Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 31(ÖS 2), ss. ÖS 153-ÖS 159, Ekim 2016
Çukurova University Journal of the Faculty of Engineering and Architecture, 31(SI 2), pp. SI 153-SI 159, October 2016

New Regression Models for Predicting the Hamstring Muscle Strength using Support Vector Machines

Boubacar SOW[1], Mehmet Fatih AKAY[*1], Fatih ABUT[1], Ebru ÇETİN[2], İmdat YARIM[2], Hacer ALAK[1]

[1]*Çukurova Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, Adana*
[2]*Gazi Üniversitesi, Beden Eğitimi ve Spor Yüksekokulu, Ankara*

Geliş tarihi: 22.08.2016 *Kabul tarihi: 24.09.2016*

Abstract

The purpose of this study is to build new prediction models for estimating the hamstring muscle strength of college-aged athletes using Support Vector Machine (SVM). The dataset is made up of 70 athletes ranging in age from 19 to 38 years who were selected from the College of Physical Education and Sport at Gazi University. The results show that the prediction model including the predictor variables gender, age, height and weight provides a valid and convenient method for estimating hamstring muscle strength within limits of acceptable accuracy. For comparison purposes, prediction models based on Multilayer Perceptron (MLP) and Single Decision Tree (SDT) have also been created, and it is seen that SVM-based models outperforms the MLP-based and SDT-based models for prediction of hamstring muscle strength.

Keywords: SVM, MLP, SDT, Hamstring muscle strength, Prediction

Destek Vektör Makinelerini Kullanarak Hamstring Kas Kuvveti Tahmini için Yeni Regresyon Modelleri

Öz

Bu çalışmanın amacı, Destek Vektör Makinesi (DVM) kullanarak üniversite çağındaki sporcuların hamstring kas kuvvetini tahmin etmek için yeni tahmin modelleri oluşturmaktır. Veri seti, yaşları 19 ve 38 arasında değişen, Gazi Üniversitesi Beden Eğitimi ve Spor Yüksekokulu'ndan seçilen 70 sporcudan oluşmaktadır. Elde edilen sonuçlara göre; cinsiyet, yaş, boy ve kilo değişkenlerini içeren tahmin modelinin, kabul edilebilir doğruluk ile hamstring kas kuvvetini tahmin etmek için geçerli ve kullanışlı bir yöntem sağladığını göstermektedir. Karşılaştırma amacıyla, Çok Katmanlı Algılayıcı (ÇKA) ve Tekli Karar Ağacı (TKA) yöntemlerine dayalı tahmin modelleri de oluşturulmuştur ve DVM tabanlı modellerin, hamstring kas gücünün tahmininde ÇKA ve TKA tabanlı modellerden daha iyi performans sergilediği görülmüştür.

Anahtar Kelimeler: DVM, ÇKA, TKA, Hamstring kas kuvveti, Tahmin

[*]Sorumlu yazar (Corresponding author): Mehmet Fatih AKAY, *mfakay@cu.edu.tr*

1. INTRODUCTION

Muscular strength is the maximum amount of force that a muscle can exert against some form of resistance in a single effort. Muscles support the skeleton and enable movement. Strong muscles in the legs such as hamstrings, but also muscles of buttocks, abdomen, chest and shoulder provide a person with the strength to stand up straight and maintain good posture. Taking place on the back of the upper leg, the hamstring muscles play a crucial role in many daily activities, such as walking, running and jumping. The hamstring muscles are responsible for the flexion of the knee as well as assisting the extension of the thigh. The hamstring muscle strength is measured in Nm [1].

In the courses of the past decades, various techniques have been proposed to quantitatively and accurately measure the hamstring muscle strength including dynamometer tests [2], tensiometer tests [3] or isokinetic tests [4]. However, among the various types, isokinetic testing has become the most popular and common technique for directly measuring the hamstring muscle strength of the upper leg. Isokinetic exercise is usually conducted by using the so-called dynamometers which sustain a constant velocity of movement. Prior to performing the isokinetic exercise, the participant is stationed in such a way that the body movement to be measured is isolated. Afterwards, the dynamometer is adjusted at different velocities and the force exerted by the participant can be measured over the entire range of movement [5].

