Early prediction of the highest workload in incremental cardiopulmonary tests

Early Prediction of the Highest Workload in Incremental Cardiopulmonary Tests

ELENA BARALIS[1], TANIA CERQUITELLI[1], SILVIA CHIUSANO[1], VINCENZO D'ELIA[1], RICCARDO MOLINARI[2] and DAVIDE SUSTA[3]

[1]Dipartimento di Automatica e Informatica, Politecnico di Torino, Torino, Italy;
[2]Sport Training Center, Eupilio, [3]Dublin City University

Keywords Incremental test, highest workload prediction, classification techniques, multivariate data, physiological signals analysis.

Abstract *Incremental tests are widely used in cardiopulmonary exercise testing both in the clinical domain and in sport sciences. The highest workload (denoted W_{peak}) reached in the test is a key information to assess the individual body response to the test and to analyze possible cardiac failures, plan rehabilitation and training sessions. Being physically very demanding, incremental tests can significantly increase the body stress on the monitored individuals, and may cause cardiopulmonary overload. This paper presents a new approach to cardiopulmonary testing that addresses these drawbacks. During the test, our approach analyzes the individual body response to the exercise and predicts the W_{peak} value that will be reached in the test and an evaluation of its accuracy. When the accuracy of the prediction becomes satisfactory, the test can be prematurely stopped, thus avoiding its entire execution. To predict W_{peak} we introduced a new index, the CardioPulmonary Efficiency Index (CPE), summarizing the cardiopulmonary response of the individual to the test. Our approach analyzes the CPE trend during the test, together with the characteristics of the individual, and predicts W_{peak}. A k-Nearest Neighbor based classifier and an ANN based classier are exploited for the prediction. The experimental evaluation showed that the W_{peak} value can be predicted with a limited error since the first steps of the test.*

1 Introduction

Cardiopulmonary exercise testing is a non-invasive and objective method of evaluating both the cardiac and pulmonary functions. Incremental tests are commonly used to progressively increase the mechanical demand that the individual cardiopulmonary system has to match until she/he can no longer maintain the current workload. The cardiopulmonary response to exercise, when skeletal muscles transform chemical energy into mechanical output, has been shown to depend on workload, and the individual performance to be largely dependent on her/his aerobic power (i.e., the ability in supplying oxygen to, and in removing carbon dioxide from, working muscles).

To analyze the individual body response to increasing strain, various physiological signals, mainly describing the respiratory function of the individual, are monitored during the incremental test. The *highest workload* (denoted as W_{peak}) achieved in the last stage of the test corresponds to the "best" cardiopulmonary adaptation for the monitored individual in terms of heart rate and ventilation, and it is an important indicator of her/his aerobic power.

Incremental tests are commonly used in the clinical domain Sill et al (2009); Spruit and Wouters (2007) and in sport sciences Pollock et al (1998), and both treadmills and cycloergometers are usually exploited for testing. In the clinical domain, incremental tests support assessing heart functions in patients with chronic heart failure Sill et al (2009), or evaluating pulmonary function in patients with chronic obstructive pulmonary disease Spruit and Wouters (2007). In sport sciences, exercise physiologists use incremental tests for endurance sports testing Pollock et al (1998). Both in the clinical domain and in sport sciences, the W_{peak} reached by the individual in the incremental test is crucial to manage rehabilitation and training sessions, by planning and adjusting the assigned workload Mezzani (2009). For example, in sport sciences, according to the American College Sport Medicine, trainers should set training intensities as fractions of the W_{peak} values reached by the athlete in previous tests Pollock et al (1998). In the clinical domain, exercise capacity has been shown to be the best predictor of survival among healthy men Myers et al (2002). Unfortunately, incremental cardiopulmonary tests, the gold standard to assess exercise capacity, are physically very demanding. Long test durations can significantly increase body stress and may cause cardiopulmonary overload on the monitored individuals. Hence, early prediction of the expected W_{peak} value during the test execution would allow to prematurely interrupt the test, thus lowering body stress. In addition, cardiopulmonary testing is time consuming and reducing the time needed to perfom one test could allow cardiopulmonary testing to become a routine procedure as suggested by experts Franklin and McCullough (2009). This issue is further discussed in Section 7.

Nomograms and other approaches predicting maximal exercise capacity based on factual data have been proposed since the origins of exercise physiology Astrand and Ryhming (1954) to overcome a series of limitations including the cost of equipment for direct measurement of power output and energy consumption. However, these approaches suffer from several limitations, e.g., they do not allow measuring a subject's improvement in performance because they cannot detect changes in individual performance in a time as short as a training programme (usually weeks).

This paper presents an innovative approach to predict the W_{peak} value that can be reached during an incremental test. By coupling factual data with the physiological signals monitored during test execution, our approach predicts W_{peak} according to the actual response of the individual *during the test*. Hence, it allows detecting changes in individual performance and adaptations to training. Both in the clinical and sport domains this information allows effectively monitoring individual progression during a rehabilitation/training programme. W_{peak} is dinamically predicted during test execution, whenever the assigned workload is increased, together with an estimate of its accuracy. When the estimated accuracy of the prediction becomes satisfactory, the test may be stopped, thus avoiding its entire execution. Hence, the proposed approach does not replace testing, but instead (i) it provides an additional tool, while testing, to reduce physical effort and (ii) it allows saving test time, thus potentially making cardiopulmonary testing more affordable without missing crucial information. To our knowledge, this work is the first study addressing dynamic early prediction of W_{peak}.

To characterize the performance of the tested individual and predict W_{peak}, we introduce a new measure, named *CardioPulmonary Efficiency Index* (CPE), which summarizes both the cardiac and pulmonary responses of the individual to the incremental test. The information provided by the CPE index during the test is coupled with some anthropomorphic data of the individual and exploited to train a classifier for the prediction of W_{peak}. We considered two different classification techniques: An instance-based learning approach based on the k-Nearest Neighbors (k-NN) algorithm Tan et al (2005), and Artificial Neural Networks (ANN) Tan et al (2005). The proposed approach has been evaluated on two sets of incremental tests collected at CSA, Sport Training Center (Italy) Sport Training Center (2011), for diverse athletes. These tests use two protocols commoly adopted in endurance sport testing. Experimental results showed that W_{peak} is predicted with a limited error since the first instants of the test.

The paper is organized as follows. Section 2 describes the test protocol and the physiological signals monitored during the test, while Section 3 introduces by means of an applicative example the motivations driving our approach. Section 4 presents the CPE index and Section 5 describes the k-NN and ANN classifiers to predict

Table 1: Monitored physiological signals

Signal name	**Abbreviation**	**Measurement unit**
Carbon dioxide production	VCO_2	ml/min
Fraction of expired air that is carbon dioxide	$FetCO_2$	%
Fraction of expired air that is oxygen	$FetO_2$	%
Heart rate	HR	bpm
Pulmonary ventilation	VE	l/min
Tidal volume	Vt	l
Oxygen consumption	VO_2	ml/min
Ventilatory equivalent ratio for carbon dioxide	VE/VCO_2	–
Ventilatory equivalent ratio for oxygen	VE/VO_2	–

W_{peak}. Section 6 reports the experiments that evaluate the prediction accuracy, whose results are discussed in Section 7. Related works are described in Section 8 and Section 9 presents future developments of the approach.

2 Test protocol and data collection

This section describes the test protocol and the physiological signals collected during the test to characterize the test execution.

Test protocol. In incremental tests, the workload is a step signal defined by two parameters: The increment of workload at each step (W_{step}) and the duration of each step (t_{step}) in which the workload is kept constant. These two parameters define the test protocol, denoted $W_{step} \times t_{step}$, meaning that every t_{step} minutes the workload is increased by W_{step} Watt.

The protocol is set before the test starts and is kept constant during the test. The test ends when the individual cannot sustain the current workload. The workload at the test end is the highest workload achieved by the individual in the test, i.e., W_{peak}.

Data collection. Each test is characterized by two kinds of information, *factual* and *dynamic* data. Factual data describe the individual performing the test. We considered individual age, Body Mass Index (BMI), and Body Surface Area (BSA).[1] These attributes are normalized by means of the z-score method, which is less sensitive to the skewed data distribution Tan et al (2005) characterizing factual data.

