Application of machine learning techniques to analyse the effects of physical exercise in ventricular fibrillation

Juan Caravaca[a] , Emilio Soria-Olivas[b], Manuel Bataller[a], Antonio J. Serrano[b], Luis Such-Miquel[c], Joan Vila-Francés[b], and Juan F. Guerrero[a]

[a]*Digital Signal Processing Group, ETSE, Universitat de València. Avda Universitat S/N, 46100 Burjassot València, Spain*

[b]*Intelligent Data Analysis Laboratory, E.T.S.E., Universitat de València. Avda Universitat S/N, 46100 Burjassot València, Spain*

[c]*Department of Physiotherapy, Universitat de València and INCLIVA .València, Spain*

Corresponding Author:

Name: Juan Caravaca

Address: Avenida de la Universidad S/N, 46100, Burjassot, Valencia (Spain)

Telephone/Fax: (34) 963544143

email: juan.caravaca@uv.es

Abstract:

This work presents the application of machine learning techniques to analyze the influence of physical exercise in the heart's physiological properties, during ventricular fibrillation. With that purpose, different kinds of classifiers (linear and neural models) were used to classify between trained and sedentary rabbit hearts. These classifiers were used to perform knowledge extraction through a wrapper feature selection algorithm. The obtained results showed the higher performance of the neural models compared to the linear classifier (higher performance measures and higher dimensionality reduction). The most relevant features to describe the benefits of physical exercise are those related to myocardial heterogeneity, mean activation rate and activation complexity.

Keywords:

Machine Learning, Classification, Knowledge Extraction, Logistic Regression, Multilayer Perceptron, Extreme Learning Machine, Ventricular Fibrillation, Physical Exercise.

1. Introduction

Several authors have proposed that physical exercise (PE) modifies the sympathetic-vagal balance of autonomic nervous system, producing an increase of parasympathetic activity that manifests in a decrease of cardiac frequency [1]. Besides, this vagal activity modification could have protective effects against the appearance of cardiac arrhythmias and death [2]. Other effects of PE are based on changes in the physiological properties of the hearth, and are called intrinsic modifications. The effects of such intrinsic modifications caused by PE, as an increment of the action potential duration, were previously reported [3].

Other previous studies have shown that physical exercise modifies Ventricular Fibrillation (VF) response by intrinsic mechanisms [4]. Those modifications were found in several parameters derived from frequency and time domains [4], and its spatial distributions [5]. There are a high amount of parameters to describe VF signals, and not all of them can describe these effects of PE in the same way [6].

The purpose of this work is to find which of the previously used parameters better explains the intrinsic modifications produced in VF by PE by extracting knowledge from machine learning classifiers. Recordings acquired from two groups of isolated rabbit hearts (trained with PE and untrained) were analysed using a wrapper feature selection algorithm [7]. This wrapper was applied to several machine learning classifiers, to find the most relevant features in the classification between records from trained and untrained rabbits.

Finding the best features to describe the benefits of PE involves a double profit. On one hand, it helps to understand the mechanisms underlying the modifications due to PE, suggesting what features of the VF signals are modified. Those features can be transferred to the heart physiological characteristics. On the other hand, it allows better analysis of these intrinsic modifications, improving future works by the use of the most relevant features.

Feature selection is a widespread application field of machine learning. There are many applications of feature selection in different areas of biomedical engineering [8]. Regarding applications of feature selection methods in VF analysis, the most common applications are related to arrhythmia discrimination [9], and classification of ECG signals [10]. This paper uses a wrapper feature selection algorithm based on the analysis of the perfomance of some classifiers using, or not, each input feature [7]. This feature selection will identify the best features to classify between two groups of VF records and, therefore, it will find the most relevant features to characterize the differences between both groups.

Next section presents data acquisition and explains the features used with the classifiers. Afterwards, these classifiers are explained: Logistic Regression (classical method in statistics) [11]; Multilayer Perceptron (the most extended neural model) [12] and, finally, the Extreme Learning Machine [13] (a new algorithm to find the parameters in Multilayer Perceptrons with one hidden layer). Next section shows the obtained results and final section presents the conclusions.

2. Methods

The proposed study consists in four stages: data acquisition, data processing, classification and knowledge extraction. Electrograms were acquired from two groups of rabbit hearts. Next, these electrograms were processed measuring four parameters from time and frequency domains. Using these parameters, 18 features were calculated and used to perform a classification between the groups in the experiment. Finally, these classifiers were analysed with a wrapper feature selection algorithm to perform knowledge extraction, analysing the relevance of the different features and performing subset selections. Figure 1 shows a diagram of the proposed study.

