GÜFBED/*GUSTIJ* (2018) CMES 2018 SEMPOZYUM EK SAYISI: 7-10
DOI: 10.17714/gumusfenbil.435897 Araştırma Makalesi / *Research Article*

Development of Physical Fitness Prediction Models for Turkish Secondary School Students Using Machine Learning Methods

Türk Ortaokul Öğrencileri için Makine Öğrenmesi Yöntemleri Kullanılarak Fiziksel Uygunluk Tahmin Modelleri Geliştirme

M.Fatih AKAY*[1,a], Ebru ÇETIN[2,b], İmdat YARIM[2,c], Özge BOZKURT[1,d], Sevtap ERDEM[1,e]
[1] *Department of Computer Engineering, Çukurova University, Adana, Turkey*
[2] *School of Physical Education and Sport, Gazi University, Ankara, Turkey*

• Geliş tarihi / *Received*: 30.06.2018 • Düzeltilerek geliş tarihi / *Received in revised form*: 01.08.2018 • Kabul tarihi / *Accepted*: 12.09.2018

Öz
Fiziksel uygunluk, belirli testlerle ölçülebilen sağlık veya beceri ile ilgili bir dizi özelliktir. Fiziksel uygunluğu korumak sağlık ve esenlik için çok önemlidir. Ancak, fiziksel uygunluğun ölçülmesi profesyonel ekipman, deneyimli personel ve çok zaman gerektirdiğinden, araştırmacıların fiziksel uygunluğu belirlemek için farklı yollara ihtiyaçları vardır. Bu çalışmanın amacı, Destek Vektör Makineleri (SVM), Radyal Tabanlı Fonksiyon Sinir Ağı (RBFNN) ve Ağaç Artımı (TB) gibi makine öğrenme yöntemlerini kullanarak Türk ortaokul öğrencilerinin fiziksel uygunluğunu tahmin etmek için yeni tahmin modelleri geliştirmektir. Veri seti 30m hız, 20m aşama koşusu, denge ve çeviklik testlerinin sonuçlarından oluşan veriyi içermektedir. Tahmin modellerini geliştirmek için kullanılan tahmin değişkenleri cinsiyet, yaş, boy, kilo, vücut yağı, 30 saniyedeki mekik ve şınav sayılarından oluşmaktadır. Tahmin modellerinin performansı Ortalama Karesel Hata (*RMSE*) kullanılarak hesaplanmıştır. Sonuçlar, SVM tabanlı tahmin modellerinin, RBFNN ve TB'ye dayanan diğer modelleri geride bıraktığını göstermektedir. Ayrıca, fiziksel uygunluk tahmini için vücut yağı, mekik ve şınav gibi tahmin değişkenlerinin birlikte kullanılması durumunda sonuçlar üzerinde önemli bir rol oynadığını gösterilmiştir.

Anahtar kelimeler: Fiziksel uygunluk, Makine öğrenmesi, Tahmin.

Abstract
Physical fitness is a set of attributes that are either health or skill-related which can be measured with specific tests. Maintaining physical fitness is essential for health and wellbeing. However, since measurement of physical fitness requires improved professional equipment, experienced staff and lots of time, researchers need different ways to determine physical fitness. The aim of this study is to develop new prediction models for predicting the physical fitness of Turkish secondary school students by using machine learning methods including Support Vector Machines (SVM), Radial Basis Function Neural Network (RBFNN) and Tree Boost (TB). The dataset comprises data of various number of subjects according to the target variables such as the test scores of the 30m speed, 20m stage run, balance and agility. The predictor variables used to develop the prediction models are gender, age, height, weight, body fat, number of curl-up and push-ups in 30 seconds. Root Mean Square Error (RMSE) has been utilized to assess the performance of the prediction models. Based on the results we can conclude that SVM based prediction models outperform other models based on RBFNN and TB. Also, the predictor variables body fat, push-up and curl-up play a significant role when used all together for physical fitness prediction.

