New Trends and Issues Proceedings on Humanities and Social Sciences

Volume 4, Issue 4 (2017) 01-05

ISSN 2421-8030
www.prosoc.eu

Selected Papers of 6th Cyprus International Conference on Educational Research (CYICER-2017), 04-06 May 2017, Acapulco Hotel and Resort Convention Center, North Cyprus.

New prediction models for the maximal oxygen uptake of college-aged students using non-exercise data

Mehmet Fatih Akay [a]*, Department of Computer Engineering, Faculty of Engineering, Cukurova University, 01000, Adana, Turkey.

Ebru Cetin [b], Department of Physical Education and Sport Teacher, Faculty of Sport Science, Gazi University, 06200, Ankara, Turkey.

Imdat Yarım [c], Department of Physical Education and Sport Teacher, Faculty of Sport Science, Gazi University, 06200, Ankara, Turkey.

Mustafa Mikail Ozciloglu [d], Department of Electrical and Electronics Engineering, Faculty of Engineering, Kilis 7 Aralik University, 79000, Kilis, Turkey.

Suggested Citation:

Akay, F. M., Cetin, E., Yarim, I. & Ozciloglu, M. M. (2017). New prediction models for the maximal oxygen uptake of college-aged students using non-exercise data. *New Trends and Issues Proceedings on Humanities and Social Sciences.* [Online]. 4(4), pp 01-05. Available from: www.prosoc.eu

Selection and peer review under responsibility of Assoc. Prof. Dr. Cigdem Hursen, Near East University

Abstract

Maximal oxygen uptake (VO2max) refers to the maximal amount of oxygen that an individual can utilize during intense or maximal exercise. VO2max plays a significant role in sport science, education and research. The direct measurement of VO2max is time-consuming, requires expensive laboratory equipment and trained staff. Because of these disadvantages of direct measurement, numerous VO2max prediction models for a variety of subject groups have been developed. The purpose of this study is to develop new Multiple Linear Regression based on VO2max prediction models for Turkish college students by using physiological and questionnaire variables. The dataset includes the data of 62 (28 females and 34 males) students, ranging in age from 18 to 27 years, from the College of Physical Education and Sports Science at Gazi University. Seven different models consisting of the predictor variables gender, age, weight, height, Perceived Functional Ability scores (PFA-1 and PFA-2), and Physical Activity Rating score (PA-R) have been used to predict VO2max. The performance of the prediction models has been evaluated by calculating their standard error of estimates (SEE's) and multiple correlation coefficients (R's). The prediction model including Gender, Age, Height, Weight, PFA-1 and PAR yields the lowest SEE with 5.14 mL.kg-1.min-1 and highest R with 0.93. It can be concluded that in situations where it is difficult to measure VO2max, the given model with MLR equation can be used to predict the VO2max of college students with acceptable error rates.

Keywords: Maximum oxygen uptake, machine learning, multiple linear regression.

* ADDRESS FOR CORRESPONDENCE: **Mehmet Fatih Akay**, Department of Computer Engineering, Faculty of Engineering, Cukurova University, 01000, Adana, Turkey.
E-mail address: mfakay@cu.edu.tr / Tel.: +90-322-3387101

1. Introduction

Maximal oxygen consumption (VO_2max) is the maximum amount of oxygen that an individual can utilize during intense or maximal exercise. VO_2max plays a significant role in sport science, education and research. In sport sciences, it is often used as an indicator for endurance capacity of athletes, representing the upper limit of their aerobic fitness (Abut, Akay & George, 2016).

In medical sciences, it can serve as a metric to estimate the disease risk of a person, suggesting an unusually large amount of cholesterol, body fat and blood pressure. The most accurate method to assess VO_2max is directly measuring the oxygen uptake during graded, maximal exertion exercise on a treadmill or cycle ergo meter in the laboratory (Hunn, Lapuma & Holt, 2002; Eler, 2016). However, this technique requires expensive laboratory equipment, a great deal of time, continuous medical supervision and highly motivated subjects (Bandyopadhyay 2013; George et al., 2009).