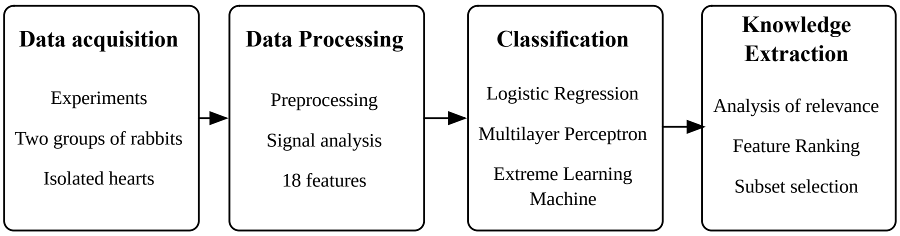

Fig. 1 Proposed study workflow

Next subsection, data acquisition, explains the first stage of the study. Further subsections explain data processing, classification and knowledge extraction stages.

2.1. Data acquisition

Twenty-one male New Zealand white rabbits (*Oryctolagus cuniculus*) were used in the present study. Animals were divided into two experimental groups: an untrained group (G1, with a sample size of 10) and a trained group (G2, with a sample size of 11). Animals in the untrained group were housed in the animal quarters for 46 days, and rabbits in the trained group were submitted to a physical exercise program. After familiarization with treadmill running for four days, animals in the trained group ran five days/week for 6 weeks at 0.33 m s^{-1}. Each training session was divided into six periods of 4 min of running and 1 min of rest [14]. The correct execution of treadmill exercise was constantly supervised, and those animals that did not adequately run on the treadmill because they either stopped frequently or ran irregularly were excluded from the study. Housing conditions and experimental procedures were in accordance with the European Union regulation on the use of animals for scientific purposes (*2003/65/CE*) and as promulgated by Spanish legislation (*RD 1201/2005*). Besides, the *University of Valencia Animal Care and Use Committee* approved all the procedures used in this study.

In order to analyse the intrinsic modifications of cardiac response in VF, isolated hearts were used to make them independents of vagal influence. Perfusion was maintained with a *Langendorff* system in order to avoid metabolic deterioration [15].

Cardiac mapping recordings were acquired with a commercial 256-channel system (*MAPTECH, Waalre, The Netherlands*). An electrode array of 240 electrodes (interelectrode distance of 1 mm) was localized on left ventricle. Each recording had 5 minutes of duration, acquired at a sampling rate of 1 kHz. VF was induced by pacing with increasing frequencies with an electrode placed in the ventricle, outside of the array capturing electrode.

2.2. Data processing

The procedure undergone to analyse the recordings involved a pre-processing stage, a frequency domain analysis and a time domain analysis. These analyses measured four parameters from which 18 features were calculated.

a) *Pre-processing stage.* Recordings were processed in consecutive segments of four seconds. The signal quality of each 4seconds-segment was analysed, discarding the signals of those electrodes in the array with low amplitude or a high presence of noise [4].

b) Frequency domain analysis. Welch spectrum was obtained for all recording electrodes in each segment, using a *Hanning* window (2 non-overlapped sections and zero padding until 4096 samples). The Dominant Frequency (DF) and the Normalized Energy (NE) were calculated [16]. The DF was determined as the frequency with maximum spectral energy. The NE was defined as the spectral energy in a window placed at DF ± 1Hz, and normalized with spectral energy in the interest band (5 - 35 Hz).

c) *Time domain analysis.* In order to analyse VF regularity and organization, two parameters were calculated: Regularity Index (RI) and Number of Occurrences (NO). The algorithm used for the RI computation [4] is a modification of the original [17], in order to adapt it to the electrophysiological characteristics of the used cardiac model. More precisely, the local activation wave duration was increased up to 50 ms. The number of occurrences (NO) was also calculated as the ratio of samples which amplitude is inside a zero centred window respect to the total number of samples [18].

With this procedure, DF, NE, RI and NO parameters were sequentially calculated for each electrode and temporal segment, obtaining one map for each parameter and time segment. The first eight features were obtained as the mean value (*mDF,mNE,mRI,mNO*) and standard deviation (*sDF,sNE,sRI,sNO*) of each parameter map. The variation coefficient of the Number of Occurrences maps (*vcNO*) was also computed.

An algorithm has been implemented to study the Regions Of Interest (ROI) [5], previously used on the VF analysis for dominant frequency [19]. To obtain the ROI, a threshold was applied to each parameter map. Later on, a ROI membership label is assigned to each electrode, according to the threshold criteria and its neighbourhood with electrodes that also passed the threshold. From this ROI analysis, three features were obtained for each DF, NE and RI parameter map [5]:

ROI spatial number (ROIsnDF, ROIsnNE, ROIsnRI): the number of ROI detected in a map, a measure of spatial fragmentation.

ROI spatial area (ROIsaDF, ROIsaNE, ROIsaRI): the percentage of the area map occupied by ROI, where higher percentages implies a higher homogeneity.

ROI electrode number (ROIenDF, ROIenNE, ROIenRI): the number of electrodes whose membership to a ROI changed between two consecutive maps, related to the temporal change of ROI size.