Dynamic data include the various physiological signals sampled during the test to analyze the individual body response under increasing strain. The individual is monitored by means of a set of sensors and a spirometer. Besides the cardiovascular parameters (e.g., the heart rate), the majority of signals describes the ventilatory function of the individual. Table 1 reports the subset of signals used in our framework. Since they differ both in scale and measurement unit, a *min-max* normalization has been performed. This normalization technique is typically exploited in time series Kasetty et al (2008), because it preserves the original data distribution.

3 Motivating example

We show by means of a real application example how the proposed W_{peak} prediction approach may support incremental test monitoring and early interruption. We selected two different individuals (test ids 130 and 279[2]) performing the incremental test with protocol 50 $W \times 2$ min. The factual data of both subjects, reported in Table 2, are characterized by identical BMI and BSA, and rather different age. Hence, traditional W_{peak} prediction approaches based on factual data (e.g., nomograms) would typically predict a higher W_{peak} for test id 130, characterized by lower age.

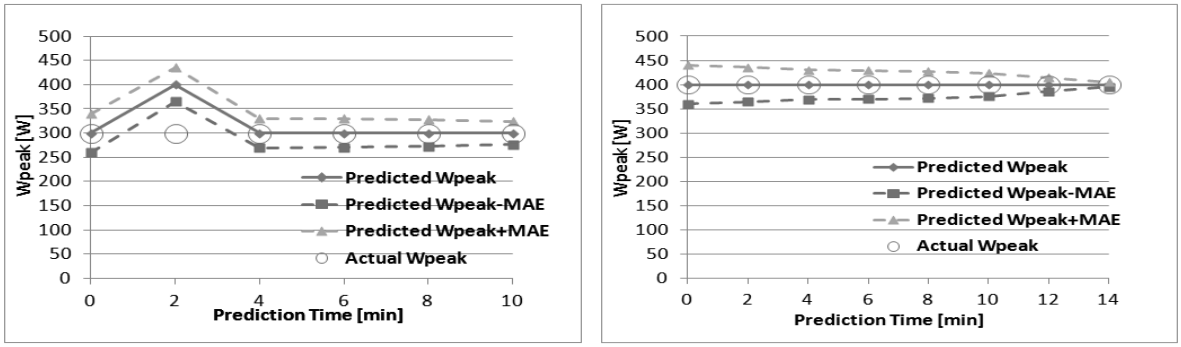
In Figure 1 the solid line plots the W_{peak} prediction performed by our approach at each workload increment during the test until its end, while the two dashed lines show the estimation of the average prediction error (Mean Absolute Error, MAE). After the initial steps, our approach correctly predicts the W_{peak} value for both individuals. In addition to factual data, our prediction is based on the analysis of the *actual individual response*

[1] The individual gender is omitted in this work, because all tests used to evaluate the approach were performed by male individuals.

[2] More information on the experimental datasets is provided in Section 6.1

Table 2: Individual characteristics

Test ID	Age	BMI	BSA	W_{peak}
130	18	21.45	1.72	300
279	27	21.45	1.72	400



(a) Test ID 130

(b) Test ID 279

Figure 1: W_{peak} prediction

elicited during the test (i.e., the dynamic data). Hence, it may unveil more subtle differences between individual responses to exercise, that can only be discovered by actually performing the test. Since the prediction error monotonically decreases and the prediction becomes rather stable after the initial steps of the test, it is not necessary to reach the natural end of the test (requiring at least 10 min). The tests can be prematurely stopped at 200 W (after 6 min), when the estimated MAE decreases below 30 W.

4 CardioPulmonary Efficiency Index (CPE)

The CardioPulmonary Efficiency Index (CPE) is a new measure that dynamically summarizes the overall cardiopulmonary response elicited by a given workload during an incremental test. Hence, it can be computed at time t during the test to evaluate the overall response up to time t. It is computed by taking into account the following two components: (i) The *exercising individual power output* (denoted ε_L), which is computed from the nominal workload kept constant by the cycloergometer, and (ii) the overall *cardiopulmonary response to exercise* for the individual (denoted ε_T), which is evaluated based on the various physiological signals monitored during the test.

The CPE value is computed from terms (i) and (ii) as

$$CPE = \frac{\varepsilon_L}{\varepsilon_T} \tag{1}$$

where both terms ε_T and ε_L are positive real values.

When cycling, human skeletal muscles always need an amount of metabolic energy higher than the produced mechanical output. Hence, the cardiopulmonary response of the individual (ε_T) is larger than the forced power output (ε_L), and the CPE index varies in the range $[0, 1)$.

The samples of every signal contributing to ε_L and ε_T are collected synchronously at each sampling time t_i. Since ventilatory signals are collected "breath by breath", the sampling rate is dependent on the respiratory rate. Hence, every cumulative measure computed by using the collected samples is normalized by dividing its value by the number of individual breaths since the beginning of the test.

4.1 Exercising individual power output (ε_L)

The nominal workload is represented as a discrete signal $Load[t]$, where $Load[t_i]$ is the workload assigned during the test at the sampling time t_i. The power output the individual is asked to produce is computed as the energy of signal $Load[t]$, by definition of energy of a discrete signal Moon and Stirling (2000), as

$$\varepsilon_L = \sum_{i=1}^{m} Load[t_i]^2 \tag{2}$$

where m is the number of samples collected for signal $Load[t]$ during the test.

4.2 Cardiopulmonary response to exercise (ε_T)

The cardiopulmonary response to exercise ε_T is computed as the overall energy of all monitored physiological signals. This value is computed by means of the Singular Value Decomposition (SVD) method Moon and Stirling (2000), which supports signal filtering to reduce the impact of signal noise and allows considering the existing correlations among signals. SVD has been effectively used in physiological signal analysis Hassanpour et al (2004); Aysin et al (2005), as well as in other domains (e.g., image Bao and Ma (2005) and microarray data analysis Wall et al (2001)). It is suited to signals characterized by sudden changes Aysin et al (2005), typically occurring after a workload increase.

Each physiological signal is represented as a discrete signal $v[t]$, being $v[t_i]$ the sample at time t_i. Signal samples are stored in a matrix $T \in R^{m,n}$, where each column contains the samples of a different signal. n is the number of signals represented in T and m is the number of samples for each signal.[3] Matrix T is factorized by means of SVD as

$$T = UHV^T \tag{3}$$

where $U^TU = I_m$ and $V^TV = I_n$. Matrices $U \in R^{m,m}$ and $V \in R^{n,n}$ represent the orthonormal basis for the column and row spaces of T, respectively. I_m and I_n are identity matrices of orders m and n. H=diag$[s_1, s_2, ..., s_n]$ includes the n singular values of T, with $s_1 \geq s_2 \geq \cdots \geq s_n > 0$ Moon and Stirling (2000).

The overall energy of the signals in matrix T can be computed from the singular values of T as Moon and Stirling (2000)

$$\varepsilon_T = \sum_{i=1}^{n} s_i^2. \tag{4}$$

To reduce the effect of signal noise on ε_T, we computed a low rank approximation of matrix T by means of a truncated SVD Moon and Stirling (2000). The smaller singular values of T are disregarded. Hence, signals in T are filtered by projecting them on the most relevant components of matrix T, i.e., a subset of the columns in matrix U. In particular, we considered only the first two largest singular values in T, i.e., $\varepsilon_T = s_1^2 + s_2^2$. The experimental evaluation reported in 6.2 showed that the resulting ε_T represents accurately the cardiopulmonary response to the exercise and supports the prediction of W_{peak} with a limited error.

4.3 Evaluation of the CPE value during the test

During the test, new samples of the monitored signals are periodically added to matrix T and the CPE value is recomputed on the updated matrix. This sequence of CPE values summarizes in a single time series the diverse signals monitored in the test. We exploited this sequence, denoted as $CPE[t]$, to characterize the test execution.

Sequence $CPE[t]$ shows a non-decreasing trend during the test. Both terms ε_L and ε_T are by their definition cumulative measures, being computed by considering the signal contributions collected from the beginning of the test up to the current time. The individual power output ε_L is characterized by a sequence of sudden increases in correspondence to the workload increments. The cardiopulmonary response to exercise ε_T is characterized by a steady growth.

To analyze the individual response to the test, we computed the variation of the CPE value between consecutive sampling instants, given by

$$CPE_\Delta[t_i] = CPE[t_i] - CPE[t_{i-1}] \tag{5}$$

$CPE_\Delta[t]$ shows a sequence of peaks, taking place when the workload is increased. We have experimentally observed that this sequence of peaks is representative of the individual response to the test. Hence, this sequence, denoted as $CPE_{peaks}[t]$, is used in this work as a fingerprint of the individual performance. As a representative example, Figure 2 plots the workload and the $CPE[t]$, $CPE_\Delta[t]$, and $CPE_{peaks}[t]$ sequences for a test with protocol 50 $W \times 2$ min and W_{peak} 450 W.

