Non-linear data mining methods to assess the impact of physical training on the cardiovascular system of subjects from different age groups

R. Pizzi, S. Siccardi, C, Pedrinazzi, O. Durin and G. Inama

***Abstract*—** A set of non-linear data mining methods have been applied to ECG signals and other cardiovascular and blood parameters to evaluate the cardiovascular response to exercise in young and master athletes, compared with control groups of untrained subjects of the same age. Methods include PNN calculation, Multiscale Entropy analysis (MSE), and a comparison between clustering and an Artificial Neural Network analysis performed by means of chaotic attractors.

After recruiting four groups of healthy athletes and sedentary subjects, with age under and over 40, we analyzed the collected data, obtaining cross-validated classifications and significant variable differences among clusters.

The analyses lead to a good stratification of the subjects, establishing some important relationships between physical activity, age, sex, and cardiovascular parameters. In particular the existence of significant differences in the cardiovascular status of these groups was shown, depending in particular on the MSE1, PNN20, VO and FC variables.

This will make it possible a follow-up of the subjects, analyzing the above specified parameters over time, in order to identify possible markers of increased arrhythmic risk, useful to prevent fatal cardiac events.

***Keywords* -** Arrhythmic risk, athletes, Artificial Neural Networks, clustering, electrocardiography, data mining, sport, training.

I. Introduction

A. *Background*

The presence of cardiac remodeling in athletes is known since the 800, when the techniques of cardiac percussion allowed to identify an increase in heart volume in patients practicing sport at high level. Later, the introduction into the clinical practice of the echocardiographic techniques has allowed a better understanding of the athlete's cardiac physiology and the impact of different sports and exercise training modes on the heart structure and function, allowing to characterize in detail the framework of the "athlete's heart". Moreover, the modern methods of non-invasive and invasive electrophysiology contribute to deepen the possible correlations between the type and intensity of athletic activity and cardiac arrhythmias [1,2,3].

Most young athletes who die suddenly, however, are suffering from structural heart disease not previously suspected or present in an asymptomatic form and hidden, but accelerated and highlighted by a particularly intense physical activity, which can be used for example as the trigger for the phenotypic expression of diseases genetically determined but previously not expressed, therefore not identifiable through clinical and instrumental methods. In particular, the hypertrophic cardiomyopathy is responsible for 30% of fatal cases in the United States, and about 25% of fatal cases in Italy is caused by arrhythmogenic right ventricular cardiomyopathy. Recent trials [4,5,6,7,8,9,10] have shown that vigorous exercise may affect a structural remodeling in particular of the right ventricle that can sometimes predispose to the development of potentially fatal ventricular arrhythmias.

In particular, physical exercise is associated with hemodynamic changes with alteration of the loading conditions of the heart. In endurance sports physical effort prevails, aimed at the production of motion (e.g. running) with limited development of muscle strength. In this case from the hemodynamic point of view there is an increase in heart rate and stroke volume, that are the two determinants of cardiac output. Peripheral vascular resistances are reduced with only a slight to moderate increase in blood pressure. In this kind of sport the load on the heart is primarily a volume load. These phenomena can lead to a dilation of the left ventricle with a proportional increase in wall thickness (eccentric left ventricular hypertrophy).

On the contrary in power sports (e.g. weightlifting, rowing, etc.) there is development of strength with limited movement. In this case the hemodynamic consequences include only a slight increase in cardiac output, favored by the increase in frequency and a more pronounced increase in blood pressure caused by increased peripheral vascular

R. Pizzi is Senior Researcher at the Department of Computer Science of the University of Milan, Italy (corresponding author, tel. 0039 02 503 30072, e-mail rita.pizzi@unimi.it). S. Siccardi is post-doc Researcher at the the Department of Computer Science of the University of Milan. G. Inama, C. Pedrinazzi and O. Durin are Cardiologists at the Division of Cardiology, Crema Hospital, Crema, Italy.

resistance. These phenomena lead to an increase of the pressure load on the heart and in the long term can facilitate the development of concentric left ventricular hypertrophy with increased wall thickness and maintenance of normal cavity size of the left ventricle.

While literature confirmed the impact of athletics on the cardiovascular system in young subjects, observations on non-young "Master" and "Senior" athletes are completely lacking. In the Cardiology Guidelines and Protocols benefits and limits of physical activity are not clearly defined, as well as the eligibility criteria in the "Master" athletes, population that increasingly turns to the Sports Physician and to the Cardiologist for a training program and sometimes for the release of a certificate of suitability for competitive sports. In particular there are no controlled and prospective works comparing exhaustively the two populations of young (< 40 years) and "Master" (> 40 years) athletes in terms of the cardiovascular adaptation to training.

The proper understanding of the physiological mechanisms of adaptation to exercise in the subject with age > 40 years also assumes an important clinical value useful in the development of the programs of cardiac rehabilitation after myocardial infarction or after coronary angioplasty and in particular in the appropriate prescription of physical training in the post-rehabilitation phase [11]. It should be noted that the most appropriate type of exercise, intensity and frequency of training sessions in cardiopathic subject are still under study. Today alongside young high-level athletes, more and more adults and elderly, also suffering from stabilized cardiovascular problems, practice athletic activity as occupation of free time in order to achieve or maintain a state of well-being and fitness.

The reduction in cardiovascular mortality and cardiac ischemic events in subjects who perform regular physical activity is mainly due to the action that exercise plays on the control of cardiovascular risk factors. In particular, physical training has proved capable of improving lipid profile, reducing blood pressure and body weight, and improving glycemic control. Physical training, in particular combined exercise training, including both aerobic activities and strength training [12,13,14], is currently recommended also in patients with coronary artery disease.

Favorable effects of exercise are known in these patients not only on metabolism, body weight and blood pressure but also on the improvement of the functional capacity, on the reduction of cardiovascular events and on the increasing of endothelial progenitor cells that appear to have an important role in angiogenesis, in the inhibition of intimal proliferation and in vascular repair. And for these reasons, physical activity and sport should always be present in the therapeutic regimen of a cardiopathic patient with custom training programs, meaning physical activity as a therapy.

And as such it must be understood, prescribing it as a drug, specifying in each case indication, route of administration and doses, without forgetting, as with all medicines, overdose problems and side effects, in order to maximize the positive effects and minimize the risks.

B. Purpose of this study

This study, conducted in collaboration between the Unit of Cardiology Ospedale Maggiore of Crema and the Department of Computer Science, University of Milan, was intended to evaluate the cardiovascular response to exercise in young and "Master " athletes even by comparison with control groups of untrained subjects of the same age. We also intended to identify the presence, in athletes, of possible markers of increased arrhythmic risk, correlating the results of the traditional cardiology methods with non-linear analysis of electrocardiographic signals and other clinical data.

C. Subjects included in the study and collected data

We considered four groups of subjects:

1) Group A: athletes practicing endurance sports (cycling or running) with age <40 years

2) Group B: "master" athletes practicing endurance sports (cycling or running) and> 40 years old

3) Group C: healthy subjects aged <40 years not practicing sports

4) Group D: healthy subjects aged> 40 years not practicing sports.

The sample under study was composed of 20 subjects in group A and 20 in group B, 10 in group C and 11 in group D.

ECG signals of the patients were collected during the exercise test.

The subjects participating in the study underwent, as well as the exercise test, also to the following diagnostic tests:

1) History and physical examination

2) Venipuncture with a dose of blood count, urea, creatinine, total cholesterol, HDL and LDL cholesterol, triglycerides, blood glucose, and other parameters.

3) 12-lead resting ECG

4) Trans-thoracic echocardiogram, with assessment of cardiac morphology and function and Doppler examination.