As result of data processing stage, 18 features were computed. Table 1 shows the computed features and its brief description. These features were computed each 4 seconds, obtaining a total of 1626 samples; 814 from control group (negative class) and 812 from trained group (positive class).

Feature	Parameter	Description
mDF	*Dominant Frequency*	*Mean value in DF parameter map*
sDF	*Dominant Frequency*	*Standard deviation in DF parameter map*
ROIsaDF	*Dominant Frequency*	*Area covered by ROI in DF parameter map*
ROIsnDF	*Dominant Frequency*	*Number of ROI in DF parameter map*
ROIenDF	*Dominant Frequency*	*Changes in area covered by ROI between consecutive DF maps*
mNE	*Normalized Energy*	*Mean value in parameter NE map*
sNE	*Normalized Energy*	*Standard deviation in parameter NE map*
ROIsaNE	*Normalized Energy*	*Area covered by ROI in parameter NE map*
ROIsnNE	*Normalized Energy*	*Number of ROI in parameter NE map*
ROIenNE	*Normalized Energy*	*Changes in area covered by ROI between consecutive NE maps*
mRI	*Regularity Index*	*Mean value in RI parameter map*
sRI	*Regularity Index*	*Standard deviation in RI parameter map*
ROIsaRI	*Regularity Index*	*Area covered by ROI in RI parameter map*
ROIsnRI	*Regularity Index*	*Number of ROI in RI parameter map*
ROIenRI	*Regularity Index*	*Changes in area covered by ROI between consecutive RI maps*
mNO	*Number of occurrences*	*Mean value in NO parameter map*
sNO	*Number of occurrences*	*Standard deviation in NO parameter map*
vcNO	*Number of occurrences*	*Variation coefficient in NO parameter map*

Table 1 Features used in the classification. Parameter related to each feature and a brief description

2.3. Classification

This section presents the classifiers used on the data. Logistic Regression (LR), which can only solve linearly separable problems (if inputs are not transformed by some function) and a non-linear classifier, the Multilayer Perceptron (MLP). MLP was trained with *Levenberg-Mardquardt* algorithm, and with a new paradigm: the Extreme Learning Machine (ELM).

2.3.1. Logistic Regression

Logistic Regression (LR) is a generalized linear model for classification. LR models the probability of an outcome in terms of some predictor variables [11]. Let $p=Pr(y=1|X)$ denote the probability of an outcome, LR models $log[p/(1-p)]$ as a function of the predictor variables. In LR, we have the relation defined by equation 1 (here $x_0=1$ and x_1 to x_n are the input variables) [11].

$$\Pr(y=1|X)=\frac{1}{1+e^{-\sum_{i=0}^{n}\omega_i\cdot x_i}} \qquad (1)$$

Each one of the parameters describes the contribution of a variable to the output model; a positive coefficient means that the explanatory variable increases the probability of the outcome, while a negative regression coefficient means that the variable decreases the probability of that outcome. Looking at the absolute value of a coefficient allows analysing the relevance of the input variable related to the coefficient if inputs are correctly normalized.

2.3.2. Multilayer Perceptron

Multilayer Perceptron (MLP) is a layered arrangement of individual non-linear computation units, called artificial neurons, organized in different layers. The neurons of a layer feed the neurons of the next layer with their outputs [12].

These neurons are grouped in layers, forming a fully connected network. The first layer contains the input nodes, usually fully connected to the next layers (hidden layers), in turn, connected to the output layer. Only one output neuron is necessary in our case, since we are tackling binary classification. Multilayer perceptron uses a learning algorithm to find the best parameters to model the relationship between input and output variables [12]. This objective can be fulfilled with a minimization procedure of a cost function that depends on the difference between the obtained output and the desired output value [12].

When the used learning algorithm consists on the minimization of a cost function, an excessive adjustment of the model parameters during the minimization of the cost function can be achieved [12]. In order to avoid this over-fitting, data set is usually divided into three subgroups: training set, validation set, and test set [12]. Other alternative to avoid over-fitting can be the use of regularization terms in the cost function.

2.3.3. *Extreme Learning Machine*

Extreme Learning Machine (ELM) was proposed by Huang et al [13]. It is a learning algorithm that is also applied to MLP with single hidden layer. ELM assumes that the weights of the hidden layer can be randomly assigned [13], thus being only necessary the optimization of the output layer weights. Such optimization can be carried out by the pseudoinverse of the *Moore-Penrose's* matrix [13]. ELM allows reducing the computational time needed for the optimization of the parameters. Gradient-descent methods, or global search methods involve a much longer time.

2.4. *Feature selection method*

The feature selection method used in this work is a wrapper applied to the previously described classifiers [7]. Here, the modification produced in the behaviour of the classifiers when each feature is individually cancelled is studied. The deviation found in the output was the criterion to perform such analysis, also called sensitivity analysis [6].

