A Classification System to Assess Low Back Muscle Endurance and Activity Using mHealth Technologies

Ignacio Diaz-Reyes[1](✉), Miguel Damas[1], Jose Antonio Moral-Munoz[2], and Oresti Banos[3]

[1] Department of Computer Architecture and Computer Technology, University of Granada, Granada, Spain
idiaz@mdurance.es, mdamas@ugr.es

[2] Department of Nursing and Physiotherapy, University of Cadiz, Cadiz, Spain
joseantonio.moral@uca.es

[3] Telemedicine Group, Center for Telematics and Information Technology, University of Twente, Enschede, Netherlands
o.banoslegran@utwente.nl

Abstract. Low back pain remains a major cause of absenteeism in the world. In addition to its socio-economic impact, the age at which the first symptoms appear is decreasing. Consequently, there are more experts who start incorporating prevention plans for the lumbar area in their work routines. In addition, the continued market growth of wearable sensors and the potential opened up by wearable technology allows experts to obtain a precise feedback from improvements in their patients in a daily basis. For this reason, this work wants to continue with the development and verification of the usefulness of mDurance, a novel mobile health system aimed at supporting specialists in the functional assessment of trunk endurance and muscle activity by using wearable and mobile devices. This work presents an extension of this system to classify low back muscle activity in the low back. mDurance has been tested into a professional football team. Clustering and data mining are applied in a new dataset of endurance and muscle activity data collected through mDurance. In addition, these results are cross-related with a questionnaire created to evaluate how the football players perceive themselves physically and mentally. The results show a clear correlation between the perception participants have about their low back endurance and the objective measurements conducted through mDurance. The results obtained through mDurance and the football players answers show a 68.3% of accuracy and 83.8% of specificity in the first approach to build a classifier to assess low back muscle endurance and activity using mDurance system.

1 Introduction

Low back pain (LBP) is considered an extremely common health problem, and the major cause of activity limitation and work absence in the world. In addition,

I. Rojas et al. (Eds.): IWANN 2017, Part II, LNCS 10306, pp. 709–721, 2017.
DOI: 10.1007/978-3-319-59147-6_60

it continues to decrease the age at which the first signs of fatigue or discomfort in the low back begins to appear [1]. Thus, experts are beginning to make plans to improve the endurance and to discharge these important muscles in daily activity. According to Kim [2], electromyography (EMG) is based on the study of muscle activity through observation and analysis of the electrical signals monitored during voluntary or involuntary muscle contractions. EMG is useful to study muscular function during sport activities, biofeedback training, daily living or detect pathological states of the musculoskeletal systems. In order to assess all those functions, EMG offers useful information about timing of muscular activity and its relative intensity. Fernandez [3] reports that muscle fatigue is manifested as a reduction in the ability to keep a certain level of strength in a sustained contraction or as the inability to achieve a level of initial strength in intermittent contractions and it is accompanied by changes in muscle electrical activity. Muscle fatigue has central and peripheral components. The first are manifested as an inability to realize an induced activity for the development of this activity. The second stands out as a muscle inability to produce a certain level of strength. In applications of electrical stimulation it is important to assess the development of muscle fatigue to prevent deterioration of the mechanical behavior of the muscle being stimulated.

Muscle fatigue can be evaluated by EMG signals. The effect of fatigue on muscle mechanics is well represented by various metrics, such as the mean frequency (MNF), median frequency (MDF), the root mean square (RMS), the average rectified value (ARV) and the maximum voluntary muscle contraction (MVC) [2,4,5]. This information is very interesting to compare the evolution of the muscle strength among sessions, as well as to measure the effectiveness of potential treatments. In the time domain the RMS of the EMG signal is considered the most reliable parameter. An increase of the RMS with advancing fatigue has been reported in many studies [6,7]. RMS is not affected by the cancellation due to motor unit action potential train superposition, which may affect other processing techniques involving rectification [7]. In [2] it is shown that muscle fatigue is accompanied with an increase in the RMS and ARV and a decrease in the MNF.