5) Evaluation of the maximum O2 consumption (VO2peak), O2 consumption at the anaerobic threshold (VO2AT) and VE/VCO2 slope (indicator of ventilatory response to exercise) during cardiopulmonary exercise testing

6) 24-hour ambulatory ECG recording, with quantification of ventricular and superventricular ectopies and search for possible pathological bradyarrhythmias or conduction defects

7) T-wave alternans.

Some people did not execute the cardiopulmonary exercise testing (n = 5) mainly due to their intolerance to the mask required for the analysis of expired gases; to the T wave alternans analysis (n = 3) mainly for intolerance to the procedure of skin abrasion in the positioning of the electrodes for the analysis of microvoltage indicated by the protocol (in such cases, in the absence of adequate skin preparation, the results of the examination does not have adequate reliability for the presence of electrical interference) and to dynamic ECG (n = 3). In such cases, the subjects were excluded from the statistical analysis only as regards the procedures that they were not subjected to.

At the end of a first data screening we selected usable data on the following subjects:

1) a group of 19 athletes, "master", i.e. over the age of 40, consisting of 6 female athletes and 13 male

2) a group of 18 young athletes, 5 females and 13 males

3) a control group of 7 sedentary, age greater than 40 years, 2 female and 5 male

4) a control group of 8 young sedentary, 3 female and 5 male

D. Inclusion and exclusion criteria

Inclusion criteria for the case of groups A and B (athletes) included the competitive sports with regularly achieved and still valid certificate of suitability for the sports of running or cycling or alternatively the usual practice of these sports with an intensity and frequency of training comparable to that request from racing, and that was quantified in a number of weekly training sessions equal to or greater than 5, with sessions lasting no less than 90 minutes. The inclusion in group A included age less than 40 years, while in group B subjects aged greater than or equal to 40 years were included.

Groups C and D (sedentary) include subjects who never achieved the certification of fitness for any competitive sports activities and who do not practice regular physical activity, intended as a number of weekly sessions of physical training less or equal to 2. The inclusion in the group C included age less than 40 years, while in group D subjects aged greater than or equal to 40 years were included.

The exclusion criteria were constituted by the presence of known heart disease, previous finding of supraventricular or ventricular arrhythmias, dysthyroidism, cardiovascular risk factors such as hypertension, diabetes mellitus, smoking, dyslipidemia, metabolic and plasma electrolyte panel, the presence of pathological conditions potentially able to lead to a reduction in functional capacity, habitual intake of drugs.

II. Materials and Methods

The analyzes focused firstly on the heart rate variability (HRV), whose importance has been recognized in several studies (see for example [15]). The first step was to derive the instantaneous beat signals by eliminating artifacts .

Then "RR intervals" were calculated, i.e. the duration of the pulse in seconds for all the tracks, with the sole exception of the artifact portions.

Using these data, the following analyses were carried out.

A. PNN Calculation

The method takes into consideration the increase in milliseconds from an RR interval to the next and calculates the distribution. Furthermore it considers individual indicators, such as PNN50 (the probability that the difference between two successive intervals is at least 50 ms.) and PNN20 (the probability that the difference between two successive intervals is at least 20 ms).

According to the literature[16], young healthy subjects had significantly higher values of the PNN compared to subjects older or suffering from heart.

After several tests, we focused on PNN20 that gave better discrimination among subjects.

B. Multiscale entropy analysis (MSE)

The method quantifies the complexity of the data series, taking into account the fact that the physiology dynamics occurs according to multiple time scales [17].

MSEX, i.e. the Multi Scale entropy with scale factor x, is calculated from the data set, constructing the series of averages of the elements taken in groups of x and calculating the sample entropy. So MSE1 = usual sample entropy, MSE5 = sample entropy of the averages of groups of 5 data , MSE20 of groups of 20 data.

The method is reported to separate healthy patients from patients suffering from Congestive Heart Failure (CHF) and Atrial Fibrillation (AF), and young patients by elderly.

C. Clustering

Using anagraphic and behavioural (sedentary/sports) data and the new processed data, we proceeded to a clustering procedure.

Clustering is particularly suitable to find regularities inside large amounts of heterogeneous data [18]. The clustering method applied to a set of discriminant variables gives a global response, i.e. it does not give information of each single variable but of the interaction between all the used parameters.

The hierarchical clustering algorithms [19,20] encountered a large popularity inside the community of biologists [21,22] and physicians [23,24,25] due to their effectiveness. Their purpose is the organization of data in a hierarchical structure, that collects similar observations into larger clusters and so on, up to the whole set of data.

More precisely, we consider a sequence of h partitions of the N data, where n_h is the cardinality of the i_{th} partition.

Then the partitions are gradually joined in pairs, using the general criterion to unify the two groups with minimum distance. Distance can be defined in many ways, suitable for the specific problem, and the hierarchical methods use several techniques for the fusion of observations.

The advantages of hierarchical clustering are bound to the possibility to observe the agglomerative process that gradually leads the units to aggregate and form groups as they move up the hierarchy. This can be of interest to assess the strength of the link between the units.

In this study we used hierarchical clustering, adopting in particular the Ward's method, after a set of tests with alternative methods, because it showed the most satisfactory classification in comparison with the available groups.

The Ward’s method joins at any step the two groups

whose fusion minimize the increase of the deviance “within”, in the meaning well-known in statistics:
Total deviance DEV_T can be decomposed in the sum of deviance “within groups, DEV_{IN}, and deviance “between groups", DEV_{OUT}:

$$DEV_T = DEV_{IN} + DEV_{OUT}$$

$$\sum_{s=1}^{p}\sum_{i=1}^{n}(x_{is}-[\bar{x}_s])^2 = \sum_{s=1}^{p}\sum_{k=1}^{g}(\bar{x}_{sk}-[\bar{x}_s])^2 n_k + \sum_{k=1}^{g}\sum_{s=1}^{p}\sum_{i=1}^{n}(x_{is}] - [\bar{x}_{sk})^2$$

where

$$DEV_{IN} = \sum_{k=1}^{g}\sum_{s=1}^{p}\sum_{i=1}^{n}(x_{is}] - [\bar{x}_{sk})^2$$

is the deviance “within groups” referred to the p variables of the k_{th} group, and $\bar{x}_{sk}$ is the average of the variable s referred to the k_{th} group;

$$DEV_{OUT} = \sum_{s=1}^{p}\sum_{k=1}^{g}(\bar{x}_{sk} - [\bar{x}_s])^2 n_k$$

is the deviance between groups.

Passing from k+1 to k groups (aggregation) the deviance “within” increases, whereas deviance “between” decreases.

At any step of the Ward’s method the groups with minimum increase of the deviance “within” are aggregated or, alternatively, the maximum decreasing of the deviance “between groups”.

For a further discussion of the method see [26,27,28,29].

D. Artificial Neural Network

Another type of classification was carried out using a custom self-organizing Artificial Neural Network.

Artificial Neural Networks and in particular the Self Organizing Maps (SOM) are non-linear models suitable to classify complex patterns [30,31].
But two main reasons exist that limit the SOM’s performances in case of strictly non-linear and time-variant input.

The first reason is that if the input topology is too tangled, the competitive layer is not able to unfold itself enough to simulate the input topology.

The second reason concerns the SOM’s convergence conditions, that are not easily verifiable. Due to the nature of the SOM’s output (non-homologous to the input), it is not possible to settle either a network error for each epoch, or the number of epochs after that the network training has to be stopped.

Another problem of the SOM, typical of any clustering algorithm, is the lack of output explication. Once a classification is obtained, the user must analyze it, comparing it to the input values in order to extrapolate a significant output.