In Logistic Regression, this analysis can be performed by means of the absolute value of the coefficient associated to each input [11]. The relevance of the different features in the Multilayer Perceptron was obtained with an specific algorithm. The authors have already used this technique with excellent results in other works, and it consists on the steps detailed above [6][20][21][22].

1. Selection of the N_b models with product between sensitivity and specificity in test subset higher than its 90th percentile.

2. Obtaining of the outputs with these N_b models, using all the dataset (denoted by o_k where k refers the pattern number).

3. Individual cancellation of each feature, x_i, and obtaining of the output for all dataset without such feature (denoted as o_k^i).

4. Computation of the model sensitivity, as indicates equation 2 (here N_p is the number of patterns, and *i* refers to the *i-th* feature):

$$S(i) = \frac{1}{N_p} \sum_{k=0}^{N_P} \left| o_k - o_k^i \right| \qquad (2)$$

5. Ranking the features according to *S(i)*, for each one of the N_b considered models. Higher values of *S* are given to the most relevant features; *S* near to zero implies that the same output is obtained either using or not that features, i.e. it is low relevant.

6. The positions are averaged for the N_p considered models.

As result, the features were ranked according to its relevance in each model. An aggregation of the obtained four rankings was calculated to estimate the joint relevance in all classifiers. The relevance of each feature in each classifier was quantified as a score inversely proportional to its order in the ranking. These scores were averaged as indicates equation 3.

$$V_{i,global} = \frac{\sum_{j=1}^{N} V_{ij} \cdot AUC_j}{\sum_{j=1}^{N} AUC_j} \qquad (3)$$

Where, $V_{i,j}$ is the obtained score by the *i-th* feature in the *j-th* classifier; AUC_j is the *Area Under the Receiver Operating Characteristics* (ROC) *Curve* of the *j-th* classifier; and N is the number of classifiers (4 in our case). This aggregation uses a weighted average, so that the best classifier has more impact than the other ones. This procedure obtains thus a new list, called global list.

3. Results

3.1. Classification achieved by classifiers

Data were divided in three data subsets, maintaining the ratio between the number of samples in positive and negative classes. Training data, 2/3 of total data, were used to adjust the classifiers; validation data (1/6) were used in the case of MLP to perform early-stopping, and test data (1/6) where used to assess the classifiers with new samples.

Logistic Regression was adjusted by classical optimization methods, as it is a generalised linear model [11]. Regarding Multilayer Perceptron, two different algorithms were tested: *Levenberg-Marquardt* and *Extreme Learning Machine*.

Levenberg-Marquardt (LM) algorithm was applied used in using *early-stopping* in order to avoid *overfitting* [12]. Different architectures were used (number of used hidden nodes ranged from 2 to 10 in each layer, and number used of hidden layers varied from one to two). Each architecture was initialized 100 times with random weights to avoid the local minima problem; this initialization was done using normal distributions with zero mean and low variance to avoid initial neurons saturation [12].

Regarding *Extreme Learning Machine*, the number of neurons was increased, as it usually needs a higher number of computing units due to the random determination of the hidden layer parameters [13]. Therefore, the hidden nodes ranged from 40 to 80. The number of used random adjustments on each architecture was 100.

The highest product between Sensitivity and Specificity in test data was used to select the best models among the different used architectures, in case of MLP. Table 2 shows the performance measures obtained with the different models. All classifiers show a good behaviour, and MLP with one and two hidden layers trained with LM, were the best performers. LR is the classifier with the lowest performance measures, due to the fact that it is unable to perform non-linear classification.

Training data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
LR	*81.33*	*82.11*	*81.33*	*79.01*	*87.11*
MLP-LM 1 layer	***98.04***	*96.07*	*96.15*	***97.99***	*99.22*
MLP-LM 2 layers	*97.55*	***97.79***	***97.79***	*97.55*	***99.75***
MLP-ELM	*89.65*	*89.95*	*89.65*	*87.59*	*95.38*

Validation data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
LR	*82.35*	*82.44*	*82.35*	*82.44*	*89.63*
MLP-LM 1 layer	***92.20***	*93.14*	*93.10*	***92.23***	***98.23***
MLP-LM 2 layers	*87.80*	***94.12***	***93.75***	*88.48*	*96.57*
MLP-ELM	*87.62*	*87.80*	*87.62*	*86.96*	*93.54*

Test data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
LR	*80.21*	*81.86*	*80.21*	*75.91*	*84.58*
MLP-LM 1 layer	***96.57***	*90.64*	*91.20*	***96.34***	*97.59*
MLP-LM 2 layers	*95.10*	***93.10***	***93.27***	*94.97*	***97.82***
MLP-ELM	*89.84*	*90.69*	*89.84*	*84.09*	*92.36*

Table 2 Performance measures of the used classifiers, using all the features. *PPV*: Positive Predictive Value; *NPV*: Negative Predictive Value; *AUC*: Area Under the ROC Curve; *MLP-LM*: MLP trained with LM algorithm; *MLP-ELM*: MLP trained with ELM.