In the $CPE_{peaks}[t]$ sequence, peaks are higher in the very first steps of the test. Then, their value rapidly decreases. The sequence length is equal to the test duration and the sequence converges to a value, which is (potentially) different among tests. The trend of the $CPE_{peaks}[t]$ sequence can be explained by means of the concept of ventilatory reserve, i.e., the difference between the maximal voluntary ventilation and the peak minute ventilation achieved in the test. The ventilatory reserve is high in the early instants and decreases until the conclusion of the test. Between two contiguous peaks, the $CPE_\Delta[t]$ sequence is characterized by a steady decrease. This trend can be interpreted as the progressive body adaptation to the current workload and the convergence to a steady state.

[3] Without loss of generality, we suppose that $m > n$, i.e., the number of samples is larger than the number of monitored signals.

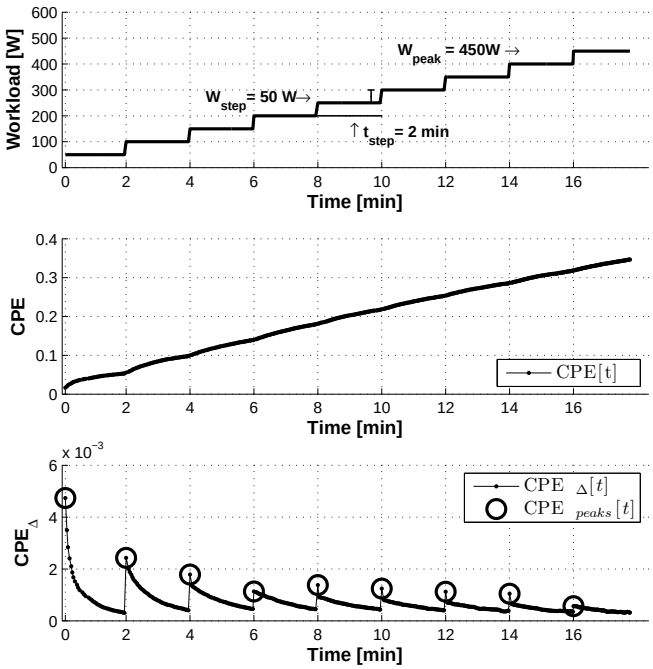


Figure 2: Workload and $CPE[t]$, $CPE_{\Delta}[t]$, and $CPE_{peaks}[t]$ sequences for an example test with protocol $W_{step} = 50\ W \times t_{step} = 2$ min and W_{peak}=450 W

5 Prediction of the highest workload in the test

The highest workload (W_{peak}) reachable in an ongoing test is periodically predicted during the test, each time the assigned workload is increased. The prediction of the W_{peak} value is based on the following information: (a) *Characteristics of the individual* doing the test, corresponding to the factual data for the individual, and (b) *cardiopulmonary response* of the individual to the test, given by the sequence of peak values $CPE_{peaks}[t]$ until prediction time, which summarizes the dynamic data of the test.

Given the predicted value of W_{peak}, a further step, which allows augmenting the available information on the individual performing the incremental test is the evaluation of her/his cardiopulmonary reserve. In particular, the information on W_{peak} may be completed by computing the potential further workload, which we denote as $W_{potential}$. At a given time t, $W_{potential}$ may be computed from W_{peak} as

$$W_{potential}[t] = W_{peak}[t] - W_{step}[t] \tag{6}$$

where $W_{peak}[t]$ corresponds to the value of W_{peak} predicted at time t and $W_{step}[t]$ to the workload reached at time t.

In this paper, we exploited two classification techniques to predict W_{peak}: (i) A classifier based on *Artificial Neural Networks* (ANN) Tan et al (2005), and (ii) an instance-based learning approach, based on the *k-Nearest Neighbors* (k-NN) classification algorithm Tan et al (2005). The prediction of the W_{peak} value based on the two classification techniques is described in the following subsections, while Section 6.4 discusses the advantages and disadvantages of both approaches.

5.1 Prediction of the W_{peak} value by means of an ANN classifier

Artificial neural networks (ANN) simulate biological neural systems. An ANN is composed by a collection of nodes (artificial neurons) and directed links connecting them. For W_{peak} prediction, we build a classifier based on a multilayer feed-forward neural network trained with backpropagation. The network consists of an input layer, n hidden layers, and an output layer. Each layer is made up of nodes. Each node in a layer takes as input a weighted sum of the outputs of all the nodes in the previous layer. It applies a nonlinear activation function to the weighted input. Currently, the sigmoid function has been used as activation function.

The prediction of W_{peak} for a new (ongoing) test Q takes place each time t in which the workload is increased. It is performed by an ANN classifier trained with a set of previous incremental tests (both factual and dynamic data), run with the same protocol of test Q, and characterized by W_{peak} values at least equal to the workload of test Q at time t.

The nodes in the input layer correspond to the factual data of the individual and the $CPE_{peaks}[t]$ sequence for the test. The nodes in the output layer represent the discrete (i.e., categorical) W_{peak} values that can be

reached in the test. The network is trained with backpropagation and learns by iteratively processing the set of training tests. For each training test, the network predicts the W_{peak} value whenever the workload is increased, and compares the predicted and the actual W_{peak}. Then, weights in the network nodes are modified to minimize the mean squared prediction error. These modifications are made in the "backwards" direction, that is, from the output layer, through each hidden layer down to the first hidden layer.

5.2 Prediction of the W_{peak} value by means of a k-NN classifier

k-NN classification algorithms predict the expected value for a new data object by using a set of examples for which the target value is known, without explicitly building a classification model from them Tan et al (2005). These algorithms exploit the *whole* training set for prediction, by searching for the k training examples that are closest to the new data object.

To predict the W_{peak} value for a new test Q, each time the workload is increased, the k-NN classifier analyzes a collection of previous tests and selects the k tests most similar to Q. Similarity evaluation considers (a) similar individual characteristics and (b) cardiopulmonary responses close to that observed in test Q. The classifier predicts the W_{peak} value for test Q based on the W_{peak} values in the k selected tests. More specifically, the prediction process entails the following steps.

Creation of the reference knowledge base for test Q. The reference knowledge base is a set of previous tests, run with the same protocol used for test Q and characterized by W_{peak} values at least equal to test Q workload at prediction time. Tests shorter than the current length of test Q are disregarded, because they are useless for the prediction.

Selection of the k-nearest tests to test Q. The similarity between test Q and each test in the knowledge base is evaluated and the k tests most similar to Q are selected. Similarity is computed by considering both the characteristics of the individuals performing tests (i.e., the factual data) and their response to exercise (i.e., the dynamic data). The approach for similarity evaluation is described in Section 5.2.1.

Prediction of W_{peak} for test Q. The W_{peak} value for test Q ($\widehat{W}_{peak,Q}$) is computed as the average W_{peak} on the selected k-nearest tests as

$$\widehat{W}_{peak,Q} = \frac{1}{k} \cdot \sum_{i=1}^{k} W_{peak,i} \tag{7}$$

Since Equation 7 provides a real value, $\widehat{W}_{peak,Q}$ is approximated to the closest workload given by the protocol. This value is the highest workload expected to be achieved by the individual in the test. For example, if the 3-nearest tests are characterized by W_{peak} equal to 300 W, 350 W, and 350 W, the average value is 333.3 W. If the protocol is 50 W×2 min, then $\widehat{W}_{peak,Q}$ is approximated to 350 W. The real value of $\widehat{W}_{peak,Q}$ provides an additional information that may support in estimating the distance of the individual from the closest workload.

5.2.1 Similarity evaluation between two tests

The similarity between a new test Q and a test θ in the knowledge base is measured by considering the similarity of their factual and dynamic data.

Similarity between factual data. It evaluates the similarity between individuals performing tests Q and θ. It is computed as the Euclidean distance Tan et al (2005) between the factual data describing them.

$$\begin{aligned} d_{factual}(Q,\theta) &= [\,(Age_Q - Age_\theta)^2 \\ &+ (BMI_Q - BMI_\theta)^2 \\ &+ (BSA_Q - BSA_\theta)^2\,]^{\frac{1}{2}} \end{aligned} \tag{8}$$

Similarity between dynamic data. It represents the similarity between the body responses Q and θ to the test strain. Dynamic data, i.e., the physiological signals, are sampled during the test and summarized in the $CPE_{peaks}[t]$ sequence. We evaluated the similarity between the dynamic data of tests Q and θ as the similarity between their $CPE_{peaks}[t]$ sequences.