In literature, only a few studies exist on VO_2max prediction of Turkish athletes. In Kaya, Akay, Cetin and Yarim (2016), Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Single Decision Tree (SDT) were used on a dataset which included the data of 48 students. Age, height, weight, body mass index (BMI), test time and maximum heart rate (HRmax) were used to predict VO_2max. That study showed VO_2max of Turkish athletes could be predicted with reasonable error rates by using SVM. Dincer, Akay, Cetin, Yarim and Daneshvar (2016) predicted VO_2max of college-aged students using Multiple Linear Regression and hybrid data, which was a combination of exercise data and questionnaire variables. 26 students from the College of Physical Education and Sports at Gazi University participated in the experiments. The dataset included gender, age, height, weight, BMI, HRmax, test time (TT), Perceived Functional Ability (PFA) and Physical Activity Rating (PA-R). This study suggested that the prediction equation, VO_2max = - (7.42 x gender) + (4.26 x age) - (1.44 x BMI) + (4.31 x HRmax) + (3.64 x TT) - (0.16 x PFA-1) + (0.75 x PFA-2) + (0.61 x PAR) - 895.26 yielded the lowest *SEE*. (Akay, Cetin, Yarim, Abut & Kaya, 2016) established new prediction equations for estimating VO_2max from gender, age, height, weight, BMI, HRmax and TT for college-aged students in Turkey. Particularly, 18 students from the College of Physical Education and Sports at Gazi University volunteered for that study. Twelve VO_2max prediction equations had been established by using MLR. The results showed that the regression equation, VO_2max = - (12.331 x gender) - (0.805 x age) + (0.883 x height) - (1.167 x weight) - (0.052 x HRmax) - (0.158 x TT) + 6.473, gave the lowest *SEE* and the highest *R* (Ozciloglu, Akay, Cetin & Yarim, 2016) developed sub maximal VO_2max prediction models for Turkish college students by using SVM, MLP and MLR. The dataset included data of 65 students from the College of Physical Education and Sport at Gazi University. To predict VO_2max, two categories of prediction models had been formed. In the first category, the common predictor variables in each model were gender, age, height and weight whereas the models in the second category had common predictor variables which were gender, age and BMI. The rest of the predictor variables for both categories were time, speed and sub maximal heart rate (HRsmax). That study showed that the models consisting of the common predictor variables together with solely time yielded the lowest *SEE*'s for prediction of VO_2max in each category by using SVM.

The purpose of this study is to develop new Multiple Linear Regression based on VO_2max prediction models for Turkish college students by using physiological and questionnaire variables. The dataset includes the data of 62 (28 females and 34 males) students, ranging in age from 18 to 27 years, from the College of Physical Education and Sports at Gazi University. Seven different models consisting of the predictor variables were gender, age, weight, height, Perceived Functional Ability scores (PFA-1 and PFA-2), and Physical Activity Rating score (PA-R) have been used to predict VO_2max. The performance of the prediction models has been evaluated by calculating their standard error of estimates (*SEE*'s) and multiple correlation coefficients (*R*'s). The results show that MLR models can be used to predict VO_2max accurately for college-aged sports students in Turkey.

The rest of the paper is organized as follows. Section 2 describes dataset generation. Section 3 introduces MLR based models. Section 4 gives results and discussion. Section 5 concludes the paper.

2. Dataset Generation

All subjects were informed prior to the maximal exercise test and they signed a consent participant form before participating in the tests. Maximal exercise test was applied to subjects to obtain their VO_2max values. During the exercise the test that was performed on a treadmill (HP COSMOS, Germany), a subject had been forced until he/she showed maximal performance. In other words, the test continued until the subject was exhausted.

During the maximal test using the maximal stepwise running exercise protocol, each subjects' HRmax was measured and registered every 15 seconds. The maximal oxygen consumption capacities of participants were measured with the Cosmed Quark CPET system (Cosmed Quark CPET; Rome, Italy) by breath-by-breath technique. In addition to HRmax, tidal volume, VO_2max and respiratory exchange ratio were also recorded every 15 seconds. VO_2max test protocol started with running at 0^o incline and at a speed of 8 km/h for women and at a speed of 10 km/h for men. Speed was incremented by 1 km/h every minute until 15km/h speed level was reached. Upon reaching 15 km/h speed, the incline started to increase by 1.5^o each minute and the test continued until the athlete got exhausted. Statistical information about the dataset is shown in Table 1.

Table 1. Statistics of variables

Predictor Variable	Minimum	Maximum	Mean	Standard Deviation
Gender	0	1.00	0.56	0.50
Age (year)	18.00	27.00	20.79	1.98
Weight (kg)	44.00	95.00	64.84	11.19
Height (cm)	153.00	193.00	172.53	7.73
PFA-1	2.00	7.00	4.79	1.47
PFA-2	1.00	9.00	3.81	2.06
PAR	1.00	10.00	5.61	2.77
VO_2max ($ml.kg^{-1}.min^{-1}$)	35.21	87.95	50.97	11.48

3. Methodology

Multiple Linear Regression (MLR) is frequently used in statistical analysis because of its flexibility. MLR is an extension of simple linear regression model in such a way that MLR uses two or more dependent variables in a prediction formula to estimate a desired variable. Even with complicated regression models including a large number of variables, MLR requires little effort to generate predictions (Slinker & Glantz, 2008).

By using combinations of the predictor variables, seven different VO_2max prediction models have been produced. The performance of the prediction models has been evaluated using *SEE* and *R,* the formulas of which are given in (1) and (2), respectively.

$$SEE = \sqrt{\frac{\sum(Y - Y')^2}{N}} \quad (1)$$

$$R = \sqrt{1 - \frac{\sum(Y - Y')^2}{\sum(Y - \overline{Y})^2}} \quad (2)$$

In (1) and (2), Y is the measured VO_2max, Y' is the predicted VO_2max, $\overline{Y}$ is the average of the measured values of VO_2max and N is the number of subjects in the dataset.

4. Results and Discussion

Table 2 shows the *SEE*'s and *R*'s of MLR based models along with the predictor variables. The prediction models are sorted by *SEE* values in rising order.