Despite a high level of accuracy, the direct measurement of hamstring muscle strength is associated with a number of practical difficulties and limitations. The equipment required for conducting the measurements is bulky, expensive and not readily available. In particular, such measurement activities are frequently conducted within the scope of research projects at educational institutions or provided as services in rehabilitation or health-care facilities. Also, it is only possible to test one participant at a time so that the practical application of direct measurement is not feasible for large populations. Regarding these difficulties, rather than directly measuring the hamstring muscle strength, it may be beneficial to predict it using machine learning methods.

In literature, to the best of our knowledge, there is only one preliminary study [6] that compares the performance of different machine learning methods for prediction of hamstring muscle strength. Particularly, in [6]; SVM, Radial Basis Function Neural Network and Decision Tree Forest have been applied to develop a model for hamstring muscle strength prediction. It has been concluded that SVM-based model yields lower *RMSE*'s than the ones obtained by using other methods. However, this study has the significant limitation that it only considers a single prediction model for performance evaluation. Additional studies focusing on developing further new models are definitely required in order to identify the best set of predictor variables for hamstring muscle strength prediction.

The aim of this study is to extend the work of [6] by building new and more comprehensive prediction models for estimating the hamstring muscle strength of college-aged athletes using SVM. The dataset included 70 volunteers who were students at the Department of Physical Education and Sport in Gazi University. The predictor variables gender, age, height, weight and sport branch were utilized to build thirty different hamstring muscle strength prediction models. The generalization error of the prediction models has been calculated by carrying out 10-fold cross-validation, and the prediction errors have been computed using root mean square error (*RMSE)* and correlation coefficient (*R)*. The results of SVM-based prediction models have also been compared with the ones obtained by MLP-based and SDT-based prediction models. The results show that the SVM-based model containing the predictor variables gender, age, height and weight yields the lowest *RMSE* and highest *R* with 15.19

Nm and 0.82, respectively. Moreover, this *RMSE* value has been found to be 2.25% lower than that of the SVM-based prediction model presented in [6], which used the same dataset as in this study. Also, it is observed that all SVM-based models perform better than the MLP-based and SDT-based prediction models, whereas MLP-based prediction models outperform SDT-based prediction models.

The rest of this paper is divided into five sections. Section 2 gives brief information about dataset generation. Section 3 summarizes the methodology used to build the prediction models. The results are discussed in Section 4. Finally, Section 5 concludes the paper.

2. DATASET GENERATION

70 students from the College of Physical Education and Sport at Gazi University were involved in the experiments. Isokinetic strength of all subjects' right upper leg hamstring muscle was measured by the isokinetic dynamometer (Isomed 2000, Germany) at 60° per second angular velocity. The hamstring muscle strength has been measured by performing the classic training which involved light run for 5 minutes. The dataset included the predictor variables gender, age, height, weight and sport branch as well as the target variable hamstring muscle strength. Table 1 gives the minimum, maximum, mean and standard deviation values for each predictor and target variable.

Table 1. Statistics of the dataset

Variables	Minimum	Maximum	Mean	Standard Deviation
Gender	0	1.00	0.36	0.48
Age (Year)	19.00	38.00	21.79	3.06
Height (m)	1.57	2.02	1.71	0.08
Weight (kg)	45.00	93.00	62.04	11.27
Sport Branch	1.00	17.00	9.31	5.03
Hamstring Muscle Strength (Nm)	50.10	195.90	111.84	36.10

3. METHODOLOGY

Four different categories of prediction models have been developed, whereby each category contains a different set and number of predictor variables. Particularly; the first, second, third and fourth categories of prediction models include the single, double, triple and quadruple combinations of predictor variables gender, age, height, weight and sport branch, respectively, which in total yield thirty different hamstring muscle strength prediction models. The performance of all models has been evaluated by using 10-fold cross validation and computing the values of *RMSE* and *R*, whose equations are given as follows:

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

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

In Eqs. (1) and (2), Y is the measured hamstring muscle strength value, Y' is the predicted hamstring muscle strength value, $\overline{Y}$ is the mean of the measured hamstring muscle strength values and n is the number of samples in a test subset.