The $CPE_{peaks}[t]$ sequence may have a different length in the two tests, because tests in the knowledge base have duration longer than, or equal to, the current duration of test Q. For test θ, the $CPE_{peaks,\theta}[t]$ sequence is computed with the samples gathered over the entire test. For test Q, still running, the $CPE_{peaks,Q}[t]$ sequence is given by the samples collected up to the current prediction time.

The similarity between sequences $CPE_{peaks,\theta}[t]$ and $CPE_{peaks,Q}[t]$ is evaluated using the Dynamic Time Warping (DTW) technique Berndt and Clifford (1994), a nonlinear time-alignment method for sequence comparison. DTW has been shown to be effective for univariate time series classification in different domains,

including physiological signal analysis Ratanamahatana and Keogh (2004). DTW supports the comparison of sequences with different lengths by looking for their best alignment. The *warping_band* is the maximum tolerated difference between sequences. Thus, this technique allows us to compare individual body responses that are similar, but locally out of phase.

The two sequences are compared by means of DTW within a time-window having approximately the duration of test Q. When test Q currently takes n minutes, the first $n \pm$ *warping_band* minutes of sequence $CPE_{peaks,\theta}[t]$ are considered. We expect that, if the two sequences show a similar trend within the time-window, they might show an analogous trend also in the next steps. Thus, test Q may have the same duration as test θ, as well as its same W_{peak}. The DTW algorithm finds the set of correspondences with the least total distance between the matched points between peaks in sequence $CPE_{peaks,Q}[t]$ and in subsequence $CPE_{peaks,\theta}[t]$ within the time-window. This distance measures the similarity between the two sequences, denoted as $d_{dynamic}(Q,\theta)$. Between two matched points, as usually done in the DTW technique Berndt and Clifford (1994), the Euclidean distance is computed.

Similarity between two tests. The similarity between tests Q and θ is computed by considering the similarity between both their factual ($d_{factual}(Q,\theta)$) and dynamic ($d_{dynamic}(Q,\theta)$) data as

$$d(Q,\theta) = w * d_{factual}(Q,\theta) + (1-w) * d_{dynamic}(Q,\theta) \tag{9}$$

where $d(Q,\theta) \in [0,1]$ since both $d_{factual}$ and $d_{dynamic}$ have been normalized in $[0,1]$, and $w \in [0,1]$. Lower values of $d(Q,\theta)$ denote a higher similarity between the two tests, and higher values of $d(Q,\theta)$ a lower similarity.

When increasing w, the k-nearest tests are mainly selected based on the characteristics of the individuals performing the tests. As a consequence, tests with body responses close to test Q, but done by individuals dissimilar from the individual of test Q, may be disregarded. On the other hand, when decreasing w, the test similarity is mainly evaluated by comparing the individual body responses to exercise. Increasing the relevance of this comparison could be misleading in the first steps of the test, when sequence $CPE_{peaks,Q}[t]$ still contains few elements. An experimental evaluation (see Baralis et al (2011b)) showed that the predicted W_{peak} value well approximates the actual one when the two terms in Equation 9 have the same weight ($w = 0.5$).

6 Experimental evaluation

The experiments performed to evaluate our approach address the following issues: (i) Prediction error, (ii) impact of test duration on the prediction error, and (iii) performance comparison with two naive classifiers. The experimental results are reported in this section, while a further discussion is reported in Section 7.

To perform the experiments, we collected two sets of incremental tests at CSA Sport Training Center (Italy) Sport Training Center (2011), by using two common protocols in endurance sport testing. The datasets are described in Section 6.1 and are available on request to the authors. The experiments have been run on a 2 GHz Intel Centrino Dual-Core PC, with 1 GB of RAM and running Linux kernel 2.6.27. The proposed approach has been implemented in C and Python programming languages.

The framework configuration depends on different parameters, set as follows for the reported experiments. (a) The first two largest singular values in matrix T are considered to compute ε_T. (b) The ANN classifier exploits the neural net operator available in the RapidMiner tool RapidMiner (Last access on December 2011). To configure this operator, we set training cycles 500, learning rate 0.3, momentum 0.2, and hidden layer 1. (c) For the k-NN classifier, the following parameters have been set. (i) To compare two $CPE_{peaks}[t]$ sequences using DTW, the *warping_band* is set to 2 min (i.e., one step for both test protocols). (ii) In evaluating test similarity, the contributions of factual and dynamic distance have the same weight ($w = 0.5$). (iii) The 5-nearest tests are selected for W_{peak} prediction ($k = 5$). The effect of varying the selected parameter values is discussed in Baralis et al (2011b) and, specifically for the ANN parameters, in Baralis et al (2011a).

6.1 Datasets

To gather representative examples, the tests have been done with two protocols commonly used in endurance sport testing to elicit the highest workload in both elite and intermediate athletes. The protocols are 50 $W \times 2$ min and 25 $W \times 2$ min, and the tests performed with these protocols have been collected in two datasets named $D_{50\times2}$ and $D_{25\times2}$, respectively. Table 3 reports the main characteristics of both datasets, and Figure 3 shows the distribution of their W_{peak} values.[4] The datasets include male athletes, both amateur and elite.

Protocol 50 $W \times 2$ min is more stressful for the athlete body, because of the higher increment of workload applied at each step. For this reason, athletes in $D_{50\times2}$ are typically younger and more trained than in $D_{25\times2}$, and they usually reach higher W_{peak} values. Protocol $D_{25\times2}$ is applicable to rather diverse athletes, due to the

[4]The distribution of the Age, BMI, and BSA values is reported in Baralis et al (2011b).

Table 3: Characteristics of the two datasets. For all parameters, the mean and the standard deviation (SD) values are reported

			Parameters			
Dataset	**No. of Tests**	**No. of Athletes**	**Name**	**Mean ± SD**	**Min value**	**Max value**
$D_{50\times2}$	231	202	Age	32.3 ± 12.2 years	14	61
			BMI	22.8 ± 2.3 Kg/cm^2	17.65	33.26
			BSA	1.8 ± 0.1 m^2	1.44	2.13
			W_{peak}	333 ± 68 W	150	500
$D_{25\times2}$	184	139	Age	38.7 ± 12.5 years	13	66
			BMI	23.9 ± 2.6 Kg/cm^2	17.48	31.8
			BSA	1.8 ± 0.1 m^2	1.44	2.38
			W_{peak}	262 ± 62 W	175	400

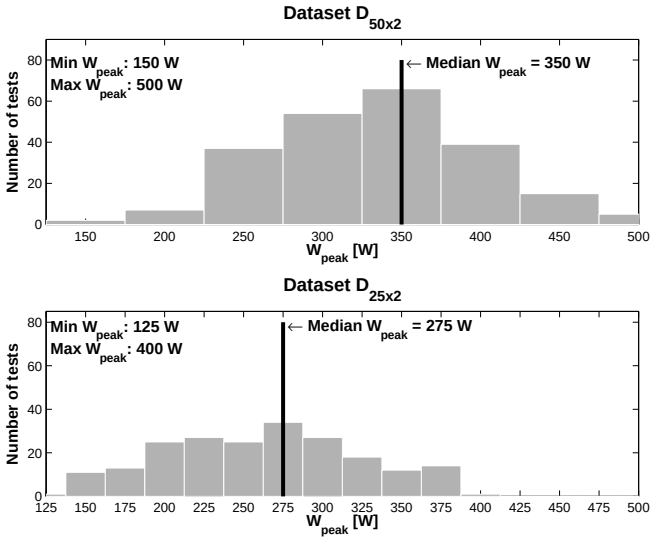


Figure 3: Distribution of the highest workload in the datasets

lower increment of the applied workload. Athletes in $D_{25\times2}$ are more heterogeneous with respect to age, BMI, and BSA values (especially when considering the joint distribution of these values).

6.2 Prediction accuracy for the W_{peak} value

The prediction error measures the ability to correctly predict the W_{peak} value for a new test. The *prediction error* is the difference between the predicted and the actual W_{peak} value for the test. The *absolute prediction error* is the absolute value of this difference. During the test, W_{peak} is periodically predicted each time an increment of workload occurs, and the corresponding error is evaluated.