***Keywords**: Physical fitness, Machine learning, Prediction.*

*[a] Mehmet Fatih AKAY; makay@cu.edu.tr; Tel: (0322) 338 71 01; orcid.org/0000-0003-0780-0679
[b] orcid.org/0000-0001-7190-4929 [c] orcid.org/0000-0003-4794-129x [d] orcid.org/0000-0001-8633-6696
[e] orcid.org/0000-0002-9332-2070

1. Introduction

Physical fitness is a state of health and well-being and, more specifically, the ability to perform aspects of sports, occupations and daily activities. Physical fitness plays an important role in our lives and can improve our health and reduce the risk of developing several diseases like type 2 diabetes, cancer and cardiovascular disease. Fitness education and student fitness assessments offer students an opportunity to assess, track, and improve their fitness level (Hoffman, 2006). However, physical fitness assessment using test protocols requires improved professional equipment, experienced staff and lots of time. Therefore, several physical fitness prediction models using statistical and machine learning methods have been proposed in literature. In (Ahmed and Loutfi, 2013), Case-Based Reasoning (CBR) approach to identify physical activity of elderly based on pulse rate has been proposed. The CBR approach has been compared with the two popular classification techniques including SVM and Neural Network (NN) on 24 subjects. In (Dijkhuis, Blssuw, Ittersum and Aiello, 2018), an activity tracker to record participants' daily step count has been used as input for a coaching session. The gathered step count data was used to train eight different machine learning algorithms to make hourly estimations of the probability of achieving a personalized, daily steps threshold for the number of 48 subjects. In (Fergus et al., 2015), a supervised machine learning approach has been adopted by a set of activities and features suitable for measuring physical activity and evaluates the use of a Multilayer Perceptron neural network for the number of 28 subjects. In (Reichherzer et all., 2017), the data analysis methods have been used to train a classifier for records with the individuals, their physical activities, and conditions under which they were performed. Four different machine learning algorithms that Decision Tree (DT), Random Forest (RF), SVM, Naive Bayesian (NB) methods were used to make predictions for 29 participants.

The limitations of the studies in literature are as follows: All the studies concentrate on classifying physical activity levels rather than predicting the actual test results. As we can observe from literature reviews, the number of subject in the datasets is limited (less than 50) and the developed models require the subject to complete several exercises in order to have physical activity level rates.

In this study, the dataset which is covering approximately 400 subjects, was analyzed using rigorous data science techniques, which led to an improved understanding of activity types and features. A series of machine learning analyses were performed to develop improved prediction accuracy. Unlike previous studies, the predictor variables and the prediction methods were expanded to have more comprehensive prediction. This study proposes to develop new prediction models for Turkish secondary school students by using SVM, RBFNN and TB.

2. Dataset Generation

The dataset comprises of different number of subjects depending on the target variables of healthy secondary school students. Different physical exercise tests were applied on the subjects for measuring their physical fitness. A consent participant form was signed by all subjects participating in this study. Participants were assigned to perform the following core stabilization assessments 30 meter (m) Speed, 20m Stage Run, Balance and Agility Tests to predict physical fitness.

3. Results and Discussion

Three different machine learning methods including SVM, RBFNN and TB have been employed in order to develop physical fitness prediction models. SVM is a state-of-the-art regression method which is widely utilized in many application areas due to its high accuracy (Chuang et al., 2011; Wang, 2005; Abut et al., 2015). RBFNN is a particular type of neural network and is becoming an increasingly popular neural network with diverse applications. RBFNN consists of three layers: an input layer, a hidden (kernel) layer with a non-linear RBFNN activation function and a linear output layer. The nodes within each layer are fully connected to the previous layer (Hannan et al., 2010). The TB is a technique for improving the accuracy of a predictive function by applying the function repeatedly in a series and combining the output of each function with weighting so that the total error of the prediction is minimized.

Eight prediction models have been produced by using combinations of the predictor variables. *RMSE*, the equation which is given below, has been used to assess the performance of the prediction models.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y - Y'\right)^2} \quad (1)$$

In (1), Y is the measured value, Y' is the predicted value, and n is the number of samples in a test subset.

Table 1 shows the physical fitness prediction models. Table 2 shows the results for each model.