Thus we adopted a model of SOM-like network (ITSOM) developed by our group that aims to overcome these limitations.

We observed that, even if the winning weights may vary at any presentation epoch, their temporal sequence tends to repeat itself. Such a sequence, provided to keep the learning rates steady (instead of gradually decreasing them), constitutes chaotic attractors [32] that repeat “nearly” exactly in time with the epochs succeeding, and that, once codified by the network, univocally characterize the input element that has determined them. An attractor can be defined as a generalization of the steady state point, and represents the trajectory in a portion of state space where a dynamical system is attracted to.

Actually the SOM learning rule makes it possible for the winning weight to represent an approximation of the input value. In this way at every epoch the new winning weight, together with the previous winner, constitutes a second-order approximation of the input value. At the n-th epoch, the set of n winning weights represents an nth-order approximation of the input value.

In this way, due to the countless variety of possible combinations among winning neurons, the configurations allow to determine finely the correct value, even in the case of tangled input topologies, despite of the small number of competitive neurons and their linear topology.

An ITSOM interesting feature is that the network does not need to be brought to convergence, as the cyclic configurations stabilize their structure within a small number of epochs, then keep it steady through time.

After interrupting the network-processing phase, an algorithm is needed that codifies the obtained chaotic configurations into a small set of outputs.

The algorithm that has shown best performances and computational load among the tested pattern recognition algorithms is based on a z-score calculus.

The cumulative scores related to each input have been normalized following the distribution of the standardized variable z given by:

$$z = (x - \mu) / \sigma$$

where μ is the average of the scores on all the competitive layer weights and σ is the root mean squared deviation. Once fixed a threshold $0 < \tau \leq 1$, we have put

$$z = 1 \ \ for \ \ z > \tau,$$
$$z = 0 \ \ for \ \ z \leq \tau.$$

In this way every winning configuration is represented by a binary number with as many 1’s and 0’s as many the competitive layer weights.
Then the task of comparing these binary numbers is straightforward.

The z-score method has shown to be steady with regard of the performances, and computationally not expensive, being linear in the number of the competitive layer weights.

But it is worth emphasizing that the z-score algorithm allows the network to reach its best performances in a very small number of epochs (often less then 15).

This allows the network to complete its work within a negligible time.

A fruitful use of this model in cardiology has already been proposed by some of the authors [33,34]

We have been able to represent the dynamical trajectories generated by the Artificial Neural Network, obtaining chaotic attractors that represent the ECG time series in the state space.

E. Software

The calculation of the instantaneous heartbeat, of RR intervals, PNN, multiscale entropy and multifractal analysis were performed using the Physionet library [35].

The analysis of the clusters was performed with R [36].

The neural network and dynamic analysis system have been developed by the group using C, Matlab and Simulink.

III. Results

A set of first considerations were carried out observing the totality of the available data.

The characteristics of the subjects included in the groups are shown in Table I.

Table I Age, sex and AP

	Group A	Group B	Group C	Group D
Age	30 $\pm$ 9	49 $\pm$ 4	32 $\pm$ 6	49 $\pm$ 7
Sex (male)	15 (75%)	14 (70%)	7 (70%)	7 (64%)
Systolic AP	114 $\pm$ 12	113 $\pm$ 25	117 $\pm$ 8	121 $\pm$ 12
Diastolic AP	73 $\pm$ 10	76 $\pm$ 7	75 $\pm$ 7	77 $\pm$ 6

As regards the echocardiographic parameters we observed a significant increase of the diameter of the tele-diastolic left ventricle in athletes of the group A vs. the age-matched sedentary subjects (50.65 + 4.7 vs. 46.8 + 3.4 mm; p = 0.01), whereas no significant differences were observed in terms of wall thickness, tele-diastolic diameter and ejection fraction. Instead, the difference in tele-diastolic volume was not observed in the groups of athletes and sedentary older than 40 years (Table II).

Table II Echocardiographic parameters

	Group A	Group B	Group C	Group D
TDD	50.65 $\pm$ 4.7	46.8 $\pm$ 3.4	50.05 $\pm$ 5.1	47.9 $\pm$ 3.9
DTS	28.4 $\pm$ 4.4	27.8 $\pm$ 3.8	29.3 $\pm$ 3.1	28.2 $\pm$ 3.5
IVS	9.2 $\pm$ 1	8.4 $\pm$ 0.7	9.4 $\pm$ 1.3	9.2 $\pm$ 1.1
PP	8.9 $\pm$ 1.2	8.5 $\pm$ 1.3	9 $\pm$ 1	9.2 $\pm$ 1.2
EF	63 $\pm$ 6	64 $\pm$ 4	63 $\pm$ 5	64 $\pm$ 4

No differences were observed in the parameters considered by comparing together the groups of sedentary subjects (C and D) and athletes (A and B) of the various age groups.

As regards the parameters of the cardiopulmonary exercise test, significant differences were observed in terms of functional capacity measured by VO2 max (p <0.01 between groups A and C, and B and D), of aerobic exercise capacity (measured by VO2AT; p = 0.02 between groups B and D and p <0.01 between groups A and C) comparing the groups of athletes and sedentary subjects of same age in both age groups. Interestingly, comparing groups of sedentary subjects according to age the functional capacity of the subjects older than 40 years is significantly reduced compared to younger subjects (p = 0.03), with no significant differences in terms of aerobic exercise capacity.

In contrast, comparing each other groups of athletes according to the age group we did not observe significant differences in terms of VO2 max and VO2AT. No difference between the groups was observed in terms of ventilatory efficiency (VE / VCO2 slope) (Table III).

Table III Cardiopulmonary parameters

	Group A	Group B	Group C	Group D
VO2max	34.1 $\pm$ 4.8	33.5 $\pm$ 6.7	26.8 $\pm$ 5.1	22.4 $\pm$ 2.9
VO2 AT	22.68 $\pm$ 5.4	23.8 $\pm$ 6.5	18.28 $\pm$ 4.3	16.7 $\pm$ 2.4
VE/VCO2	23.25 $\pm$ 5.1	24.7 $\pm$ 3.1	25.8 $\pm$ 2.8	28.1 $\pm$ 4

As regards the blood chemistry data, no difference between the groups was observed apart from an increase of HDL values among athletes and sedentary older than 40 years (65.8 + 18 mg / dl vs. 53.8 + 11.6; p <0.01).

Regarding the T-wave alternans data, we observed a trend towards a higher percentage of positive tests in athletes compared to sedentary subjects, despite the lack of statistical significance. Considering the total sedentary subjects (group C + D) and athletes (groups A + B) the percentage of positive tests were respectively 7 vs. 20% (p = 00:22). This trend was also highlighted by considering separately the sedentary subjects and athletes under age of 40 (17 vs. 0%, p = 0.27) and over 40 (12 vs. 23%, p = 0.5).

With respect to the data derived from the dynamic ECG, there is a high number of subjects both sedentary and athletes that don't present during the 24 hours recording ventricular or supraventricular ectopic beats, making the statistical comparison of mean values pointless. It is, however, interesting to note that among athletes there are 3 subjects presenting a number of ventricular ectopy exceeding 1000/24 h, while this cutoff level is not present in any sedentary subject.

A. Clustering

The first classification was achieved processing the data on the basis of the following variables:

1) athlete yes / no
2) age

3) PNN20
4) MSE1
5) MSE5
6) MSE20
7) Gender
8) Edited data for the presence of artifacts yes / no

The analysis was subsequently refined to isolate the variables able to cluster the data in a satisfactory manner.

A satisfactory result was obtained considering the variables:

1) athlete yes / no
2) age
3) MSE1

The clustering is represented in Fig. 1.