Cohen's kappa coefficient was obtained in order to check the agreement level among those classifiers. *Cohen's kappa* coefficient is a statistical measure of inter-rater agreement for qualitative items [23]. It is a more robust measure than ordinary agreement calculation, since it deals the agreement occurring by chance. Table 3 shows the obtained kappa coefficients among all the used classifiers, using the test subset. The highest agreement was observed among the MLPs, trained with LM or ELM, due to the fact that those classifiers use have similar architectures. In the same way, the low *kappa* obtained between LR and the other classifiers is due to the existent differences in its architectures.

Cohen's Kappa	*MLP-LM 1*	*MLP-LM 2*	*MLP-ELM*
LR	*0.5950*	*0.5668*	*0.6735*
MLP-LM 1	-	***0.8917***	*0.7531*
MLP-ELM	-	*0.7243*	-

Table 3 *Cohen's kappa* coefficient between the best classifiers. *MLP-LM 1*: MLP trained with LM using one hidden layer; *MLP-LM 2*: MLP trained with LM using two hidden layers.

3.2. Feature selection analysis

The wrapper explained in section 2.4 was applied to the classifiers. This scheme consists on the analysis of the individual effect of each feature. In case of MLP classifiers, there is not only a single model, since several architectures and initializations were used. Therefore, this analysis is applied to the best N_b architectures and initializations with product between sensitivity and specificity higher than its 90th percentile. Figure 2 shows boxplots of several performance measures of the N_b selected models for MLP-LM with one and two hidden layers and MLP-ELM; note that the selected MLP trained with LM (with one or two hidden layers) has higher rates than the ELM alternative.

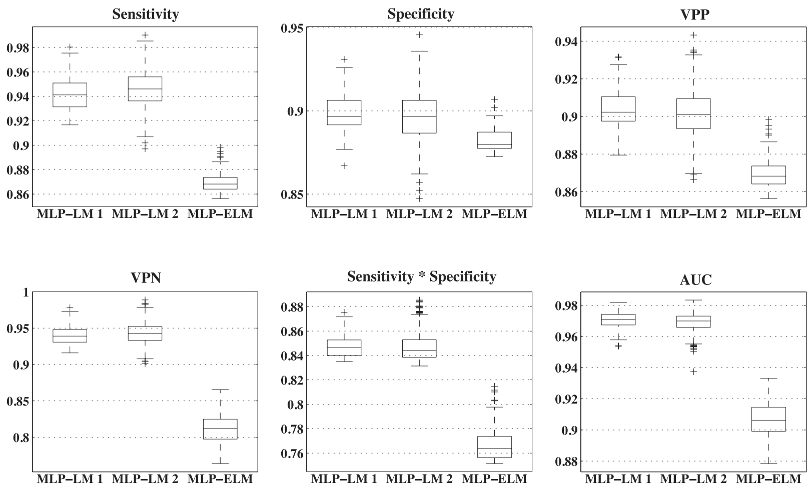

Fig. 2 Performance measures of the N_b selected models to perform the feature selection.

Figure 3 shows the results of the feature selection algorithm. The relevance obtained for each one of the 18 features with the four used models, and the relevance obtained in the global ranking is represented. Here, the most relevant feature has a score of 100, the second a punctuation of 94.44, and so on.

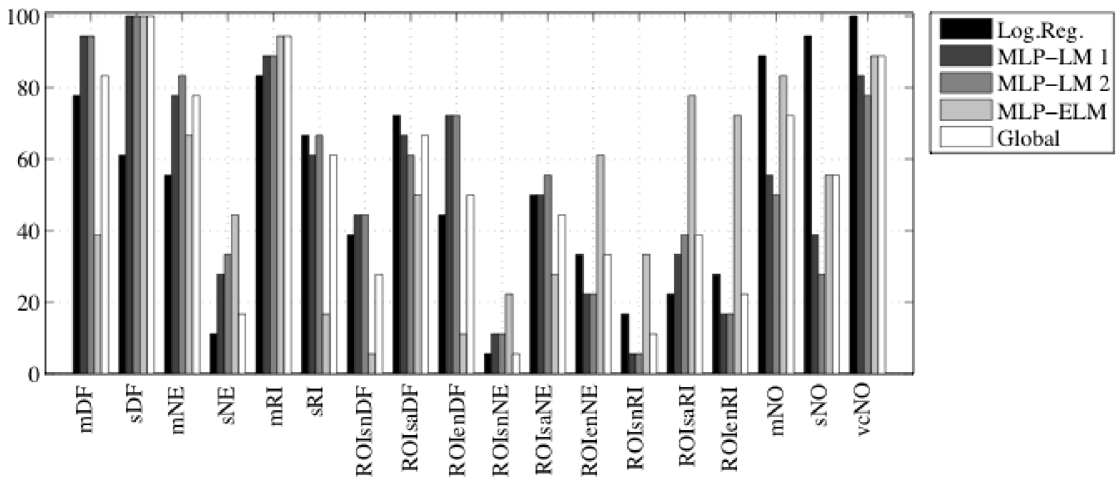

Fig. 3 Scores obtained by each feature with each classifier, and the aggregation of the scores in all the classifiers (global score).