The leave-one-out cross-validation method Tan et al (2005) is used for error evaluation. At each workload increment, i.e., at each prediction time t_p, the subset of tests still running is selected from the dataset. In turn, a different test is picked out of this subset, while the remaining tests are used as knowledge base to predict its final W_{peak}. To perform the prediction, the chosen test is described by its $CPE_{peaks}[t]$ sequence until time t_p, and the other tests by their $CPE_{peaks}[t]$ sequence up to the test end. The *Mean Absolute Error* (MAE) Tan et al (2005) at time t_p is the average of the absolute prediction errors computed for all tests in the subset.

In all charts, the origin of the time axis (time = 0 min) corresponds to the first step of the test, in which workload 50 W is assigned in dataset $D_{50\times2}$ and 25 W in dataset $D_{25\times2}$. In the corresponding datasets, the average test lengths (average W_{peak} values) are 13.32 min (333 W) and 20.96 min (262 W), respectively. The experimental results for the ANN and k-NN classifiers are denoted as *ANN-based* and *k-NN-based*, respectively.

Figures 4(a) and 5(a) report the mean absolute error (MAE) by varying the prediction time. In dataset $D_{50\times2}$ (Figure 4(a)), the MAE value is below 40 W, except for the first step, in which it is sligthly above 40 W for the k-NN classifier. The ANN classifier always provides lower values of prediction error. In dataset $D_{25\times2}$ (Figure 5(a)), the MAE value is always below 40 W for the k-NN classifier. For the ANN-classifier, it is below 50 W at the first step, almost 30 W after 6 min, and then it rapidly decreases after 10 min. These errors correspond to over(under)estimating the test duration of less than 1 step (2 min) when the protocol is 50 W $\times$ 2 min and less than 1.6 steps (3.2 min) when the protocol is 25 W $\times$ 2 min. For dataset $D_{50\times2}$, we observe that the prediction error of both classifiers is always lower than the actual increment of workload at each step.

In both datasets, the MAE value decreases when postponing the prediction time and progressively tends to zero. The error is higher in the early steps, when the running test is described by a short $CPE_{peaks}[t]$ sequence and the knowledge base used for prediction contains the majority of the dataset. Because of these two conditions, the prediction of W_{peak} is initially affected by a larger error. Later, instead, the prediction becomes more accurate, because the running test is more precisely represented by a longer $CPE_{peaks}[t]$ sequence.

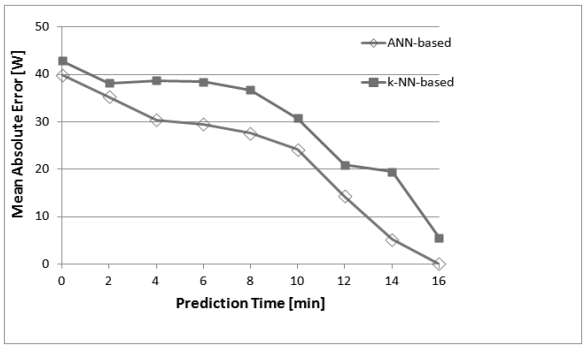
The MAE curves are more irregular and not strictly decreasing in dataset $D_{25\times2}$ (Figure 5(a)), especially in the first steps. This trend is due to the lower increment of workload applied in protocol 25 W $\times$ 2 min, which has less capability to differentiate the athlete body responses to exercise, especially in the first steps of the test. Consequently, tests reaching different W_{peak} values may be initially characterized by similar $CPE_{peaks}[t]$ sequences. Furthermore, athletes in $D_{25\times2}$ are more heterogeneous, being the protocol applicable to rather diverse athletes.

6.3 Impact of the test duration on the prediction error

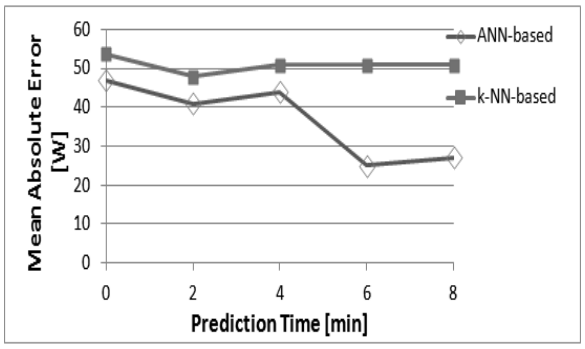
To analyze the effect of the test length on the prediction error, we considered separately tests reaching the same W_{peak}. Figures 4(b)-4(f) and 5(b)-5(f) plot the corresponding prediction errors for datasets $D_{50\times2}$ and $D_{25\times2}$, respectively.

In both datasets, and for both classifiers, longer tests are in general affected by a higher error in the very first instants. The knowledge base used for prediction initially contains the entire dataset, while a limited subset of tests reaches high W_{peak} values. Few long tests are usually available, because they are only sustainable by well trained athletes. In addition, in the first steps the running test is described by a short $CPE_{peaks}[t]$ sequence, while the final sequence is significantly longer. The prediction error is usually reduced after few steps, as soon as the increments of workload sufficiently characterize the $CPE_{peaks}[t]$ sequence describing the running test, thus significantly improving the prediction performance.

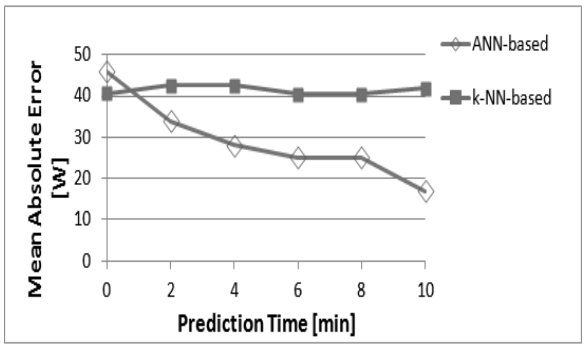
For dataset $D_{50\times2}$ (see Figures 4(b)-4(f)), the experiments are only reported for W_{peak} values characterizing at least 10 tests, because for a smaller number of tests the prediction error on a single test may significantly affect the average error. In general, the ANN classifier achieves a more accurate prediction than the K-NN classifier. For both classifiers, the MAE value is always lower than 50 W, except for tests with W_{peak} 450 W. For these longer tests, the MAE value at the first step is close to 80 W for the k-NN classifier and to 60 W for the ANN classifier. The prediction error falls below 50 W at the next steps. The initial error is caused by the small fraction of tests achieving W_{peak} 450 W (only 7.36%, while for the other W_{peak} values the test percentage is between 16.01% and 29.87%).



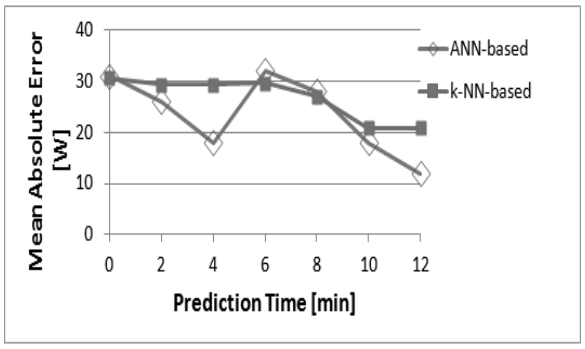
(a) MAE for all tests



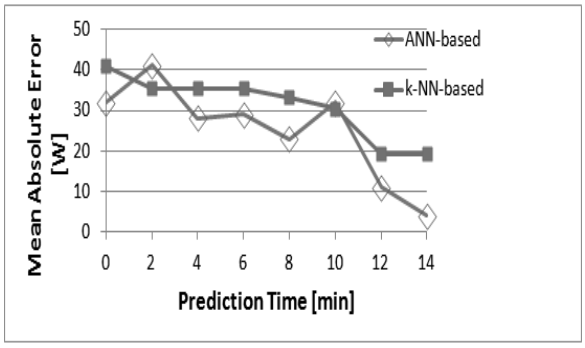
(b) MAE for W_{peak} = 250W (37 tests)