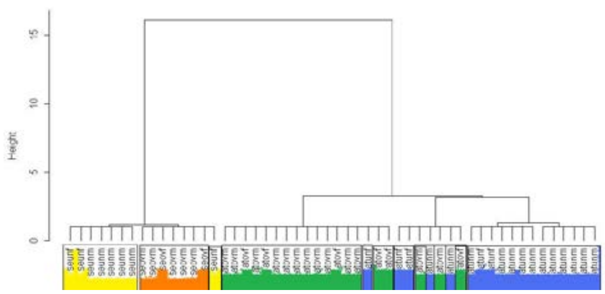

Fig. 1 Clustering using athlete yes/no, age, MSE1.
Sedovm and sedovf = Sedentary over 40
Sedunm and sedunf = Sedentary under 40
Atovm and atovf = Athletes over 40
Atunm and atunf = Athletes under 40

As seen from Fig. 1, 6 clusters are highlighted:

1) The sedentary under 40, characterized by low MSE1 (maximum = 0.182. minimum = 0.081, mean = 0.15, standard deviation = 0.034). The PNN20, as for most of the clusters, is dispersed: minimum = 0.9, maximum = 19.8, mean = 11.3, standard deviation 7.66.

2) The sedentary over 40, with highest MSE1 (maximum = 0.381, minimum = 0.122, mean = 0.222, standard deviation = 0.09). We note that also the PNN20 is greater than in the previous cluster, even if the dispersion is remarkable: PNN20 minimum = 1.7, maximum = 52.1, mean 16.3, standard deviation 16.6. In this group also appears a sedentary under 40, which has MSE1 = 0.307, sharply higher than that of the previous group.

3) Athletes over 40, with MSE1 low (maximum = 0.264 minimum = 0.001, mean = 0.127, standard deviation = 0.071). The other indicators tend to be low : PNN20 minimum = 2.6, maximum = 23.9, mean = 11, standard deviation = 6.8. An under 40 athlete is in this group with MSE1 = 0,192, that is intermediate between those of the two following clusters, in which there are other athletes under 40.

4) A cluster consists of athletes with high MSE1 (maximum = 0.568 minimum = 0.297, mean = 0.399, standard deviation = 0,11). It is the group with the highest MSE1 sample. The other indicators are high: PNN20 minimum = 9.5 ,maximum = 52.4 , men = 24.4, standard deviation 14. Three of the members are over 40 and four under 40. From the available data no other characteristics are detected.

5) A cluster of athletes under 40 with low MSE1 (maximum = 0.162, minimum = 0.069, mean = 0.115, standard deviation = 0.035). The other indicators are low, in particular in the PNN20 that this group is not dispersed as in the other (minimum = 2.3, maximum = 8.7, mean = 2.5, standard deviation = 5.97). Also note that this group has an average age of 35 years old, so it is older than the next cluster, which contains the other group of athletes under 40.

6) A cluster of young athletes very less than 40 years (mean age 19 years), with MSE1 higher than the previous cluster (maximum = 0.249, minimum = 0.095, mean = 0.137, standard deviation = 0.058) and PNN20 values: minimum = 4, maximum = 26.4, mean = 8.5, standard deviation = 11.38.

The clusters discriminate on the basis of the presence or not of a sporting activity, whereas it becomes immediately evident that sex and other variables are not discriminating, with the exception of MSE1.

The subjects aged less than 40 years have low MSE1, whether athletes (cluster 5) or non-athletes (cluster 1). Nevertheless athletes also have low PNN20 (range (2.3 - 8.7), mean = 5.97 , standard deviation = 2.53), which discriminates against them by non-athletes (range (0.9 - 19.8) , mean = 11.03, standard deviation = 7.66) according to this parameter. We note that in general PNN20 is a variable general enough dispersed in all the clusters, however, in the cluster 5 standard deviation is relatively low.

Those athletes aged very less than 40 years (cluster 6) have instead MSE1 apparently dispersed, but the mean value becomes low by eliminating one subject with MSE = 0.249 . The subject does not present other evident special features, at least in the data examined so far).

This seems to indicate that MSE1 changes after the adult age, independently from the fact that the subject is an athlete or not.

In young athletes instead MSE1 assumes variable values, indicating the hypothesis that in time the full physical maturity causes a cardiovascular modification that lowers MSE1 independently from the physical activity.

Conversely, it appears that MSE1 rises in non-athletes over the age of 40, while the athletes of the same age group have a MSE1 almost always low, as if with age physical activity tended to preserve a low MSE1: the comparison between the non-athletes over 40 (cluster 2) and athletes (cluster 3) shows MSE1 lower in athletes. Even PNN20 is smaller, however the dispersion is very high, especially in non-athletes.

Interestingly, there exists a separate group of athletes (Cluster 4) with very high MSE1 with other indicators, in turn, with high values independently from the age. The data are quite dispersed: PNN20 minimum = 9.5, maximum = 52.4, mean = 24.4 standard deviation = 14, MSE1 maximum = 0.568, minimum = 0.297, mean = 0.399, standard deviation = 0.11.

Another feature of the cluster is that 6 of the 7 components have around age 40 years (39 to 45), the other component is 26 years old and the values of PNN20 and MSE1 are the maximum of the cluster.

Eliminating this component reduces the dispersion of the variables, especially the MSE1: PNN20 range (9.5-26),mean = 19.79, standard deviation = 7.4; MSE1 range (0.3-0.53), mean = 0.09, standard deviation = 0.09.

B. Artificial Neural Network

Another classification has been carried out by means of the Artificial Neural Network described above.

Table IV ANN attractors

Athletes under 40

Attractor	N.	Av. age	Av. PNN20	Av. Mse1	Av. Mse5	Av. Mse20
C(0 – 1)	3	24	7.033	0.115	0.20	0.5
F3	4	31	25.7	0.26	0.35	0.35
F4	1	19	26.4	0.249	0.473	0.4
C(5 – 9)	**6**	**36**	**2.269**	**0.159**	**0.27**	**0.6**
F6	2	20	6.58	0.137	0.19	0.45
F8	2	40	20.42	0.295	0.486	0.89

Athletes over 40

Attractor	N.	Av. age	Av. PNN20	Av. Mse1	Av. Mse5	Av. Mse20
C(0 – 7)	8	48	12.93	0.216	0.276	0.69
C(1 – 4)	5	48	11.43	0,112	0.138	0.31
F2	2	51	14.72	0.094	0.256	0.53
F3	1	56	2.62	0.158	0.182	0.46
C(6 – 9)	3	44	8.135	0.165	0.196	0.53

Sedentary

Attractor	N.	Av. age	Av. PNN20	Av. Mse 1	Av. Mse5	Av. Mse20
C(0-others)	**7**	**40**	**19.26**	**0.219**	**0.31**	**0.5**
C(1 – 8)	**5**	**34**	**11.65**	**0.164**	**0.283**	**0.58**
C(6 – 7)	2	39	5.08	0.16	0.319	0.64
F9	1	43	4.18	0.15	0.17	0.26

The most significant results with the Artificial Neural Network system have been obtained by employing a configuration of 500 input neurons, 10 mapping units, learning rate 0.001, delta = 0.1 and are summarized in Table IV. Two kinds of chaotic attractors are found: "fixed point", consisting mainly of a single point and prefixed by F in the table, and "cyclic", consisting of at least two points and prefixed by C in the table. For instance, F3 is the attractor consisting mainly of the repetition of point 3 only, while C(0-1) is the attractor consisting mainly of the repetition of points 0 and 1.

We can also draw several objective considerations.