Table 4 shows the six most relevant features for each classifier, and aggregating the relevance in the four classifiers. Note that MLP-LM with 1 and 2 hidden layers obtain similar feature rankings, while LR is the classifier with the more differenciated feature ranking. MLP-

ELM presents a feature ranking more similar to han to LR. Thsese similarities among the used classifiers were previously found in terms of *kappa* coefficient, and are related to the analogies among its structures. The features of highest performance are *sDF* and *mRI*. The features related to DF and RI parameters have a high relevance, as shows the global list where these parameters appear in four of the six most important features.

Classifier	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
LR	*vcNO*	*sNO*	*mNO*	*mRI*	*mDF*	*ROIsaDF*
MLP-LM 1	*sDF*	*mDF*	*mRI*	*vcNO*	*mNE*	*ROIenDF*
MLP-LM 2	*sDF*	*mDF*	*mRI*	*mNE*	*vcNO*	*ROIenDF*
MLP-ELM	*sDF*	*mRI*	*vcNO*	*mNO*	*ROIsaRI*	*ROIenRI*
Global	***sDF***	***mRI***	***vcNO***	***mDF***	***mNE***	***sRI***

Table 4 Six most important features for each considered model, and aggregating the relevance in the four models (global list).

A feature subset selection is done with the results shown in figure 2. This subset was selected adding features in the order established on the rankings. The criteria to stop adding new features uses the quantity *S(i)*, computed as was indicated in equation 2. This quantity was accumulated for the N_b best architectures of all MLP-LM and MLP-ELM models. Finally, this score was normalized. The subset includes the minimum number of features needed to accumulate a normalized score higher than a threshold (th_S), established in $th_S=0.5$.

This procedure was firstly done with the LR model, scoring features according to the absolute value of its coefficient in the model, instead of *S(i)*. These scores were normalized and accumulated until pass th_s. As result, a subset of two features was selected, which misclassifies all negative patterns in the dataset. These results showed that the used subset selection algorithm does not provide accurate results with linear models.

Next, this subset selection algorithm was applied to MLP. The obtained subset lengths depended on the training algorithm; it consisted on six features for MLP trained with LM (one and two hidden layers) and nine features for MLP trained with ELM. Table 4 summarizes the obtained results with those subset selections. A higher performance of MLP when is trained with LM algorithm was observed, comparing to ELM algorithm. Besides, the results provided by the subset with MLP with two hidden layers (*Se=97.06 Sp=95.07 PPV=95.19 NPV=96.98 AUC=99.15*), overcome the results using all the features (*Se=92.68 Sp=94.61 PPV=94.50, NPV=92.34 AUC=97.72*) particularly in test data. This improvement suggests that the features discarded by this subset are noisy features that trigger the misclassification of some patterns.

Training data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
MLP-LM 1 (6)	*93.38*	*90.42*	*90.71*	*93.16*	*98.13*
MLP-LM 2 (6)	***98.28***	***97.54***	***97.57***	***98.27***	***99.46***
MLP-ELM (9)	*86.70*	*87.25*	*86.70*	*83.96*	*92.90*

Validation data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
MLP-LM 1 (6)	***92.68***	*90.69*	*90.91*	***92.50***	*97.72*
MLP-LM 2 (6)	*92.20*	***94.61***	***94.50***	*92.34*	***98.22***
MLP-ELM (9)	*84.08*	*84.39*	*84.08*	*83.17*	*91.02*

Test data

Classifier	*Sensitivity*	*Specificity*	*PPV*	*NPV*	*AUC*
MLP-LM 1 (6)	*94.12*	*92.12*	*92.31*	*93.97*	*97.37*
MLP-LM 2 (6)	***97.06***	***95.07***	***95.19***	***96.98***	***99.15***
MLP-ELM (9)	*90.59*	*92.16*	*90.59*	*79.32*	*91.77*

Table 5 Results obtained using the feature subsets; the number in brackets is the subset length.

4. Discussion and conclusions

A comparison among three classification algorithms was carried out. Two different algorithms for training Multilayer Perceptrons were compared, *Levenberg-Mardquardt* algorithm (LM) and Extreme Learning Machine (ELM). These models were compared using a linear classifier (Logistic Regression) as reference.