Table 2. *SEE* and *R* values of VO_2max prediction models

Models	Predictor Variables	Results	
		SEE	*R*
Model 6	Gender, Age, Height, Weight, PFA-1, PAR	5.135	0.926
Model 1	Gender, Age, Height, Weight, PFA-1, PFA-2, PAR	5.432	0.917
Model 4	Gender, Age, Height, Weight, PFA-2, PAR	5.724	0.908
Model 7	Gender, Age, Height, Weight, PAR	5.793	0.905
Model 3	Gender, Age, Height, Weight, PFA-1	6.741	0.869
Model 2	Gender, Age, Height, Weight, PFA-1, PFA-2	7.408	0.839
Model 5	Gender, Age, Height, Weight, PFA-2	8.717	0.769

The following discussions can be made regarding the results obtained:

- The outcomes indicate that the models consisting of the common predictor variables together with PFA-1 and PAR yields the lowest *SEE*'s and highest *R*'s for prediction of VO_2max.
- When Model 3, including the common predictor variables and PFA-1, and Model 6, including the common predictor variables PAF-1 and PAR, are compared it can be observed that PAR provides a significant improvement for prediction of VO_2max. In more detail, the inclusion of PAR in the aforementioned model leads in 23.82% reduction in *SEE* for predicting VO_2max.
- Likewise, the inclusion of PAR in Model 5 yields in 34.33% reduction in *SEE* for predicting VO_2max.
- The prediction equation, VO_2max = (15.47 x gender) - (0.12 x age) + (0.04 x height) - (0.45 x weight) + (1.74 x PFA-1) + (1.45 x PAR) + 49.74 yields the lowest *SEE* with 5.14 $mL.kg^{-1}.min^{-1}$ and highest *R* with 0.93.

5. Conclusion

In this study, Multiple Linear Regression has been used to create seven new different VO_2max prediction models for Turkish college students by using physiological and questionnaire variables. It can be concluded that in situations where it is difficult to measure VO_2max, the given equation can be used to predict VO_2max of college students with acceptable error rates. Also, the prediction models including the variable PAR gives significant improvements for VO_2max prediction. Future work can involve using different machine learning methods combined with feature selection algorithms to advance the accuracy of VO_2max prediction.

Acknowledgment

The authors would like to thank Cukurova University Scientific Research Projects Center for supporting this work under grant no. FBA-2016-5537.

References

Abut, F., Akay, M. F. & George J. (2016). Developing new VO(2)max prediction models from maximal, sub maximal and questionnaire variables using support vector machines combined with feature selection. *Journal of Computers in Biology and Medicine, 79,* 182-192.

Akay, M. F., Cetin E., Yarim I., Abut F. & Kaya K. (2016). *New regression equations for estimating the maximal oxygen uptake of college of physical education and sports students in Turkey.* In Proceedings of 5th Cyprus International Conference on Educational Research, Kyrenia, North Cyprus., 31 March - 2 April 2016 (pp.4-7).

Bandyopadhyay, A. (2013). Validity of 20 meter multi-stage shuttle run test for estimation of maximum oxygen uptake in male university students. *Indian Journal of Physiology Pharmacol, 57*(1), 77–83.

Dincer, O. F., Akay, M. F., Cetin, E., Yarim I. & Daneshvar, S. (2016). *New prediction equations for estimating the maximal oxygen consumption of college-aged students using hybrid data*. In Proceedings of the 1st International Mediterranean Science and Engineering Congress, Adana, Turkey, 26-28 Oct 2016 (pp.1-1).

Eler, S. (2016). Effects of short term camp periods on aerobic and anaerobic performance parameters in ice hockey national team athletes. *International Journal of Environmental and Science Education, 11*(5), 973-977.

George, J. D., Paul, S. L., Hyde, A., Bradshaw, D. I., Vehrs, P. R., Hager, R. L. & Yanowitz, F. G. (2009). Prediction of maximum oxygen uptake using both exercise and non-exercise data. *Journal of Measurement in Physical Education and Exercise Science, 13*(1), 1-12.

Hunn, H. M., Lapuma P. T. & Holt, D. T. (2002). The influence of pre-test anxiety, personality and exercise on VO_2max estimation. *Journal of Exercise Physiology, 5*(1), 5–14.

Kaya, K., Akay, M. F., Cetin, E. & Yarim, I. (2016). *Development of new prediction models for maximal oxygen uptake using artificial intelligence methods.* In Proceedings of the International Conference on Natural Science and Engineering, Kilis, Turkey, 19-20 March 2016 (pp. 986-988).

Ozciloglu, M. M., Akay, M. F., Cetin, E. & Yarim, I. (2016). *Development of new maximum oxygen uptake prediction models for turkish college students using support vector machines and submaximal data"*. In Proceedings of the Fourth International Symposium on Engineering, Artificial Intelligence and Applications, Kyrenia, North Cyprus, 2-4 Nov 2016 (pp.19-20).

Slinker, B. K. & Glantz, S. A. (2008). *Multiple linear regression accounting for multiple simultaneous determinants of a continuous dependent variable. American Heart Association, 117*(13), 1732-1737.