Firstly we note again that sex is not a discriminant variable: there is no evidence of attractors consisting mainly of males or females.

Additionally, there is no classification within the group of non-athletes related to the age, while athletes have cardiovascular characteristics differentiated depending on the age.

Moreover, it is to be noted that the comparison with the procedures of clustering shows a correspondence between an attractor of the ANN (attractor 5) and the cluster 5 (athletes younger than 40 years old, low MSE1): attractor 5 includes 5 of the 7 subjects of the cluster 5, plus a subject that is not part of the cluster.

The variables values for the attractor are the following: MSE1 range (0.08 - 0.37), mean = 0.159, standard deviation = 0.1, PNN20 range (2.27-24.23), mean = 8.2, standard deviation = 10.15.

We note that MSE1 and PNN20 have values quite close to those of cluster 5, however, the presence of the outlying subject in the attractor increases the standard deviation.

Removing this subject, which has the highest values in all the variables, we get MSE1 range (0.08-0.162), mean = 0.117, standard deviation = 0.03, PNN20 range (2.27-7.68), mean = 5.0, standard deviation = 2.33, much better aligned to the values of the cluster 5.

Fig. 2 shows a graph of the attractor C(1-8) and Fig. 3 shows a graph of the attractor C(0-1).

In addition, there is also correspondence between the attractor 0 and the cluster 2 (non-athletes with low MSE1), which have 5 elements in common of 7 and 8 elements respectively. Variables for the attractor have the following values: MSE1 range (0.08-0.38), mean = 0.22, standard deviation = 0.1; PNN20 range (1.7-52.1), average = 19.26, standard deviation = 16.45.

For the cluster 2 we have: MSE1 range (0.122-0.381), mean = 0.22, standard deviation = 0.09; PNN20 range (1.7 - 52.1), average = 16.3, standard deviation = 16.6.

Finally, many attractors of non-athletes are composed mainly from elements belonging to the cluster of non-athletes under the age of 40 years. For example, four of the five subjects of the attractor 1-8 (Fig. 3) are part of the cluster 1.

The results of the Artificial Neural Network (ANN) procedure show that athletes and non-athletes are clearly discriminated on the basis of the already mentioned variables, as the tables show that these groups have no common attractors.

The above considerations show that the adopted data analysis procedures are mutually congruent in discriminating the studied subjects, confirming the validity of the obtained results.

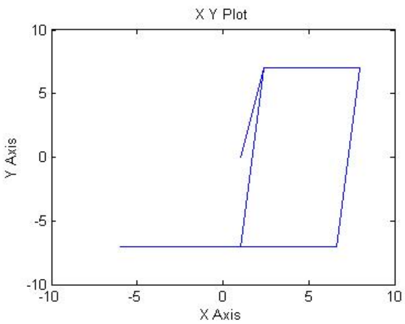

Fig. 2 A graph of the attractor C(1-8)

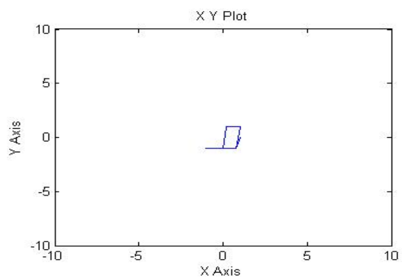

Fig. 3 A graph of the attractor C(0-1)

To summarize this first analysis, the above considerations show that both the existence of sensitive variables (athlete / no, age, MSE1) and of significant groups of patients related to the variability of these parameters.

The subjects aged less than 40 years have low MSE1 , whether athletes or non-athletes. However, athletes have PNN20 low.

- This would suggest that MSE1 modifies itself above the adult age independently from the physical activity.
- In young athletes, instead, MSE1 assumes variable values, indicating the hypothesis that over time the full physical maturity results in a cardiovascular modification that lowers MSE1 regardless of physical activity.
- Conversely, it appears that MSE1 rises in non-athletes over the age of 40, while the athletes of the same age group have MSE1 almost always low, as if with age, physical activity tended to preserve a low MSE1.
- The sex is never a discriminating variable.
- It is also worth observing that the ANN classification stresses the lack of age classification within groups of non-athletes, while athletes result to have cardiovascular characteristics differentiated by age.

C. Adding other variables

Starting from this preliminary analysis, we were interested in evaluating whether and to what extent, within the identified clusters, other variables may indicate subgroups of individuals with characteristics that could represent common cardiovascular prognostic indicators.

Among the previously mentioned collected parameters we considered just those available for all patients:
TDD (telediastolic diameter), IVS (interventricular septum) and PW (posterior wall) thickness, ETT (exercise tolerance test), EF (ejection fraction), maximum O2 consumption (VO2peak),% VO2, O2 consumption at the anaerobic threshold (VO2AT), VE/VCO2 slope (indicator of ventilatory response to exercise), peak RER (respiration exchange ratio), maximum workload, HB (hemoglobin), HCT (hematocrit), creatinine, cholesterol (total, HDL, LDL), triglycerides, blood glucose, BNP (brain natriuretic peptide), BMI (body mass index), DTS (Duhe treadmill score), mean Holter HR (cardiac frequency), min-max-mean FC, FC%, besides of the previously considered: athlete (yes / no), age, PNN20, MSE1, MSE5, MSE20.

D. Comparison among clusters

To examine the significance of the differences between the variable values moving from one cluster to another, we used a Student's t significance test.

The following table shows the significant differences between the six clusters for the main variables. We reported just the variables with significant value (Table V).

We can draw the following observations:
- None of the considered variables is significantly different between the two clusters of sedentary subjects (clusters 1 and 2)
-some variables (VO2peak, workload, FCmean, FCmin) differ significantly between the two clusters of sedentary subjects and various clusters of athletes
-some variables (TDD, FCmin, DTS) differ significantly between the sixth cluster (very young athletes) and many of the other
-some variables (FCmin, FCmean, VO2peak,% VO2, VO2AT) have significant differences between the cluster 2 (sedentary over 40) and 3 (athletes over 40) .

E. Clustering with the new variables

As a following step, we performed a new clustering procedure using this time all the available variables.

We found that the best result identifies 5 new clusters:

A first cluster of 16 patients, 6 of which are over 40 athletes, 10 athletes under 40 years (mean age 33.9, 15 males, 1 female)

A second cluster of 12 patients, 6 of which are athletes over 40, 6 athletes under 40 years (mean age 40.4, 10 female, 1 male)

A third cluster of 12 patients, 8 of athletes over 40, 4 under 40 athletes (mean age 46, 12 males)

A fourth cluster of 10 patients, 6 of sedentary subjects over 40, 4 sedentary under 40 (mean age 44.1, 9 males, 1 female)

A fifth cluster of 11 patients, 5 of which are sedentary subjects over 40, 6 sedentary under 40 (mean age 38.5, 5 males, 6 females).

Table V
Variable differences among clusters.
Significances at the 0.1% are indicated with **

Var	Significantly different clusters	Var	Significantly different clusters
TDD	Cl1-Cl6 0,001234 Cl2-Cl6 0,003948	HB	Cl3-Cl6 0,009329
VO2 AT	Cl2-Cl3 0,000143**	Tot Col	Cl2- Cl6 0,005123 Cl3-Cl6 0,00952
VO2 peak	Cl1-Cl3 0,003012 Cl1-Cl6 0,003548 Cl2-Cl6 7,34E-05** Cl2-Cl3 2,3E-06** Cl2-Cl4 0,001829 Cl2-Cl5 0,004963	FC mean	Cl2-Cl3 0,00574 Cl2-Cl5 0,002248
% VO2	Cl1-Cl3 0,000618** Cl2-Cl3 0,003959 Cl3-Cl6 0,007783	FC min	Cl2-Cl6 0,001545 Cl2-Cl3 0,006866 Cl2-Cl5 0,000548** Cl2-Cl6 0,000377** Cl3-Cl6 0,00504 Cl4-Cl6 0,007178
Work load	Cl1-Cl3 0,000365** Cl1-Cl5 0,000595** Cl1-Cl6 0,001346 Cl2-Cl5 0,004411 Cl2-Cl6 0,003899	DTS	Cl3-Cl6 0,002852 Cl5-Cl6 0,000909**

A positive aspect of this clustering is that it reflects the separation between athletes and sedentary subjects.