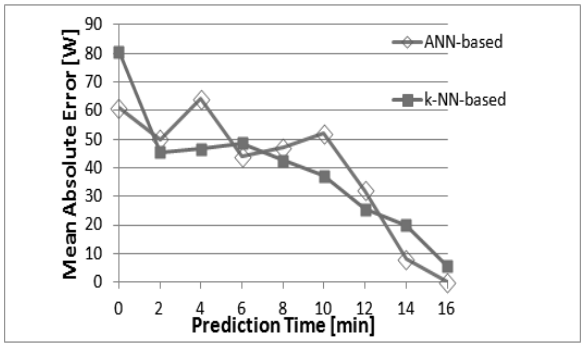
(c) MAE for W_{peak} = 300W (54 tests)



(d) MAE for W_{peak} = 350W (69 tests)

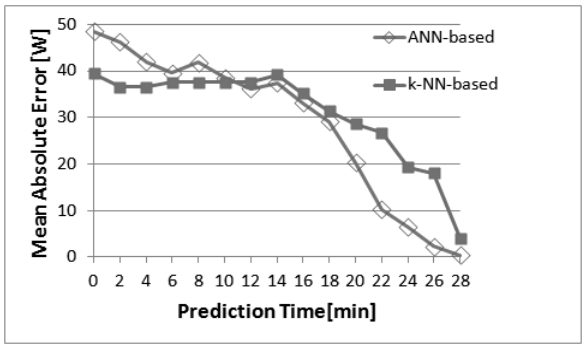


(e) MAE for W_{peak} = 400W (42 tests)

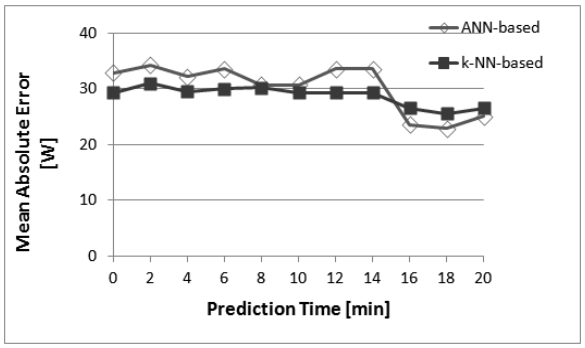


(f) MAE for W_{peak} = 450W (17 tests)

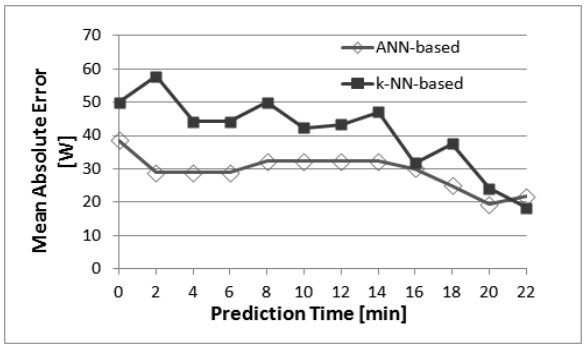
Figure 4: MAE for tests in dataset $D_{50\times2}$



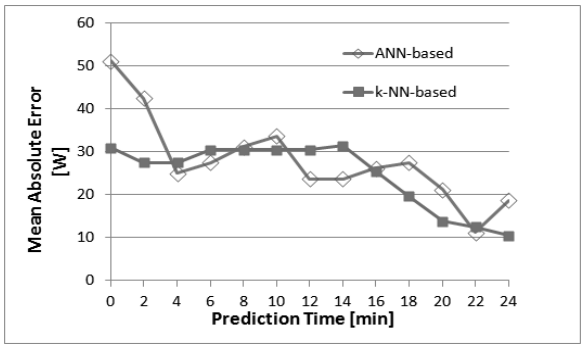
(a) MAE for all tests



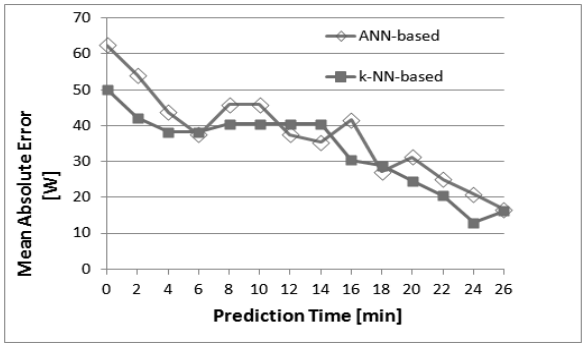
(b) MAE for W_{peak} = 275W (36 tests)



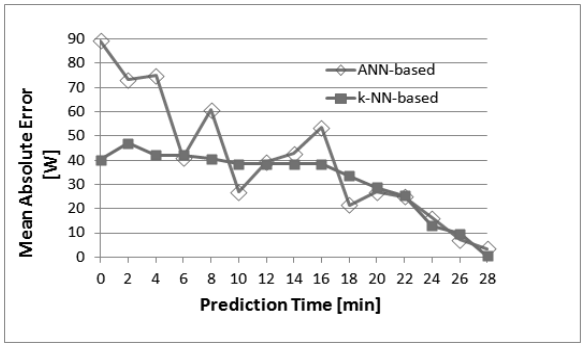
(c) MAE for W_{peak} = 300W (27 tests)



(d) MAE for W_{peak} = 325W (18 tests)



(e) MAE for W_{peak} = 350W (12 tests)



(f) MAE for W_{peak} = 375W (14 tests)

Figure 5: MAE for tests in dataset $D_{25\times2}$

For dataset $D_{25\times2}$ (see Figures 5(b)-5(f)), to analyze the error on longer tests, only results for tests with $W_{peak} \geq 275$ W are reported. Results for W_{peak} 400 W are omitted, because only a single test achieves this value. The MAE curves are more irregular for the ANN classifier, and not strictly decreasing. The ANN classifier is more accurate on tests with W_{peak} 300 W, while the k-NN on tests with W_{peak} 375 W. For tests with W_{peak} 300 W, 325 W, and 350 W the two classifiers achieve similar performance. As discussed in Section 6.2, protocol 25 W$\times$ 2 min has less capability to differentiate the athlete body response to exercise, and athletes in this dataset are more heterogenous. In addition, a lower number of tests is available for each W_{peak} value. Under these conditions, building an accurate "global" classification model, which summarizes the information available in the training set and provides the correct classification of new tests, is more difficult. The k-NN classifier, instead, extracts from the entire training set a "local" subset of instances which are exploited for prediction. Hence, in these cases, its prediction may be both more accurate and more stable.

6.4 Classifier comparison

As shown in the previous sections, both the ANN and k-NN classifiers provide an effective W_{peak} prediction. Since the two approaches are characterized by different features, each one may be more appropriate in a different working condition. In the following, we compare the features of the two approaches in the context of the maximum workload prediction.

Classification accuracy. ANN classification algorithms are very effective in a wide range of applications and typically show high tolerance to noisy data. When the training set contains a sufficient number of tests, they provide a very accurate W_{peak} predition, in general more accurate than the k-NN approach. When a low number of training tests is available (e.g., lower than 20 tests), building an accurate "global" classification model is more difficult. The k-NN classifier, instead, extracts from the training set a small subset of instances close to the test instance, which are exploited to provide a more accurate and stable "local" prediction. For example, in Figure 5(f), the ANN classifier shows lower accuracy and a more irregular MAE curve because only 14 tests were available for training.

Interpretability. Both in the clinical and sport domains, doctors and trainers may be interested in analyzing the reasons motivating a given prediction. While neural network models are intrinsically characterized by very poor interpretability, the k-NN classifier may provide some insight by pointing at the subset of k individuals, whose factual data and cardiopulmonary response to the exercise are more similar to the individual performing the test.

Model evolution. When new training objects are available, the ANN model cannot be incrementally updated and should be completely rebuilt. Since the k-NN algorithm analyzes the whole training set to perform each prediction, incremental update by adding new examples to the training set is straightforward.