Three machine learning methods including SVM, MLP and SDT have been utilized to build the hamstring muscle strength prediction models. In the field of sport physiology, SVM has been reported to be a promising method that has shown satisfactory performance for a variety of problems [7–9]. The accuracy of an SVM model is largely dependent on the selection of the model parameters such as C, ε and the type and parameters of kernel function. As the kernel function, the radial basis function kernel has been chosen which requires the optimization of γ. Hence, one needs an effective search algorithm to find the best values of the triple (C, ε, γ). In this study, grid search and cross validation have been used in order to determine the best values of the mentioned parameters and to overcome the problem of generalization [10].

MLP is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. In contrast to single-layer perceptrons which can only represent linear decision surfaces, MLP's can represent non-linear decision surfaces. The performance of the MLP network has been improved by varying the number of hidden layers, convergence tries and maximum iterations. The logistic sigmoid activation function has been used for the hidden layers for prediction of the hamstring muscle strength [11].

Decision trees predict the value of the target variable by using the values of predictor variables and building regression models in form of a tree structure. It divides a dataset into smaller and smaller parts while at the same time an associated decision tree is progressively built. In order to get more accurate value of the predicted variable, three important parameters including minimum rows in a node, minimum size node to split and maximum tree levels of SDT-based prediction models have been optimized [12].

Table 2 shows the list of intervals for values of the utilized parameters for SVM-based, MLP-based and SDT-based prediction models.

Table 2. List of intervals for values of the utilized parameters for SVM-based, MLP-based and SDT-based prediction models

Method	Parameter	Range
SVM	Cost (C)	[2^{-3}-2^{12}]
	Epsilon (ε)	[0.0001-150]
	Gamma (γ)	[2^{-9}-2^{7}]
MLP	Number of hidden layers	[2-20]
	Number of convergence tries	[1-25]
	Maximum iterations	[5000-10000]
SDT	Minimum rows in a node	[3-22]
	Minimum size node to split	[5-20]
	Maximum tree levels	[10- 20]

4. RESULTS AND DISCUSSION

Table 3 through Table 5 show the validation results (i.e. the values of *RMSE* and *R*) for all hamstring muscle strength prediction models.

The results reveal that, in general, among the thirty prediction models developed, prediction models including predictor variables gender and weight lead to relatively lower *RMSE*'s and higher *R*'s whereas prediction models including age and sport branch lead to relatively higher *RMSE*'s and lower *R*'s for prediction of hamstring muscle strength. On the other hand, the results suggest that the category of prediction models with four predictor variables on the average shows the best prediction performance, whilst the category of prediction models with one predictor variable exhibits the worst prediction performance. The average *RMSE*'s of SVM-based prediction models for each category are illustrated in Figure 1.

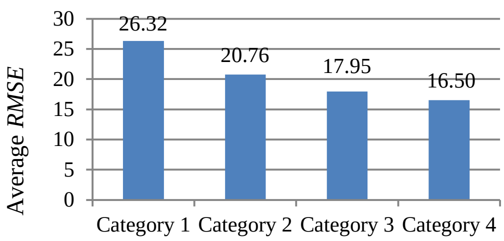

Figure 1. The average *RMSE*'s of SVM-based prediction models for each category

Table 3. Validation results for prediction models with two predictor variables

Predictor Variables	**SVM**		**MLP**		**SDT**	
	RMSE	***R***	***RMSE***	***R***	***RMSE***	***R***
Gender, Weight	15.31	0.82	16.29	0.79	18.88	0.72
Gender, Height	17.26	0.77	18.49	0.73	23.89	0.56
Gender, Sport Branch	19.05	0.72	20.81	0.66	24.75	0.52
Gender, Age	18.69	0.73	19.79	0.70	24.95	0.52
Sport Branch, Weight	18.70	0.73	19.78	0.70	22.42	0.61
Height, Weight	19.32	0.71	20.94	0.66	25.40	0.50
Weight, Age	20.05	0.69	21.34	0.65	23.70	0.56
Sport Branch, Height	22.93	0.59	24.84	0.52	28.53	0.37
Height, Age	25.68	0.49	27.44	0.41	30.95	0.25
Sport Branch, Age	30.62	0.27	32.64	0.17	37.14	0