Another interesting aspect is that clusters reflect a clear division by sex, although this variable was not used in the analysis.

The following charts show the evolution of the variables on the basis of the clusters.

In principle, they seem to reflect the division of patients into clusters. In in the clusters of athletes (1-2-3) we note (Fig.4) the dependence on the age (rising from the cluster 1-3 and on sex (female athletes are concentrated in cluster 2).

For example, the workload is maximum for cluster 1 (male athletes), minimum for cluster 2 (female athletes), intermediate for cluster 3 (males athletes older on average).

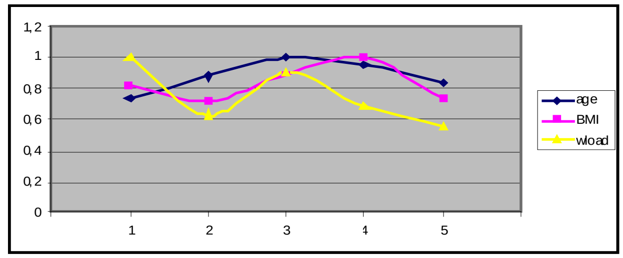

Fig.4 Dependence of clusters on age, BMI, workload

The graph regarding TDD, IVS, PW, DTS shows a good correlation between IVS and PW in clusters 1, 2 and 5; between IVS and TDD clusters 3 and 4; between PW and DTS in cluster 3. In other cases these variables are more distant (Fig.5).

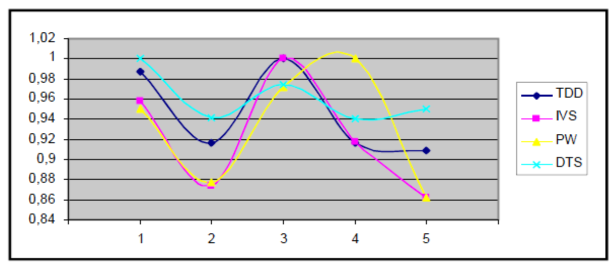

Fig. 5 Dependence of clusters onTDD,IVS, PW,DTS

The graph VO2AT, VO2peak, %VO2 shows a good correlation between the variables for clusters 1 and 2; VO2AT and VO2peak between clusters 4 and 5. In other cases, these variables are more distant (Fig.6.)

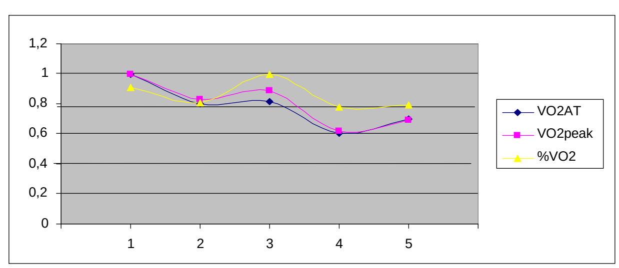

Fig. 6 Dependence of clusters on VO2AT,VO2peak,%VO2

The graph HB, HCT shows a good correlation between the variables for clusters 1, 4 and 5. In other cases, these variables are more distant (Fig.7).

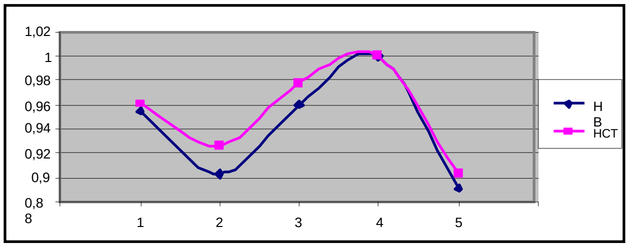

Fig. 7 Dependence of clusters on HB, HCT

The graph FCmean, FCmin shows a good correlation for all clusters (Fig.8).

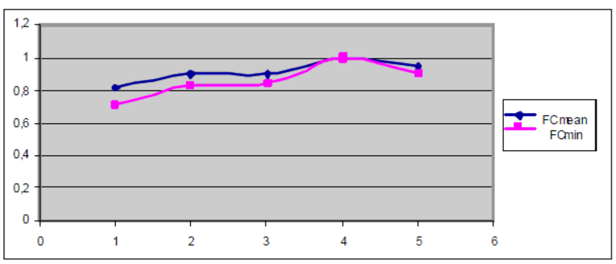

Fig. 8 Dependence of clusters on FCmean and FCmin

The graph PNN20, MSE1, MSE5, MSE20 shows a good correlation between MSE1 and MSE5 in all clusters; among all the variables for clusters 2 and 4. In other cases these variables are more distant (Fig.9).

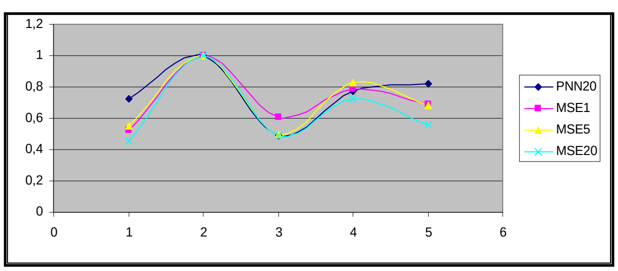

Fig. 9 Dependence of clusters on PNN20, MSE1, MSE5, MSE20

To examine the significance of the differences between values of each variable moving from one cluster to another, we used a paired Student's t.

The table shows, for each pair of clusters, the variables whose mean differ significantly (p <0.01) (Table VI).

Table VI
Variable differences among clusters
Results with p < 0.001 are indicated with *

	c2	c 3	c4	c5
c1	**VO2 peak, work load*, FCmin, MSE20**	**age, VO2AT, tot col*, HDL*, triglyc., FCmean, FCmin**	**VO2AT*, VO2peak*, %VO2, workload*, tot col, HDL, triglyc., FCmean*, FCmin***	**TDD, VO2AT *, VO2peak, workload*, Fcmean *, Fcmin***
c2		**TDD, IVS, workload*, tot col* , HDL*, PNN20, MSE5**	**VO2peak, FCmin**	**VO2 peak**
c3			**VO2AT*, VO2peak*, workload, triglyc., FCmin**	**TDD, IVS, VO2peak*, workload*, HCT, tot col*, HDL***
c4				**PW, HB, HCT, triglyc.**

We observe that:
- Cluster 1 (young male athletes) is significantly different from all other clusters for respiratory variables (VO) and at least one FC variable. It also shows significant differences in workload for all clusters except for cluster 3 (male athletes over 40).
- The differences between cluster 2 (female athletes), and two clusters of sedentary subjects are limited to a few variables (a respiratory variable and FC)

The differences between clusters of athletes and sedentary subjects involve frequently cholesterol or variables connected to it.

IV. Conclusions

The above described study involved athletes and sedentary subjects of different ages and had as its objective the stratification of the subjects on the basis of a number of cardiovascular and clinical variables in order to search for possible sensitive parameters for evaluating the risk of severe tachyarrhythmia events.

