntRSAR are presented. In this case there are four runs with an overall classification equal to 100%.

Fig. 4 presents the summary of the results obtained with the three methods. For GenRSAR and AntRSAR only the best solutions are taken into consideration. The overall classification accuracy is 100% for QRA and AntRSAR while it is close to 95% for GenRSAR. This means that the variables selected by QRA and AntRSAR are more informative in order to characterize the three classes of subjects than the features selected by GenRSAR.

Even if the results of the AntRSAR are comparable to those of QRA, we decided that QRA is the preferable method because it is deterministic, that is it returns always the same subset of features.

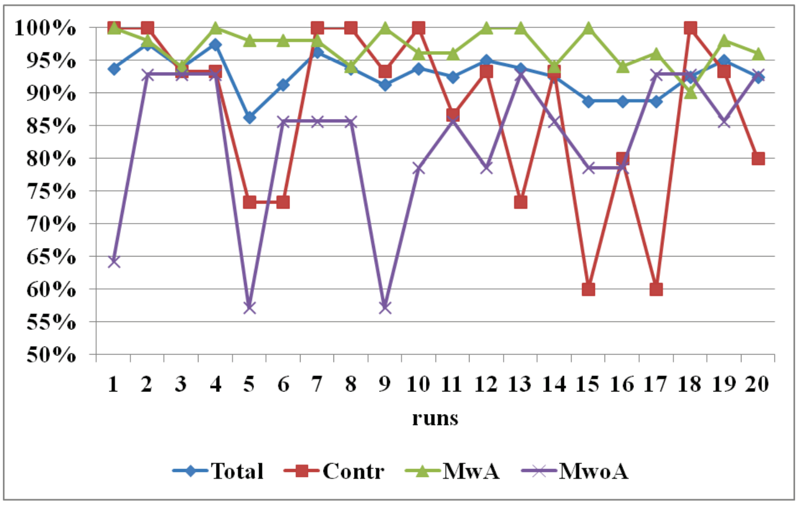

Figure 2. *Percentage of correct classification for the whole population and considering each subjects' class individually obtained with the feature subsets selected in all the runs of GenRSAR.*

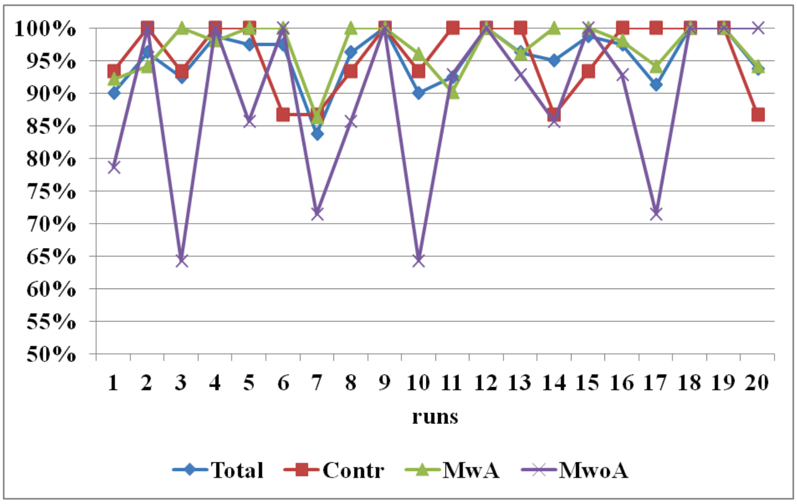

Figure 3. *Percentage of correct classification for the whole population and considering each subjects' class individually obtained with the feature subsets selected in all the runs of GenRSAR.*

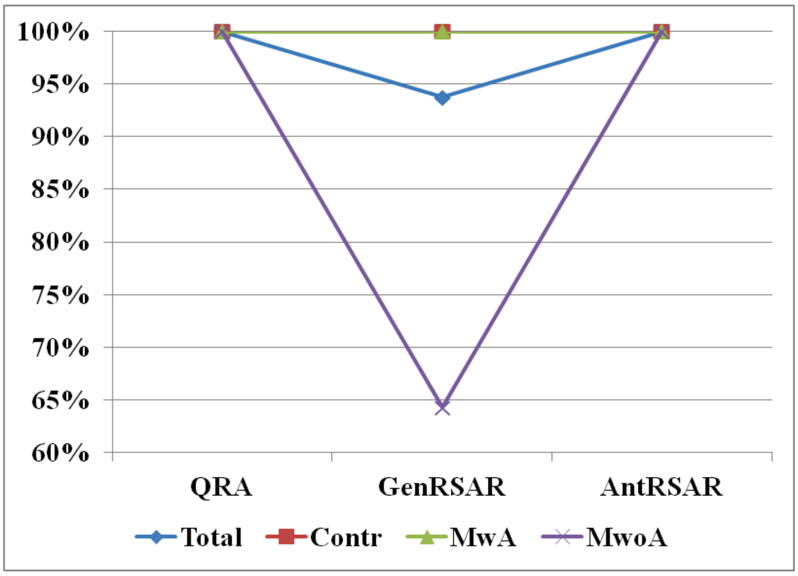

Figure 4. *Summary of the results of applying the ANNs to the feature subsets obtained with the three methods, in terms of percentage of correct classification for the whole population and considering each subjects' class individually.*

References

Chen YM, Miao DQ, Wang RZ, Wu KS. A rough set approach to feature selection based on power set tree. Knowledge-Based Systems 24(2):275-281, 2011.

He Z, Yu W. Stable feature selection for biomarker discovery. *Computational biology and chemistry*, 34(4):215-225, 2010.

Headache Classification Committee of the International Headache Society. Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. *Cephalalgia*, 8(7):1-96, 1988.

Jensen R, Shen Q. Finding rough set reducts with ant colony optimization. In proceedings of the 2003 UK Workshop on Computational Intelligence, 2003, 15-22.

Jensen R, Shen Q. Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches. Wiley-IEEE Press, Hoboken, 2008.

Liboni W, Molinari F, Allais G, Mana O, Negri E, Grippi G, Benedetto C, D'Andrea G, Bussone G. Why do we need NIRS in migraine?. *Neurological Sciences*, 28(2):S222-S224, 2007.

Molinari F, Liboni W, Grippi G, Negri E. Relationship between oxygen supply and cerebral blood flow assessed by transcranial Doppler and near-infrared spectroscopy in healthy subjects during breath-holding. *Journal of Neuroengineering and Rehabilitation*, 3:16, 2006.

Nowak A, Kacinski M. Transcranial doppler evaluation in migraineurs. *Neurologia i Neurochirurgia Polska*, 43(2):162-172, 2009.

Pawlak Z. Rough sets. *International Journal of Parallel Programming*, 11(5):341-356, 1982.

Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507-2517, 2007.

Somol P, Novovi J, Pudil P. Notes on the evolution of feature selection methodology. *Kybernetika*, 43(5):713-730, 2007.

Tietjen GE. Migraine as a systemic vasculopathy. *Cephalalgia*, 29(9):987-996, 2009.

Vernieri F, Tibuzzi F, Pasqualetti P, Altamura C, Palazzo P, Rossini PM, SilvestriniM. Increased cerebral vasomotor reactivity in migraine with aura: an autoregulation disorder? A transcranial Doppler and near-infrared spectroscopy study. *Cephalalgia*, 28(7):689-695, 2008.