Table 4. Validation results for prediction models with three predictor variables

Predictor Variables	**SVM**		**MLP**		**SDT**	
	RMSE	***R***	***RMSE***	***R***	***RMSE***	***R***
Gender, Weight, Age	15.48	0.81	16.81	0.78	20.15	0.68
Gender, Height, Weight	15.36	0.82	16.53	0.79	18.93	0.72
Gender, Sport Branch, Weight	15.68	0.81	17.39	0.76	22.41	0.61
Gender, Sport Branch, Height	17.33	0.77	19.03	0.72	25.70	0.49
Gender, Height, Age	16.78	0.78	17.97	0.75	20.03	0.69
Gender, Sport Branch, Age	18.99	0.72	20.47	0.67	23.93	0.55
Sport Branch, Height, Weight	17.89	0.75	19.50	0.70	24.07	0.55
Sport Branch, Weight, Age	19.39	0.71	21.52	0.64	25.44	0.50
Height, Weight, Age	19.78	0.70	20.90	0.66	25.57	0.49
Sport Branch, Height, Age	22.88	0.59	24.26	0.54	27.22	0.42

Table 5. Validation results for prediction models with four predictor variables

Predictor Variables	**SVM**		**MLP**		**SDT**	
	RMSE	***R***	***RMSE***	***R***	***RMSE***	***R***
Gender, Height, Weight, Age	15.19	0.82	16.29	0.79	18.93	0.72
Gender, Sport Branch, Height, Weight	15.84	0.80	17.09	0.77	19.59	0.70
Gender, Sport Branch, Weight, Age	15.89	0.80	17.66	0.76	20.62	0.67
Gender, Sport Branch, Height, Age	17.62	0.76	18.52	0.73	23.93	0.55
Sport Branch, Height, Weight, Age	17.97	0.75	19.83	0.69	24.96	0.52

In more detail, independent of which regression method is utilized for model development, the results reveal that:

- Firstly, within the category including a single predictor variable, prediction model with variable gender yields the lowest *RMSE* and the highest *R*, whilst the one with variable age shows the worst performance in terms of *RMSE* and *R*.
- Secondly, the prediction model with variables gender and weight, and the prediction model with variables age and sport branch occupy the first and last place, respectively, in terms of the performance regarding the *RMSE* and *R* for the category with two predictor variables.
- Thirdly, within the category with three predictor variables, prediction model with variables gender, weight and height

outperforms the others whereas the prediction model including sport branch, height and age yields the worst performance.

- Finally, among the category with four predictor variables, the prediction model with variables gender, weight, height and age yields the lowest *RMSE* and highest *R* whilst that one with variables sport branch, height, weight and age yields the highest *RMSE* and the lowest *R*.

Regarding the performance of prediction methods, it is seen that SVM-based prediction models give the lowest *RMSE*'s and the highest *R*'s for prediction of hamstring muscle strength for all thirty prediction models developed. SDT-based prediction models, in contrast, yield the highest *RMSE*'s and the lowest *R*'s, independent of which category of prediction models has been evaluated.

Compared with the *RMSE*'s obtained by MLP-based and SDT-based prediction models, the percentage decrement rates in *RMSE*'s obtained by SVM-based prediction models are in average 7.39% and 20.98%, respectively, as illustrated in Figure 2. The ranking of regression methods in terms of leading from lowest *RMSE*'s to highest ones can be listed as SVM, MLP and SDT; independent of which category of prediction models is considered. The execution times of SVM-based prediction models vary between 1 and 5 seconds. The MLP-based prediction models have execution times ranging from 1 to 2 seconds. Finally, the SDT-based prediction models have execution times less than one second.

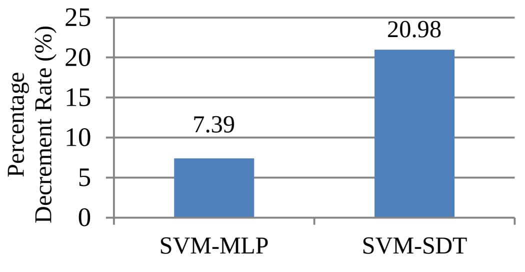

Figure 2. Percentage decrease rates in *RMSE*'s of hamstring muscle strength prediction with SVM compared to *RMSE*'s obtained by MLP and SDT

