www.elsevier.com/locate/cbm

Developing New VO_2max Prediction Models from Maximal, Submaximal and Questionnaire Variables Using Support Vector Machines Combined with Feature Selection

Fatih Abut, Mehmet Fatih Akay, James George

PII: S0010-4825(16)30278-5
DOI: http://dx.doi.org/10.1016/j.compbiomed.2016.10.018
Reference: CBM2527

To appear in: *Computers in Biology and Medicine*

Cite this article as: Fatih Abut, Mehmet Fatih Akay and James George Developing New VO_2max Prediction Models from Maximal, Submaximal an Questionnaire Variables Using Support Vector Machines Combined with Feature Selection, *Computers in Biology and Medicine* http://dx.doi.org/10.1016/j.compbiomed.2016.10.018

Developing New VO_2max Prediction Models from Maximal, Submaximal and Questionnaire Variables Using Support Vector Machines Combined with Feature Selection

Fatih ABUT[1], Mehmet Fatih AKAY[1], James GEORGE[2]
[1]Dept. of Computer Engineering, Çukurova University, Adana, TURKEY
[2]Dept. of Exercise Sciences, Brigham Young University, Provo, UT, USA

Abstract

Maximal oxygen uptake (VO_2max) is an essential part of health and physical fitness, and refers to the highest rate of oxygen consumption an individual can attain during exhaustive exercise. In this study, for the first time in the literature, we combine the triple of maximal, submaximal and questionnaire variables to propose new VO_2max prediction models using Support Vector Machines (SVM's) combined with the Relief-F feature selector to predict and reveal the distinct predictors of VO_2max. For comparison purposes, hybrid models based on double combinations of maximal, submaximal and questionnaire variables have also been developed. By utilizing 10-fold cross-validation, the performance of the models has been calculated using multiple correlation coefficient (*R*) and root mean square error (*RMSE*). The results show that the best values of *R* and *RMSE*, with 0.94 and 2.92 mL kg^{-1} min^{-1} respectively, have been obtained by combining the triple of relevantly identified maximal, submaximal and questionnaire variables. Compared with the results of the rest of hybrid models in this study and the other prediction models in literature, the reported values of *R* and *RMSE* have been found to be considerably more accurate. The predictor variables gender, age, maximal heart rate (MX-HR), submaximal ending speed (SM-ES) of the treadmill and Perceived Functional Ability (Q-PFA) questionnaire have been found to be the most relevant variables in predicting VO_2max. The results have also been compared with that of Multilayer Perceptron (MLP) and TreeBoost (TB), and it is seen that SVM significantly outperforms other regression methods for prediction of VO_2max.