Table 1. Physical fitness prediction models

Model Number	Predictor Variables
1	Gender, Age, Weight, Height
2	Gender, Age, Weight, Height, Curl-up
3	Gender, Age, Weight, Height, Push-up
4	Gender, Age, Weight, Height, Body Fat
5	Gender, Age, Weight, Height, Curl-up, Push-up
6	Gender, Age, Weight, Height, Curl-up, Body Fat
7	Gender, Age, Weight, Height, Push-up, Body Fat
8	Gender, Age, Weight, Height, Curl-up, Push-up, Body Fat

Table 2. *RMSE* values of physical fitness prediction models

Model No	30m Speed (s)			20m Stage Run (s)			Balance (s)			Agility (s)		
	SVM	**RBF**	**TB**	**SVM**	**RBF**	**TB**	**SVM**	**RBF**	**TB**	**SVM**	**RBF**	**TB**
1	0.51	0.55	0.58	2.11	2.17	2.17	6.65	7.41	9.85	1.76	1.83	1.86
2	0.47	0.53	0.50	2.06	2.13	2.14	6.60	6.99	9.83	1.63	2.48	1.72
3	0.49	0.50	0.53	2.07	2.17	2.11	6.62	7.10	9.82	1.68	1.87	1.75
4	0.48	0.51	0.54	2.06	2.22	2.22	6.70	6.94	9.80	1.74	1.87	1.85
5	0.44	0.47	0.46	2.04	2.14	2.09	6.62	7.65	9.81	1.58	1.69	1.64
6	0.44	0.48	0.47	2.04	2.14	2.16	6.67	7.43	9.84	1.61	1.76	1.72
7	0.46	0.58	0.49	2.05	2.18	2.15	6.62	7.44	9.81	1.68	1.84	1.76
8	0.43	0.49	0.45	2.03	2.22	2.12	6.63	9.69	9.80	1.59	1.90	1.65

SVM based models usually outperform RBFNN based and TB based models by yielding lower *RMSE*'s. The worst performance has been observed in TB based models. The predictor variables number of curl-up and push-ups in 30 seconds and body fat play a significant role in physical fitness prediction.

4. Conclusion

This is an initial study showing that SVM is a viable method that can be safely used to predict the physical fitness of Turkish secondary school students. In this context, several models have been developed to predict the results of 30m Speed, 20m Stage Run, Balance and Agility tests. However, future work is definitely required to improve the accuracy of physical fitness prediction models. Future work will include developing deep learning based physical fitness prediction models as well as integrating feature selection algorithms.

Acknowledgment

The current study has been financed by Cukurova University Research Projects Center under grant no. FYL-2018-10430.

References

Abut, F., Akay, M.F., Sow, B. and George, J., 2015. Support Vector Machines for Prediction of Endurance Times Involving Isometric Side Bridge Exercise Test, Third International Symposium on Engineering, Artificial Intelligence & Applications (ISEAIA2015), North Cyprus, pp. 7-9.

Ahmed, M.U., Loutfi, A., 2013. Pyhsical Activity Identification using Supervised Machine Learning and based on Pulse Rate, International Journal of Computer Science and Applications, Publication 4(7).

Chuang, C.C., Lee, Z.C., 2011. Hybrid robust support vector machines for regression without liers, Applied Soft Computing.

Dijkhuis, T.B., Blaauw, F., Ittersum, M. W., Aiello, M., 2018. Personalized Physical Activity Coaching : A Machine Learning Approach, in Sensors, 18(2), p.623.

Fergus, P., Hussain, A., Hearty, J., Fairclough, S., Boddy, L., Machintoch, K.A., Stratton,G., Ridgers, N.D. and Radi, N., 2015. A Machine Learning Approach to Measure and Monitor Physical Activity in Children to Help Fight Overweight and Obesity, Intelligent Computing Theories and Methodologies: 11th International Conference, ICIC 2015, Fuzhou, China, pp. 676-688.

Hannan, S.A., Manza, R.R., Ramteke, R.J., 2010. Generalized Regression Neural Network and Radial Basis Function for Heart Disease Diagnosis, International Journal of Computer Applications, Volume 13, Issue 2, ISBN: 10.5120/1325-1799.

Hoffman, J., 2006. Norms for Fitness, Performance and Health, Human Kinetics Publishers, ISBN-13: 9780736054836, 97–104.

Reichherzer, T., Timm, M., Earley, N., Reyes, N., 2017. Using Machine Learning Techniques to Track Individuals & Their Fitness Activities, In A. Bossard, G. Lee, & L. Miller (Eds.), Proceedings of 32nd International Conference on Computers and Their Applications, Honolulu, Hawaii, USA, pp. 119–124.

Wang, L., 2005. Support Vector Machines: Theory and Applications, Springer Science & Business Media, Volume 177, 1st Edition: ISBN: 978-3-540-32384-6.