Results provided by LM overcome those achieved by ELM. LM obtains higher prediction power when new data is presented, i.e. higher performance measures in test dataset. Besides, this higher performance is achieved with a fewer number of hidden nodes (in case of MLP with one hidden layer), showing the benefits of optimizing all parameters in a Multilayer Perceptron instead of randomizing the hidden layer weights as proposes ELM. The linear classifier provided lower performance measures in all data subsets, showing non-linear components among the used features and the group (trained or untrained rabbits).

Knowledge was extracted from these classifiers, to analyse the relevance of the 18 features in the classification. A wrapper feature selection algorithm was applied, where the output deviation produced when each feature is individually cancelled provided a score to rank the features. This score was obtained for the best architectures in case of MLP (trained both with LM and ELM). These best structures provided better classifications when where trained with LM algorithm. Besides, the aggregated score (global ranking) showed the joint relevance in all classifiers. This global score has provided a ranking of features with physiological meaning, showing the goodness of the feature ranking method.

These rankings allowed the creation of feature subsets. This method has improved the behaviour of the MLP with 2 hidden layers respect to the performance using all the input space. Besides, using all the input space it was not clear which number of hidden layers has performed a better classification (i.e. some performance measures were higher for single hidden layer and others for the two layered version). The dimensionality reduction performed by our subset selection algorithm has clarified this fact. With the selected subset, the best model was the MLP using two hidden layers and trained with LM algorithm.

In this way, not all the used features can describe the modifications between the two groups with the same relevance. The knowledge extraction performed showed that some features allow a better analysis of the intrinsic modifications produced by PE in VF.

The high relevance of the parameters based on the mean value of RI and NE shown that these intrinsic modifications are also based on changes in regularity of activations, measured in time and frequency domains. An increase in regularity implies the existence of local activation waves with a decrease of complexity, which is related to the existence of lesser complex activation patterns [24]. Moreover, the upper relevance of the features related to standard deviation of NE and NO showed that the dispersion in the wave morphology and spectral complexities are essential to characterize the benefits of PE. In that way, these results suggest that PE modifies the complexity of activation patterns by intrinsic mechanisms.

Regarding mean and standard deviation of DF, are related to ventricular refractoriness and action potential duration. It is known that physical exercise influences in ventricular refractoriness [25]. The modification of these physiological features helps to stabilize cardiac electrical refractoriness, and thus also helps to prevent sudden cardiac death, caused in most cases by reentrant arrhythmias as VF [26].

Features related to ROI, i.e. related to the parameters spatial distribution, where ranked with low scores; except those that involve DF. Such DF features, especially ROIsnDF and ROIenDF, where the most relevant within the ROI features. The spatial homogeneity of refractoriness period (related to DF) is highly connected with fibrillatory rhythms, where higher homogeneity of refractoriness period involves lesser facilities in initiation of fibrillatory arrhythmias [25].

In that way, can be assumed that physical exercise produces intrinsic modifications in VF response, and such modifications can be better analysed with features related to activation rate and activation complexity rather than with other VF features.

5. Future work

Feature selection is an unsolved problem, and the high amount of existing techniques to perform feature selection implies the existence of several future work directions. In that way, other wrapper methods can be applied to the same classifiers used in this work. Besides, there are alternatives to the wrapper feature selection. There are many different feature selection techniques based on filter methods, which uses statistical measures of relevance.

As alternative, the wrapper used in this work can be applied to other classifiers. Classification is another important application field of machine learning, and there is a high diversity of classifiers. For instance, Support Vector Machines or Decision Trees could be applied to this classification problem. Such algorithms also allow the study of input relevance by the application of wrapper feature selection methods.

AKNOWLEDGEMENTS

This work was supported in part by the Spanish Ministry of Science and Innovation: Plan Nacional de I+D+I, Acción Estratégica: *"Deporte y Actividad Física"*, project DEP2010-22318-C02-02.

REFERENCES

[1] C.G. Blomqvist , B. Saltin, Cardiovascular adaptations to physical exercise, Rev. Physiol. 45, 169-189 (1983).

[2] G.E. Billman, The effect of carbacholandcyclic GMP on susceptibility to VF, FASEB J. 4, 1668-1673 (1990).

[3] A.J. Natali, L.A. Wilson, M. Peckham, D.L. Turner, S.M. Harrison, and E. White, Different regional effects of voluntary exercise on the mechanical and electrical properties of rat ventricular myocytes, J. Physiol. (Lond) 541, 863-875 (2002).

[4] J. Guerrero, A. Rosado-Muñoz, A.J. Serrano, M. Bataller, J. Chorro, A. Alberola, L. Such, Modifications on regularity and spectrum of ventricular fibrillation signal induced by physical training, Computers in Cardiology, 321-324 (2009).

[5] J. Guerrero, A. Rosado-Muñoz, A.J. Serrano, M. Bataller, J. Chorro, A. Alberola, L. Such, Analysis of spatial and temporal evolution of regularity maps during ventricular fibrillation, Computers in Cardiology, 325-328 (2009).