Our study has highlighted some interesting aspects about the adaptation of the cardiovascular exercise and the peculiarities of the adaptation, respectively, in subjects aged less than 40 years and in the master athletes.

As a first step, the study identified six clusters grouping the subjects according to the variation of some parameters, especially MSE1 and PNN20, found analysing the signals of an exercise test ECG.

The clusters are identified in this way:

CL1: non-athletes <40 and low MSE1
CL2: non-athletes > 40 and high MSE1 and high PNN20
CL3: athletes > 40 and low MSE1 and low PNN20
CL4: athletes and max MSE1 and low PNN20
CL5: athletes <40 and low MSE1 and low PNN20
CL6: athletes << 40

The classification shows that the parameter MSE1 is low in athletes both < and > 40 and sedentary <40, and higher in sedentary> 40. It can be concluded that sport seems to keep down the parameter MSE1 regardless of age.

The application of a self-organizing ANN does not contradict the classification made by the clustering, but reveals the following additional facts: the sedentary subjects are not separated by age, but result to be clearly separated by athletes.

Sex is not a discriminating variable in either the clustering or the ANN classification.

In the second part of the study, to the first parameters derived from the exercise test ECG a number of parameters derived from other diagnostic tests have been considered.

Adding the other variables in the study, these have been used to classify the subject within the clusters previously formed. In particular, it is possible to discriminate:

-depending on the VO variables
athletes > 40 from non-athletes > 40
non-athletes <40 from athletes > 40
non-athletes > 40 from athletes <40
-depending on the workload
non athletes <40 from all athletes
depending on DTS
athletes <40 from athletes << 40
-depending on FC
non-athletes> from 40 athletes <40

No variables are significantly different between non-athletes <40 and > 40.

Then we run a second clustering procedure, which divides sharply athletes and sedentary males and females without distinguishing them by age:

c1 athletes l <> 40 male (except1)
c2 athletes <> 40 female (except 1)
c3 athletes <> 40 male
c4 non-athletes <> 40 male (except 1)
c5 non-athletes <> 40 male and female

The following groups can be discriminated on the basis of the variables:

-workload: between male athletes and female athletes
-VO parameters, workload, FC parameters : between male athletes and non-athletes, both male and female
-VO parameters: between male athletes and female athletes.

Furthermore male athletes differ compared to all the other groups according to different variables, in particular VO variables and workload.

Finally, male athletes are distinguished from non-athletes according to the FC parameters.

We can conclude that the study has allowed a stratification of the subjects on the basis of physical activity, age and sex, and showed the existence of significant differences in the cardiovascular status of these groups through the variability of a set of cardiovascular parameters, in particular MSE1, PNN20, VO and FC variables.

The complex interaction between these variables required the use of non-linear analysis techniques, namely clustering and ANN, which made it possible to perform a set of classifications.

Furthermore, our data showed that, as amply demonstrated in the mentioned literature, the athletes have a tele-diastolic left ventricular diameter higher than sedentary subjects. In our study, due to the peculiar sport practiced, we did not observe a concomitant increase in wall thickness.

Actually in endurance sports the heart overload is primarily a volume overload, and as such it determines a form of eccentric hypertrophy, characterized mainly by an increase in cavitary volume. It is interesting to note, however, that this form of adaptation has not been observed in subjects with age greater than 40 years, and this opens interesting perspectives in terms of pathophysiology. In fact in these subjects remodeling due athletic activity could be offset by a form of concentric remodeling that is often observed in individuals of the same age and is due to several factors, including for example hypertension.

On the other hand, this different remodeling could be due to a reduction in the intensity of training, reducing the intensity of the stimulus to eccentric ventricular remodeling. In this regard, further studies with larger number of cases could help to provide further clarification with respect to such evidence. Another important aspect to consider is that the sport activity is able to determine an increase of the work capacity both overall and aerobic, through a better muscle conditioning with a consequent increase of the capacity of peripheral oxygen extraction, and a more efficient metabolism at a myocytes level, but also through an improvement of cardiac performance, as shown in the cited literature. Our study showed that while in sedentary individuals with age above 40 years, the exercise capacity tends to decrease, with reduction of VO2 max, in athletes master there are significant differences in this parameter compared to younger athletes. The fact that this difference between athletes and sedentary is observed only in terms of VO2 max and VO2 AT seems to support the hypothesis that this difference is due to"peripheral" factors, then linked to muscle conditioning and metabolism of myocytes.

As for parameters related to cardiovascular risk such as blood pressure and cholesterol, in our study there were no significant differences between the groups with regard to these parameters, except for HDL cholesterol between athletes and sedentary subjects above 40 years.

This finding has not been confirmed in subjects aged less than 40 years, probably because of the already higher values of HDL and of the level of physical activity in sedentary individuals already higher than that of the subjects over 40 years. There were no difference in systolic and diastolic blood pressure and LDL cholesterol.

In this sense, however, we must consider that our study included subjects with normal lipid and normotensive, and consequently any changes in such parameters determined by exercise are certainly minor compared to those observed in patients with hypertension or dyslipidemia, and therefore they are not able to determine a statistically significant difference in a small group of subjects. Our study also highlights interesting aspects also regarding arrhythmic risk stratification.

In the group of athletes we observed a higher percentage of patients with ventricular extrasystoles than in non-athletes. This fact, in our study, is associated with the finding of a higher percentage of subjects with positive T-wave alternans test among athletes compared with sedentary subjects. This figure, which does not reach statistical significance probably due to the small size of the sample, however could suggest the presence of an electrical remodeling at the level of cardiomyocytes induced by endurance sports, that is compatible with the above cited mechanical remodeling. This electrical remodeling might be one of the factors behind the increased tendency of athletes to develop hyperkinetic ventricular arrhythmias, regardless of the presence of structural heart disease.

These data may also support the hypothesis formulated in the literature that intensive exercise can determine by itself a not completely physiological and potentially arrhythmogenic cardiac remodeling.

In conclusion, our study provided interesting information about the peculiar adaptation of the cardiovascular exercise in different age groups, a subject about which the data in the literature are very scarce. It also highlighted a higher presence in athletes of parameters indicative of a possible increased risk of arrhythmia, as the positivity of T-wave alternans, finding that opens interesting perspectives from the point of view of the pathophysiology of ventricular arrhythmias and sudden death in athletes.

Using non-linear analysis on the data of the exercise test also allowed the identification of a significant breakdown in clusters among the various groups of subjects according to the degree of physical activity, confirming its significant role in determining the functional and cardiovascular parameters.

Taking into account the achieved stratification it will be possible, following the subjects over time, to identify cardiovascular prognostic indicators that may help to prevent possibly fatal cardiac arrhythmias.

These results can be better confirmed on samples of greater size, and carrying out ad hoc analyses in order to clarify their clinical and pathophysiological significance.

REFERENCES

[1] H. Bjornstad, L. Storstein, “Ambulatory electrocardiographic findings in top athletes, athletic students and control subject,” *Cardiology*, vol. 84, no. 1, pp. 42-50, 1994.

[2] H. Heidbüchel, J. Hoogsteen, R. Fagard R, “High prevalence of right ventricular involvement in endurance athletes with ventricular arrhythmias. Role of an electrophysiologic study in risk stratification,” *Eur Heart J*, vol. 24, no. 16, pp. 1473-1480, 2003.

[3] G. Inama, C. Pedrinazzi, O. Durin, M. Nanetti, R. Pizzi, “Microvolt T-wave alternans for risk stratification in athletes with ventricular arrhythmias: correlation with programmed ventricular stimulation,” *Annals of Noninvasive Electrocardiology*, vol. 13, no. 1, pp. 14-21, 2008.