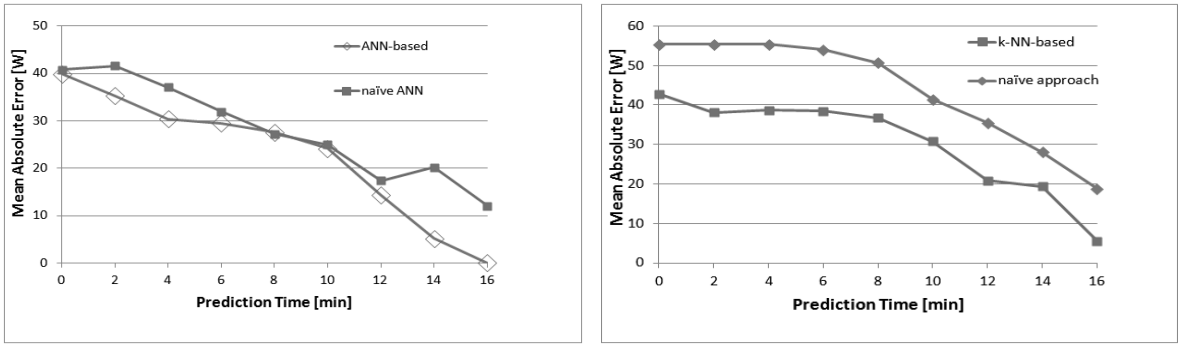
Training and classification time. Neural networks require a long training time (i.e., several hours) to build the classification model. Furthermore, devising the appropriate parameter setting may be complex and time-consuming Baralis et al (2011a). Once the model is defined, classification is very efficient (i.e., few seconds). k-NN does not build a classification model at training time. Hence, classification typically requires a longer time (roughly 15-20 seconds), because the distance between each training instance and the test instance has to be computed at run time. k-NN classification time grows with the size of the training set.

6.5 Comparison with two naive classifiers

To our knowledge, an approach for the early prediction of the W_{peak} value in incremental tests has not been proposed yet. Thus, we compared our approach against two naive classifiers. Dataset $D_{50\times2}$ is used as a reference example for comparison, being the dataset including more tests. The leave-one-out cross-validation method has been exploited for prediction error evaluation.

The naive ANN classifier is a multilayer feed-forward neural network trained with backpropagation, that receives as input data (i) the factual data describing the athlete and (ii) all the physiological signals sampled during the test (instead of the $CPE_{peaks}[t]$ sequence summarizing them). The experimental results in Figure 6(a) show that the ANN naive classifier is less accurate than the ANN classifier based on the $CPE_{peaks}[t]$ sequence. Hence, the $CPE_{peaks}[t]$ sequence provides an effective summarization of athlete performance, which can be profitably used to predict the W_{peak} value.

The second naive classifier that we exploited works as follows. Consider a new test Q, where the workload currently assigned in the test is W_Q. The naive classifier selects all tests with $W_{peak} \geq W_Q$ from the dataset with the same protocol as test Q. Then, it computes the average W_{peak} on the selected tests and predicts this value as the W_{peak} for test Q. Experimental results in Figure 6(b) show that the k-NN classifier always performs better than the naive approach, by providing lower MAE values. Thus, the selection of the k-nearest



(a) Comparison against a naive ANN classifier

(b) Comparison against a naive k-NN classifier

Figure 6: Comparison against two naive classifiers in dataset $D_{50\times2}$

tests, based on the analysis of factual and dynamic data describing the test, properly support the prediction of the W_{peak} value.

7 Discussion

The experimental evaluation has been run on two protocols commonly adopted in endurance sport testing. For both protocols, tests with different lengths can all be prematurely stopped even in their early steps, with a limited prediction error on the final W_{peak}value. For example, a 200 W workload is usually tolerated by endurance athletes with average performance, while the athlete body stress may significantly increase for higher workloads. When the workload is 200 W (i.e., after 6 min for protocol 50 W × 2 min and after 14 min for protocol 25 W × 2 min), the mean absolute error (MAE) on the predicted W_{peak} value is at the worst case below 40 W for both protocols 50 W × 2 min and 25 W × 2 min. In dataset $D_{50\times2}$, this error does not affect the evaluation of the athlete performance, because it is within the volitional exhaustion range, i.e., the increment of workload assigned when the athlete decides to interrupt the test, and corresponding to 50 W. Also the prediction error achieved in dataset $D_{25\times2}$ is acceptable, because of the characteristics of the protocol. Protocol 25 W × 2 min requires a more gradual adaptation of the athlete body to the physical strain, due to the lower increment of the applied workload. As a consequence, a prediction error below 40 W does not significantly affect the evaluation of the overall athlete response to the test. For a discussion about the impact of the adopted protocol on the subjects evaluation, see Revill et al (2002).

In dataset $D_{50\times2}$, tests representing the average performance (W_{peak} 333 W) end between 10 min (W_{peak} 300 W) and 12 min (W_{peak} 350 W). By interrupting these tests when workload is 200 W (i.e., after 6 min), the test length is shortened by 40%-50% with a significant reduction of the athlete body stress. For longer tests, for example with W_{peak} 450 W, the reduction of the test length is around 62%. In dataset $D_{25\times2}$, the average W_{peak} is 262 W, corresponding to a test duration between 18 min (W_{peak} 250 W) and 20 min (W_{peak} 275 W). By ending these tests when the workload is 200 W (i.e., after 10 min), the test is shortened by 22%-30%. The reduction for longer tests, for example with W_{peak} 375 W, is about 50%. In both datasets, the prediction error progressively decreases for workloads higher than 200 W. Trainers can thus decide when to stop the test, based on the trade-off between test duration and closer approximation of the final workload.

8 Previous related work

Incremental tests to measure the maximal workload (exercise capacity) provide a fundamental information to assess an individual's fitness, predict life expectancy and better prescribe workout intensities to stimulate beneficial adaptations. Hence, the importance of this procedure has been recognised by the sport and medical communities and cardiopulmonary tests are the gold standards to test both athletes and patients Pollock et al (1998). Exercise capacity has been shown to be the best predictor of survival among healthy men Myers et al (2002). Unfortunately, incremental cardiopulmonary tests are not suitable for all and early prediction could be very useful to avoid overstressing the subjects being tested. In addition, cardiopulmonary testing is time consuming and reducing the time needed to perfom one test could make more affordable cardiopulmonary testing as a routine procedure Franklin and McCullough (2009).

Ways of predicting maximal exercise capacity have been proposed since the origins of exercise physiology Astrand and Ryhı (1954) to indirectly assess the highest workload and the maximal aerobic capacity achieved in the test, thus

avoiding the need of the equipment required to execute the test. Nomograms and short supramaximal protocols, based on the linear relationship between heart rate, workload, and oxygen consumption, have been proposed Arts et al (1993), as well as techniques to estimate from other measures, and thus without a direct measurement Luttikohlt et al (2006), the peak workload elicited by the individual. A series of predicting regression equations, based on factual data such as age, gender, life styles and habits (investigated through specifically designed questionnaires) have been proposed more recently to predict cardiovascular events in the general population according to exercise capacity Myers et al (1994); Nogueira and Pompeu (2006). In Kim et al (2007) a comparison between nomograms is presented, together with a discussion of their use as clinical tools. An evaluation of their accuracy is presented in Patton et al (1982).

A major factor limiting the application of nomograms and other ways of predicting exercise capacity based on factual data is that they are not applicable to measuring a subject's improvement in performance. In our approach, instead, an effectively trained subject will show adaptations to the early stages of testing and W_{peak} will be predicted accordingly. Finally, factual parameters based indices aim at providing a different information (e.g., cardiovascular risk of events stratification) and cannot detect changes in performance in a time as short as a training programme (usually weeks). Hence the need of a more dynamic tool to avoid missing fundamental information about adaptations to training, be more accurate in prescribing exercise intensity, and monitor progression over the course of a training programme. As stated in the recent American College Sport Medicine "the exercise prescription is best adjusted according to individual responses because of the individual variability" Stand (2011).

To our knowledge, the early prediction of the highest workload reached *during* incremental tests has not been investigated. Differently from previous aproaches, our technique does not replace testing, but instead provides an additional tool to reduce the physical effort of the individual and reduce the duration of an ongoing test. Thus, our approach can make cardiopulmonary testing more affordable, while preserving the relevant information that its execution provides. The idea of exploiting the cardiopulmonary efficiency index (CPE) to summarize both the cardiac and pulmonary response of the individual to exercise and to predict the W_{peak} value reached by the individual in the incremental test was first introduced in Baralis et al (2010). This paper significantly improves over the previous approach by (i) presenting a new approach for W_{peak} prediction based on artificial neural networks (ANN) and (ii) analyzing a new set of incremental tests run by using a different test protocol (50 W$\times$ 2 min).

In this paper, the physiological signals monitored during the test are analyzed by means of data mining classification techniques based on an instance based learning approach and an artificial neural network Tan et al (2005). The k-NN-based classifier uses previously labeled data for W_{peak} prediction, without building a classification model from them. The ANN-based classifier, instead, exploits a multilayer feed-forward neural network trained with backpropagation. Given the high variability of individual characteristics and body responses, traditional prediction techniques (e.g., the ARIMA model) Chatfield (2004) are not suitable for our analysis.