[6] J. Caravaca, A.J. Serrano-Lopez, E. Soria-Olivas, P. Escandell-Montero, J.M. Martinez-Martinez, J.F. Guerrero-Martinez, Analysis of ventricular fibrillation signals using feature selection methods, 3rd International Workshop on Cognitive Information Processing, 28-30 (2012).

[7] R. Kohavi, G.H.John, Wrappers for feature subset selection, Artificial Intelligence 97, 274-324 (1997).

[8] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics Bioinformatics 23, 2507-2517 (2007).

[9] C. Zhang, J. Zhao, J. Tian; F. Li, H. Jia , Support vector machine for arrhythmia discrimination with TCI feature selection, IEEE 3rd International Conference on Communication Software and Networks, 111-115 (2011).

[10] Z. Zhen-Xing, L. Sang-Hong, L.S. Joon, Comparison of feature selection methods in ECG signal classification, Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, 73, (2010).

[11] D.G Kleinbaum, M. Klein, Logistic Regression: A Self-Learning Text, Springer, 2002.

[12] S. Haykin, Neural Networks and Learning Machines, third ed., Englewood Cliffs, Prentice-Hall (2009).

[13] G.B. Huang; Q-Y. Zhu; C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks. Procedings IEEE International Joint Conference on Neural Networks 2, 985-990 (2004).

[14] L. Such, A.M. Alberola, L. Such-Miquel, L. López, I. Trapero, F. Pelechano, M.C. Gómez-Cabrera, A. Tormos, J. Millet, F.J. Chorro, Effects of chronic exercise on myocardial refractoriness: a study on isolated rabbit heart, Acta Physiol. (Oxf) 193, 331-339 (2008).

[15] M. Skrzypiec-Spring, B. Grotthus, A. Szeląg, R. Schulz, Isolated heart perfusion according to Langendorff—Still viable in the new millennium, Journal of Pharmacological and Toxicological Methods 55, 113-126 (2007).

[16] P. Sanders, O.Berenfeld, H. Mélèze, J. Pierre, R.Vaidyanathan, L.Hsu, S.Garrige, Y.Takahashi, M.Rotter, F.Sacher, C.Scavée, R. Ploutz-Snyder, J. Jalife, M. Haïssaguerre, Spectral Analysis Identifies Sites of High-Frequency Activity Maintaining Atrial Fribrillation in Humans, Circulation 112-6, 789-797 (2005).

[17] L. Faes, G. Nollo, R. Antolini, F. Gaita, F. Ravelli. A method for quantifying atrial fibrillation organization based on wave morphology similarity. IEEE Trans. Biomed. Eng. 49, 1504-1513 (2002).

[18] G. Calcagnini, F. Censi, A. Michelucci, P. Bartolini, Descriptors of wavefront propagation. IEEE Eng. Med. Biol. Mag. 25, 71-78 (2006).

[19] B. Choi, W. Nho, T. Liu, G. Salama, Life span of ventricular fibrillation frequencies. Circ. Res. 91, 339-345 (2002).

[20] E. Soria, J.D. Martin, J. Caravaca, A.J. Serrano, M. Martinez, R. Magdalena, J. Gomez, M. Heras, G. Sanz, Survival prediction in patients undergoing ischemic cardiopathy, International Joint Conference on Neural Networks, IJCNN, 2979-2982 (2009).

[21] O. Pastor-Bárcenas, E. Soria-Olivas, J.D. Martín-Guerrero, G. Camps-Valls, J.L. Carrasco-Rodríguez, S. del Valle-Tascón, Unbiased sensitivity analysis and pruning techniques in neural networks for surface ozone modelling. Ecological Modelling 182, 149-158 (2005).

[22] J. Gómez, J.D. Martín-Guerrero, E. Soria-Olivas, J. Vila-Francés, J.L. Carrasco, S. del Valle-Tascón, Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration, Atmospheric Enviroment 40-32, 6173-6180 (2006).

[23] J.A.Cohen, Coefficient of agreement for nominal escales, Educational and Psychological Measurement 20, 37-46 (1960).

[24] K.T. Konings, J.L. Smeets, O.C. Penn, H.J. Wellens, M.A. Allessie. Configuration of unipolar atrial electrograms during electrically induced atrial fibrillation in humans, Circulation 95, 1231-1241 (1997).

[25] L. Such, A.M. Alberola, L. Such-Miquel, L. López, I. Trapero, F. Pelechano, M.C. Gómez-Cabrera, A. Tormos, J. Millet, F.J. Chorro, Effects of chronic exercise on myocardial refractoriness: a study on isolated rabbit heart, Acta Physiol. 193, 331-339 (2008).

[26] G.E. Billman, Aerobic exercise conditioning: a nonpharmacological antiarrhythmic intervention, J Appl. Physiol. 92, 446-454 (2002).