[4] P. Palatini, C. Miraglino, “Prevalence and possible mechanisms of ventricular arrhythmias in athletes,” *Am Heart J.* , vol. 110, no. 3, pp. 560-567, 1985.

[5] F. Furlanello, R. Bettini, A. Bertoldi, G. Inama, “Arrhythmia patterns in athletes with arrhythmogenic right ventricular dysplasia,” *Eur Heart J.* , vol.10(D), pp. 6-19, 1989.

[6] A. Biffi, A. Pelliccia, “Long-term clinical significance of frequent and complex ventricular tachyarrhytmias in trained athletes,” *J. Am. Coll.Cardiol*, vol. 40, no. 3, pp. 446-452, 2002.

[7] H. L. Kennedy, “Ventricular ectopy in athletes. Do not worry ..... More good news,” *J. Am. Coll. Cardiol*, vol. 40, no. 3, pp. 453-456, 2002.

[8] D. Corrado, C. Basso, “Does sports activity enhance the risk of sudden death in adolescent and young adults? ,” *J. Am. Coll. Cardiol.* , vol. 42, no. 11, pp. 1959-1963, 2003.

[9] A. Biffi, L. Verdile, “Arrhythmias and relationship to physical training in athletes,” *Ital Heart J. Supp.* , vol. 5, no. 1, pp. 143S - 144S, 2004.

[10] G. Inama, C. Pedrinazzi, O. Durin, M. Nanetti, G. Donato, R. Pizzi, “Ventricular arrythmias in competitive athletes: risk stratification with T-wave alternans”, Heart International, vol. 3, no. 1, pp. 58-67, 2007.

[11] A. Pelliccia, R. Fagard, H. H. Bjørnstad, “Study Group of Sports Cardiology of the Working Group of Cardiac Rehabilitation and Exercise Physiology; Working Group of Myocardial and Pericardial Diseases of the European Society of Cardiology, Recommendations for competitive sports participation in athletes with cardiovascular disease: a consensus document from the Study Group of Sports Cardiology of the Working Group of Cardiac Rehabilitation and Exercise Physiology and the Working Group of Myocardial and Pericardial Diseases of the European Society of Cardiology,” *Eur Heart J.* , vol. 26, no.14, pp. 1422-1445, 2005.

[12] D. P. Zipes, A. J. Camm, M. Borggrefe, “ACC / AHA / ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death,” *J Am Coll Cardiol*, vol. 48, no. 5, pp. 1064-1108, 2006.

[13] Organizing Committee for the Fitness for Sports Cardiology ANCE-ANMCO-FMSI-SIC-SICSPORT, *Protocols for cardiological assessment of their suitability to competitive sport 2009*. Rome: International Scientific Publishing House, 2009.

[14] F. Giada, G. Inama, M. Bertaglia, “Management of patients with palpitations”, Consensus document of the AIAC (Italian Association of Arrhythmiology and Cardiostimulation), *G Ital Cardiol* , vol. 11, no. 4, pp. 329-340, 2010.

[15] The Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, “Heart Rate Variability Standards of Measurement, Physiological Interpretation, and Clinical Use,” *Circulation*, vol. 93, pp. 1043-1065, 1996. Available: http://circ.ahajournals.org/content/93/5/1043.full

[16] J. E. Mietus, C.K. Peng, I. Henry, R. L. Goldsmith, A. L. Goldberger, “The PNNx files: re-examining a Widely used heart rate variability measure,” *Heart*, vol. 88, no.4, pp. 378-380, 2002.

[17] M. Costa, A. L. Goldberger, C. Peng, “Multiscale entropy analysis of physiologic time series,” *Phys Rev Lett*, vol. 89, no. 62, pp. 1-4, 2002.

[18] Rui Xu and Wunsch D., ”Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, pp. 645-678, 2005.

[19] Rohlf, F.J. (1982),”Single link clustering algorithms", in *Handbook of Statistics*, vol. 2, P.R. Krishnaiah and L.N. Kanal, Eds. Amsterdam: North-Holland, 1982, pp. 267-284.

[20] Murtagh, F. (1983), “A Survey of Recent Advances in Hierarchical Clustering Algorithms,” *Computer Journal*, vol. 26, pp. 354–359, 1983.

[21] R. Pizzi, T. Rutigliano, P. Guadalupi , M. Pregnolato, "Molecular Cooperation to Reinforce Immune Response during Carcinoma: a Structural Bioinformatics Analysis," in *8th Int Conf on Applied Mathematics, Simulation and Modeling*, Florence , 2014, pp. 346-353

[22] R. Pizzi, T. Rutigliano, A. Ferrarotti, M. Pregnolato, "Computational Prediction of Binding Affinity between Psychotropic Drugs and Neural Cytoskeleton Elements," in *8th Int Conf on Applied Mathematics, Simulation and Modeling*, Florence, 2014, pp. 132-138.

[23] R. Pizzi, S. Siccardi , C. Pedrinazzi, O. Durin and G. Inama, "Cardiovascular Modifications and Stratification of the Arrhythmic Risk in Young and Master Athletes," *American Journal of Biomedical Engineering*, vol.4, no.3, 2014.

[24] R. Pizzi, G. Inama, O. Durin, C. Pedrinazzi, "Non-linear analysis of ECG signals as a method for the prognosis of letal tachyarrhythmias," in *IV Int. Conf. on Information and Communication Technologies*, Cairo, 2006.

[25] R. Pizzi. S. Siccardi, C. Pedrinazzi, O. Durin, G. Inama, "Data Mining methods for the stratification of the arrhythmic risk in young and master athletes," in *3rd Int Conf on Circuits Systems Communications Computers and Applications*, Florence, 2014, pp. 235-244.

[26] V. Batagelj, *Generalized Ward and related clustering problems, in Classification and Related Methods of Data Analysis*, North-Holland, 1988, pp. 67-74.

[27] T. Hastie, R. Tibshirani, J. Friedman, *Hierarchical clustering in : The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition, New York: Springer, 2009, pp. 520–528.,

[28] A. D. Gordon, *Classification*, Second Edition, London: Chapman and Hall / CRC, 1999.

[29] J. H. Ward, Jr. Hierarchical grouping to optimize an objective function, J. Am Stat. Ass, 58(301), 236–244. 1963,

[30] T. Kohonen, 1983, *Self-Organisation and Association Memory*, 3rd edition, New York:Springer-Verlag.

[31] T. Kohonen, "Physiological Interpretation of the Self-Organizing Map Algorithm," *J. Neural Net*., vol. 6, no. 7, pp. 895-905, 1993.

[32] R. Pizzi, M. de Curtis, C. Dickson, "The Binding Problem: evidence of chaotic attractors in cortical fast oscillations tested by an Artificial Neural Network, in *Advances in Soft Computing*, Physica Springer-Verlag, 2003.

[33] O. Durin, C. Pedrinazzi, G. Donato, R. Pizzi, G. Inama, "Usefulness of nonlinear analysis of ECG signals for prediction of inducibility of sustained ventricular tachycardia by programmed ventricular stimulation in patients with complex spontaneous ventricular arrhythmias," *Annals of Noninvasive Electrocardiology*, vol. 13, no. 3, pp. 219-227, 2008.

[34] R. Pizzi, G. Inama, O. Durin, C. Pedrinazzi, "Non-invasive assessment of risk for severe tachyarrhythmias by means of non-linear analysis techniques," *Chaos and Complexity Letters*, vol. 3, no. 3, pp. 229-250, 2007.

[35] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. Ch. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215-e220, 2000.

[36] R Development Core Team. (2005) *R: A language and environment for statistical computing*. Available: http://www.R-project.org .