5. CONCLUSION AND FUTURE WORK

In this study, thirty new prediction models have been developed by using the single, double, triple and quadruple combinations of predictor variables gender, age, height, weight and sport branch for prediction of hamstring muscle strength of the upper leg. Using 10-fold cross validation, the performance of the prediction models are assessed by calculating performance metrics such as *RMSE* and *R*. The results reveal that among the prediction models, the SVM-based prediction model containing the variables gender, age, height and weight gives the lowest *RMSE* and the highest *R*, and can be used as an alternative way to direct measurement. Also, it is observed that all SVM-based prediction models yield lower *RMSE*'s and higher *R*'s than MLP-based and SDT-based models for prediction of hamstring muscle strength. MLP-based and SDT-based prediction models, in turn, occupy the second and the third place in terms of performance regarding the *RMSE* and *R*, respectively. Hence, SVM-based prediction models can be considered as a feasible alternative way to the direct hamstring muscle strength measurement.

For future work, other candidate potential predictors of hamstring muscle strength such as the length and width of the bone and leg fat-free mass can be included in prediction models to test whether a further improvement in prediction accuracies can be achieved. Also, new unapplied machine learning algorithms combined with various feature selection algorithms can be utilized to determine the relevant and irrelevant predictors of hamstring muscle strength.

6. REFERENCES

1. Dervišević, E., Hadžić, V., 2012. Quadriceps and Hamstrings Strength in Team Sports: Basketball, Football and Volleyball, Isokinetics and Exercise Science, vol. 20, no. 4, pp. 293–300.

2. Ford-Smith, C.D., Wyman, J.F., Elswick, R.K., Fernandez, T., 2001. Reliability of Stationary

Dynamometer Muscle Strength Testing in Community-dwelling Older Adults, Archives of Physical Medicine and Rehabilitation, vol. 82, no. 8, pp. 1128–1132.

3. Clarke, H.H., 2013. Comparison of Instruments for Recording Muscle Strength, Research Quarterly, American Association for Health, Physical Education and Recreation, vol. 25, no. 4, pp. 398–411.

4. Montgomery, L.C., Douglass, L.W., Deuster, P.A., 1989. Reliability of an Isokinetic Test of Muscle Strength and Endurance, The Journal of Orthopaedic and Sports Physical Therapy, vol. 10, no. 8, pp. 315–322.

5. Kılınç, B.E., Kara, A., Camur, S., Oc, Y., Celik, H., 2015. Isokinetic Dynamometer Evaluation of the Effects of Early Thigh Diameter Difference on Thigh Muscle Strength in Patients Undergoing Anterior Cruciate Ligament Reconstruction with Hamstring Tendon Graft, Journal of Exercise Rehabilitation, vol. 11, no. 2, pp. 95–100.

6. Akay, M.F., Abut, F., Çetin, E., Yarım, İ., Sow, B., 2015. Data-Driven Modeling of Quadriceps and Hamstring Muscle Strength Using Support Vector Machines, in Third International Symposium on Engineering, Artificial Intelligence & Applications, pp. 2–4.

7. Abut, F., Akay, M.F., 2015. Machine Learning and Statistical Methods for the Prediction of Maximal Oxygen Uptake: Recent Advances, Medical devices, vol. 8, pp. 369–379.

8. Akay, M.F., Abut, F., Özçiloğlu, M., Heil, D., 2016. Identifying the Discriminative Predictors of Upper Body Power of Cross-country Skiers using Support Vector Machines Combined with Feature Selection, Neural Computing and Applications, vol. 27, no. 6, pp. 1785–1796.

9. Acikkar, M., Akay, M.F., Ozgunen, K.T., Aydin, K., Kurdak, S.S., 2009. Support Vector Machines for Aerobic Fitness Prediction of Athletes, Expert Systems with Applications, vol. 36, no. 2, pp. 3596–3602.

10. Hsu, C.W., Chang, C.C., Lin, C.J., 2003. A Practical Guide to Support Vector Classification, Dept. of Computer Science, National Taiwan University, Taipei, 2003.

11. Hiregoudar, S.B., Manjunath, K., Patil, K.S., 2014. A Survey: Research Summary on Neural Networks, International Journal of Research in Engineering and Technology, vol. 3, no. 15, pp. 385–389.

12. Banfield, R.E., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P., 2007. A Comparison of Decision Tree Ensemble Creation Techniques, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, pp. 173–180.