Data mining techniques have been widely used in the healthcare domain Green et al (2006); Chuang (2011); Lin et al (2008) to analyze physiological signals and to support clinicians during the diagnosis. Time series classification techniques have been devised to deal with specific physiological signals and to predict the next signal value based on its previous values (e.g., EEG and ECG beat classification Tomioka et al (2007); Yu and Chou (2008)). Differently from these works, our approach does not analyze the monitored signals to estimate their next trend, but instead to predict the maximum value of a different signal, i.e., the workload. Several techniques have been proposed for the general task of multivariate time series classification Yang and Shahabi (2004). In incremental tests, the body response to the increments of workload is peculiar for each individual, and general techniques do not support well these domain specific characteristics (e.g., the joint analysis of factual and dynamic information).

9 Conclusions and Future Work

The highest workload reached in incremental tests is a crucial information to evaluate individual characteristics and plan training intensities. The proposed approach allows a reliable W_{peak} prediction since the early steps of the test, thus allowing a reduction of its duration. As future developments of this work, we will address the following issues.

(i) *Support for individual comparison.* Muscular efficiency has been defined as the ratio between the mechanical power output and the consumed metabolic energy, usually computed from data collected at a steady state. By summarizing both cardiac and pulmonary responses to exercise, the CPE index provides to exercise physiologists additional information on the individual performance in the test. The final CPE value, coupled with the analysis of the CPE trend during the test, may be used to compare the performance of different individuals.

(ii) *Extension to other application domains.* Our approach may be applied on a wide range of protocols characterized by different increments of workload and step durations. Thus, it might be effectively exploited also in the clinical domain to analyze the exercise of pulmonary and cardiac patients. In this domain, protocols are

usually characterized by lower workload increments and shorter step durations.
(iii) *Support for the prediction of additional parameters.* The proposed prediction approach may become a building block to develop a predictor for crucial physiological parameters such as the maximal oxygen consumption (VO_2max) during the test.

10 Acknowledgments

The authors would like to thank Piera Gueli for implementing the ANN-based classifier.

References

Arts F, Kuipers H, Jeukendrup A, Saris W (1993) A short cycle ergometer test to predict maximal workload and maximal oxygen uptake. Int journal of sports medicine 14:460–460

Astrand P, Ryhming I (1954) A nomogram for calculation of aerobic capacity (physical fitness) from pulse rates during submaximal work. Journal of Applied Physiology 7:218–221

Aysin B, Chaparro L, Grave I, Shusterman V (2005) Orthonormal-basis partitioning and time-frequency representation of cardiac rhythm dynamics. IEEE Transaction on Biomedical Engineering 52(5):878–889

Bao P, Ma X (2005) Image adaptive watermarking using wavelet domain singular value decomposition. IEEE Transaction on Circuits and Systems for Video Technology 15(1):96–102

Baralis E, Cerquitelli T, Chiusano S, D'Elia V, Molinari R, Susta D (2010) Predicting the highest workload in cardiopulmonary test. In: CBMS, IEEE, pp 32–37

Baralis E, Cerquitelli T, Chiusano S, Molinari R, Susta D (2011a) Technical report n. TR-2-2011. Early prediction of the highest workload in incremental cardiopulmonary tests. [Online] Available: http://dbdmgpolitoit/wordpress/research/wpeak/

Baralis E, Chiusano S, D'Elia V, Molinari R, Susta D (2011b) Technical report n. TR-1-2011. Early prediction of the highest workload in incremental cardiopulmonary tests. [Online] Available: http://dbdmgpolitoit/wordpress/research/wpeak/

Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: KDD Workshop, AAAI Press, pp 359–370

Chatfield C (2004) The Analysis of Time Series: An Introduction. Chapman & Hall/CRC

Chuang CL (2011) Case-based reasoning support for liver disease diagnosis. Artif Intell Med 53:15–23

Franklin B, McCullough P (2009) Cardiorespiratory fitness: an independent and additive marker of risk stratification and health outcomes. Mayo Clin Proc 84:776–779

Green M, Björk J, Forberg J, Ekelund U, Edenbrandt L, Ohlsson M (2006) Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artif Intell Med 38:305–318

Hassanpour H, Mesbah M, Boashash B (2004) Time-Frequency Feature Extraction of Newborn EEG Seizure Using SVD-Based Techniques. EURASIP Journal on Applied Signal Processing 16:2544–2554

Kasetty S, Stafford C, Walker GP, Wang X, Keogh E (2008) Real-time classification of streaming sensor data. In: Proceedings of the 20th IEEE ICTA, IEEE Computer Society, Washington, DC, USA, Volume 01, pp 149–156

Kim E, Ishwaran H, Blackstone E, Lauer M (2007) External prognostic validations and comparisons of age- and gender-adjusted exercise capacity predictions. J Am Coll Cardiol 50:1867–1875

Lin CS, Chiu JS, Hsieh MH, Mok MS, Li YC, Chiu HW (2008) Predicting hypotensive episodes during spinal anesthesia with the application of artificial neural networks. Comput Methods Prog Biomed 92:193–197

Luttikohlt H, McNaughton L, Midgley A, Bentley D (2006) A prediction model for peak power output from different incremental exercise tests. Int journal of sports physiology and performance 1(2):122 – 136

Mezzani Aea (2009) Standards for the use of cardiopulmonary exercise testing for the functional evaluation of cardiac patients: a report from the exercise physiology section of the european association for cardiovascular prevention and rehabilitation. European Journal of Cardiovascular Prevention and Rehabilitation 16:249–267

Moon T, Stirling W (2000) Mathematical methods and algorithms for signal processing. Prentice Hall, Upper Saddle River, NJ

Myers J, Do D, Herbert W, Ribisl P, Froelicher V (1994) A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data. The American Journal of Cardiology 73:591–596

Myers J, Prakash M, Froelicher V, Do D, Partington S, Atwood J (2002) Exercise capacity and mortality among men referred for exercise testing. New England Journal of Medicine 346:793–801

Nogueira F, Pompeu F (2006) Maximal workload prediction models in the clinical cardio-pulmonary effort test. Arq Bras Cardiol 87:137–145

Patton J, Vogel J, Mello R (1982) Evaluation of a maximal predictive cycle ergometer test of aerobic power. European journal of applied physiology 49:131–140

Pollock M, Gaesser G, Butcher J (1998) The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in health adults. Medicine & Science in Sports & Exercise 30:975–991

RapidMiner (Last access on December 2011) Available at http://rapid-i.com/content/view/181/190/. URL Availableathttp://rapid-i.com/content/view/181/190/

Ratanamahatana C, Keogh E (2004) Making Time-series Classification More Accurate Using Learned Constraints. In: Fourth SIAM International Conference on Data Mining, Lake Buena Vista, Florida, SIAM, pp 11–22

Revill S, Beck K, Morgan M (2002) Comparison of the Peak Exercise Response Measured by the Ramp and 1-min Step Cycle Exercise Protocols in Patients With Exertional Dyspnea. Chest 121(4):1099 – 1105

Sill JM, Morris MJ, Johnson JE, Allan PF, Grbach VX (2009) Cardiopulmonary exercise test interpretation using age matched controls to evaluate exertional dyspnea. Mil Med 174(11):1177–82

Sport Training Center C Eupilio (2011) http://www.csa4sport.it/

Spruit MA, Wouters EFM (2007) New modalities of pulmonary rehabilitation in patients with chronic obstructive pulmonary disease. Sports Med 37(6):501–18

Stand AP (2011) Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise

Tan P, Steinbach M, Kumar V (2005) Introduction to Data Mining. Addison Wesley

Tomioka R, Aihara K, Müller KR (2007) Logistic regression for single trial eeg classification. In: Schölkopf B, Platt J, Hoffman T (eds) Advances in Neural Information Processing Systems 19, MIT Press, Cambridge, MA, pp 1377–1384

Wall M, Dyck P, Brettin T (2001) SVDMAN-singular value decomposition analysis of microarray data. Bioinformatics 17(6):566–568

Yang K, Shahabi C (2004) A pca-based similarity measure for multivariate time series. In: MMDB '04: Proceedings of the 2nd ACM international workshop on Multimedia databases, ACM, New York, NY, USA, pp 65–74

Yu S, Chou K (2008) Integration of independent component analysis and neural networks for ECG beat classification. Expert Systems With Applications 34(4):2841–2846