Feature Selection Methods to Extract Knowledge and Enhance Analysis of Ventricular Fibrillation Signals

Juan Caravaca
University of Valencia, Spain

Antonio J. Serrano-López
University of Valencia, Spain

Emilio Soria-Olivas
University of Valencia, Spain

José M. Martínez-Martínez
University of Valencia, Spain

Pablo Escandell-Montero
University of Valencia, Spain

Juan F. Guerrero-Martínez
University of Valencia, Spain

INTRODUCTION

Nowadays there is a problem involving sudden cardiac arrest, it is the single most common cause of death in developed countries (The American Heart Association, 2007). Ventricular fibrillation (VF) is a malignant arrhythmia that consists on a rapid and disorganized pacing of the heart that inhibits the cardiac output and can result in cardiac arrest and death in, as short as, four minutes (Kari, Sears, & DeAntonio, 2012).

Physical exercise has both beneficial and adverse effects on sudden cardiac death risk. Most studies have found inverse associations between increasing regular physical activity and sudden cardiac arrest, and these results were most consistent with moderate levels of exercise (Rajay & Albert, 2012). The main mechanism by which physical exercise regulates the cardiac activity involves the sympathetic-vagal balance of autonomic nervous system, called extrinsic mechanisms. These modifications produce a decrease of cardiac frequency that could have protective effects against cardiac arrhythmias and cardiac sudden arrest.

There are other kinds of mechanisms, which do not involve the autonomic nervous system, for the cardiac response regulation by physical exercise. These mechanisms are related to changes in the electrophysiological characteristics of the heart, called intrinsic mechanisms, as can be an increment of action potential duration in ventricular cardiocytes (Nataly el al., 2002). Other previous studies have found that physical exercise modifies VF response by intrinsic mechanisms. These modifications were found in terms of several parameters extracted from time and frequency domains (Guerrero et al., 2009b), and from their spatial homogeneity (Guerrero et al 2009a). Nevertheless, the obtained results suggest that not all the features can describe these intrinsic modifications in the same way, being some features more relevant than the other ones.

This article encompasses the identification of the most relevant features to describe the intrinsic effects that physical exercise produces in cardiac response during VF. Several kinds of features will be analysed, derived from time and frequency domains. These features will be ranked according to their relevance

DOI: 10.4018/978-1-4666-5888-2.ch548

to distinguish between subjects that were trained with physical exercise and other untrained subjects. Different feature selection techniques will be employed, and the obtained feature rankings will be aggregated to perform knowledge extraction.

BACKGROUND

Fibrillatory rhythms, as atrial and ventricular fibrillation, have traditionally been analysed with cardiac mapping systems (Nash et al., 2006). These systems capture signals in hundreds of points of the cardiac tissue by means of an electrode array placed either in the epicardium or in the endocardium. The electrode array simultaneously captures a high number of electrograms, which can be used to extract some parameters to describe the fibrillatory process. Such description is usually performed through parameters that measure the regularity, despite fibrillation has been traditionally considered a chaotic and irregular process (Jalife, 2000). The regularity can be measured in time and frequency domains. Besides, there are other parameters related to the fibrillatory frequency content that can provide information related to the activation rate of the tissue.

The Dominant Frequency (DF) of a fibrillation process is defined as the highest peak on its power spectrum, and it is both related to the mean activation period and to the cardiocytes refractoriness period (Sanders et al., 2005). DF is one of the most commonly used parameters in fibrillatory rhythms due to its direct relation with these physiological characteristics.

Other spectral parameter is commonly used in order to measure the spectral regularity of a fibrillatory electrogram: the Normalized Energy (NE). It is defined as the ratio between the spectral energy in a frequency window (DF$\pm$1Hz) and the spectral energy in the interest frequency band (Sanders et al., 2005). NE is related to spectral complexity; a higher value of NE implies that spectral activity is focused on DF, i.e. the activity is lesser complex.

It is also common to use some parameters to describe the temporal regularity level of the fibrillatory electrograms. The Regularity Index (RI) compares the morphological regularity of all activation waves found in a temporal window, analysing the similarity level among the activation waves (Faes, Nollo, Antolini, Gaita, & Ravelli, 2002).

In this sense, the parameter Number of Occurrences (NO) also allows analysing the complexity of a fibrillatory electrogram. It is defined as the percentage of samples which amplitude falls in a zero centred window, respect to the total number of samples (Calcagnini, Censi, Michelucci, & Bartolini, 2006).

As these parameters are usually computed for all channels in an electrode array, a representation (called parameter map) of the values of the parameters in each electrode can be carried out. It is usual to use the mean value or standard deviation of the parameter in a map to provide information about the global parameter value in each temporal window. The main drawback of this technique is that the information about the spatial uniformity of the parameters is lost (Guerrero et al., 2009a).

The Regions of Interest (ROI) analysis of a parameter map allows studying the spatial uniformity of a parameter with a representation of the electrodes that pass a threshold value. With that technique, the spatial uniformity of a parameter can be analysed in several ways. The ROI spatial area (ROIsa) is the percentage of a map covered by electrodes that passed the threshold, and it is related to the spatial homogeneity of the parameter. The ROI spatial number (ROIsn) is the number of electrode groups that passed the threshold, and it expresses the spatial fragmentation of the parameter. Finally, the ROI electrode number (ROIen) is the number of electrodes whose threshold criterion changed between two consecutive maps (Guerrero et al., 2009a). These three features derived from ROI can be computed for each parameter map (as can be DF, NE or RI).

Feature selection (FS) is a preprocessing step in data mining, used to extract knowledge from a dataset and to reduce its dimensionality selecting only the best features. There are multiple types of feature selection techniques, based on subset selection (multivariate, searching for the best feature subset) or based on rankings (univariate, studying each variable individually) (Wald, Khoshgoftaar, Dittman, Awada, & Napolitano, 2012). A lower robustness of a FS method leads to not robust results in feature selection. Such robustness can be increased by means of ensemble feature selection methods, which consists on performing the FS multiple times and aggregating the results (Abeel, Helleputte, Van de Peer, Dupont, & Saeys, 2010). There are three kinds of ensemble FS: based on data diversity (appli-