Supplementary material
Technical Report n. TR-1-2012
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], Davide Susta [3]

I. Introduction

This appendix reports additional experiments to support the evaluation of the proposed approach for the early prediction of the W_{peak} value. These experiments have not been included in the paper due to the lack of space.

The appendix includes the following four sections. Section II describes the dataset used for the experiments. Section III discusses the accuracy yieleded by different designs of the ANN and k-NN classifiers. Section IV analyses the effect of the normalization methods on the prediction error, while Sections V analyzes the impact of the singular values on the prediction error.

II. Dataset

In this report we consider, as representative dataset, tests done with protocol 50 $W \times 2$ min. Table II reports the main characteristics of the dataset, including male athletes.

III. Accuracy of different classifier designs

Figures 1 and 2 show, for ANN and k-NN classifiers, respectively, the Mean Absolute prediction Error (MAE) yielded by different designs of the ANN and k-NN classifiers. For the k-NN classifier (Figure 1), we compared the prediction error when considering (i) the real estimated W_{peak}, (ii) the rounded estimated W_{peak}, computed by approximating the real estimated W_{peak} to the closest workload given by the protocol, and (iii) by exploiting a majority voting schema. In the last case, the classifier predicts a categorical W_{peak} value, equal to the most frequent W_{peak} in the k-nearest tests to the current test. For k-NN classifier the MAE value is always higher in cases (i) and (iii) than in case (ii). Since the configuration (ii) shows the best prediction accuracy, the approximation process performed to estimate the W_{peak} value is beneficial for the k-NN classifier.

[1] Department of Control and Computer Engineering, Politecnico di Torino, Italy (email: elena.baralis@polito.it; tania.cerquitelli@polito.it; silvia.chiusano@polito.it; vincenzo.delia@polito.it)

[2] CSA, Sport Training Center, Eupilio, Como, Italy (email:riccardo@csa4sport.it)

[3] School of Health and Human Performance, Faculty of Science and Health, Dublin City University, Ireland (email:davide.susta@dcu.ie)

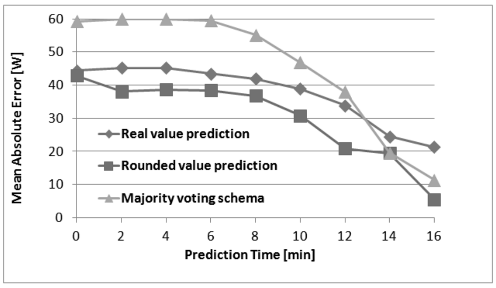

Fig. 1. MAE of different kNN-based classifier designs.

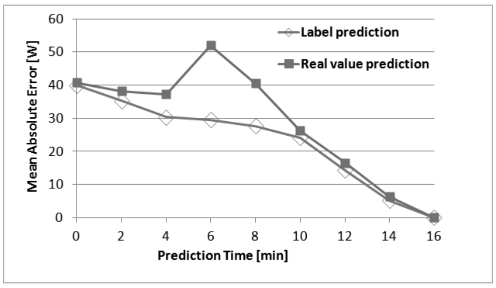

Fig. 2. MAE of different ANN-based classifier designs.

For the ANN-based classifier (Figure 2), we compared the prediction error when considering W_{peak} as either (i) a real or (ii) a categorical value. The experiments show that the MAE value is always significantly higher in case (i) than in case (ii).

For each classification approach (i.e., ANN and k-NN) we select its best configuration: (i) ANN-based classifier predicts the W_{peak} as categorical label, while (ii) kNN-based classifier predicts the rounded estimated W_{peak}.

IV. Effect of the normalization methods on the prediction error

We performed a set of experiments to evaluate the effect of the normalization methods for factual data, on the prediction error (see Figure 3). The k-NN classifier is discussed as reference example.

Figure 3 reports the MAE values by varying the prediction time when either (i) min-max or (ii) z-score methods are exploited to normalize factual data. These data are characterized by a skewed data distribution (i.e., most val-

			Parameters			
Dataset	**No. of Tests**	**No. of Athletes**	**Name**	**Mean ± SD**	**Min value**	**Max value**
$D_{50\times 2}$	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

TABLE I
CHARACTERISTICS OF THE DATASET $D_{50\times 2}$.

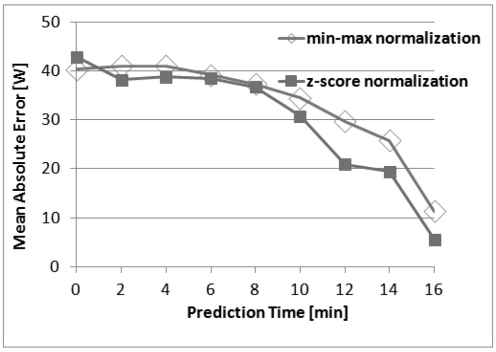

Fig. 3. kNN-based classifier: Impact of the normalization methods on the prediction error.

ues are closed to the median but less frequent values can dominate the min-max normalization results). Since the z-score method is less sensitive to the skewed data distribution, the k-NN classifier based on this method achieves higher accuracy than the other configuration. The differences on prediction error increase for high prediction times.

V. Impact of the number of singular values on the prediction error

We performed a set of experiments to analyze the effect of the approximation provided by the truncated SVD on the MAE value. Experiments have been run with 1, 2, 3, and 9 singular values for matrix T. Increasing the number of singularities reduces the action of signal filtering performed by means of the SVD technique. In our approach, the maximum number of singular values is 9, being 9 the physiological signals used to compute ε_T. The use of all available singularities corresponds to a non-truncated SVD solution, where no filtering step is performed.

In Figure 4 we report the MAE values for a different number of singular values. We discussed the kNN-based classifier as reference classifier. The plots are zoomed on the prediction time range where the difference was more significant (see Figure 5). The experiments show that, in both datasets, our approach is only weakly sensitive to the number of singular values. The MAE value is slightly higher when increasing this number (e.g., 3 or 9 in Figure 5), suggesting that a filtering stage is important to deal with possible noise components.

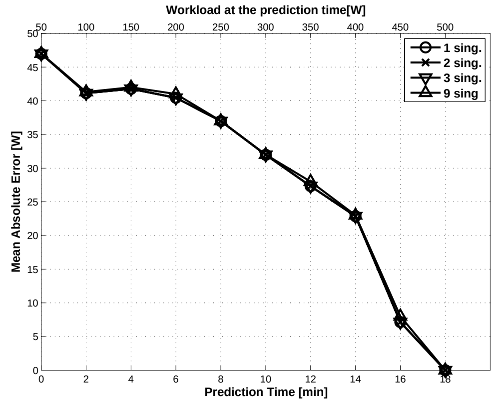

Fig. 4. kNN-based classifier: Impact of the number of considered singular values on the MAE. Prediction time in [0-18] min.

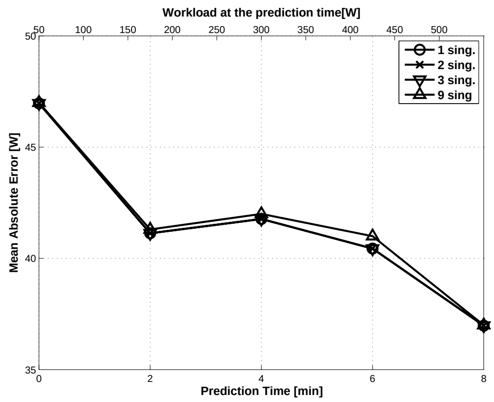

Fig. 5. kNN-based classifier: Impact of the number of considered singular values on the MAE. Prediction time in [0-8] min.