On the other hand, the design and implementation of wearable EMG systems for health monitoring has got lot of attention throughout the world, specially in the sport, the physiotherapy and scientific community during the last years. These wearable sensors allow experts to obtain a precise feedback from improvements in their patients daily [8]. The studies that use surface EMG in sport sciences are mostly related to determination of the mechanism of contraction and relaxation of muscles while also dealing with evolution of injuries. The data obtained from these studies can be used in the following areas [9]: the evaluation of the technical development, the establishment of the suitable exercise programs and the follow up of the development of the athletes. Techniques that take full advantage of the extraordinary amount of data that such sensors/systems can gather are lacking. Several analysis provide a very large data volume coming from EMG register and physical examinations. The analysis and treatment of these data is difficult and time consuming.

For this reason, mDurance was presented as a novel mobile health system aimed at supporting specialists in the functional assessment of trunk endurance and muscle activity by using wearable and mobile devices [10,11]. The present study aims to analyze the endurance and the muscle activity of the low back during an endurance muscle test execution by mDurance. This work studies the effects of the muscle fatigue of a football team. In addition, this study is intended to be the starting point for the development of a new classification system of low back muscle activity, by applying clustering and data mining in the new data generated by mDurance.

The design of this new classification system can be applied in several investigations, including the functional evaluation of muscular processes and verification of the response of rehabilitation therapy in sports medicine, occupational medicine and physical medicine and rehabilitation, the study and ergonomic analysis of the workload and the prevention of muscular fatigue in activities of human occupation. The rest of the paper is structured as follows. Section 2 presents a description of the main features of the study. The fundamental results and discussion about this work are outlined in Sect. 3. Final conclusions and remarks are summarized in Sect. 4.

2 Methodology

2.1 Subjects

Fifteen professional soccer players from 19 to 32 years old (mean $\pm$ SD; age 24.05 ± 3.32 years old; height 178.07 ± 5.21 cm; weight 74.39 ± 5.19 kg) were recruited to be evaluated by one external physical therapist using mDurance. Before performing the evaluation, the volunteers were informed about the research aims, risks and benefits of participation. All subjects were tested during the 2016/2017 Spanish competitive soccer preseason.

2.2 Instrumentation

For this work, mDurance was used to measure the low back endurance and muscle activity. The mDurance system consists of a wearable inertial sensor to track the patient trunk posture and a portable electromyography sensor to seamlessly measure the electrical activity produced by the trunk muscles. All the information registered through these sensors is intelligently managed by a mobile or tablet application by mHealthDroid [12]. In addition, all data is stored in a new storage system in the cloud, formed by a back-end service and API Rest application. The back-end platform is here used for data engine can apply data mining techniques and applications.

2.3 Test Procedure

Different tests are available to assess the trunk endurance in people with or without LBP. This kind of tests are performed by a specialist, and they normally

consist of the measurement of the time a person can hold a specific posture involving the trunk muscles. During the execution of the test, the health professional has to control the patient position, and with the help of mDurance, the specialist decides when the test ends, according to some established termination criteria. The results obtained for a given patient help experts determine their status and muscular capacity, as well as their ability to hold a posture normally related to daily living activities. To assess the low back stabilization, mDurance uses a functional trunk endurance test widely used: the static trunk extensor endurance test (STEET), also known as Sorensen test [13] (see Fig. 1).

Fig. 1. STEET procedure using mDurance.

In the STEET, the subject has to maintain a horizontal unsupported posture with the upper body extending beyond the edge of the bench. Special remarks are that two chances are given to the individual to execute the STEET. The position is held up to a maximum of 240 s. A detailed description of this test, including posture, procedure and finalization criteria, is available in our previous work [10].

The STEET test was executed once per week to minimize the fatigue effects on the performance of the volunteers. The test was realized both before and after the training sessions up to a maximum of three times, in order to get a higher variety in the results for reducing the muscle fatigue intervals of each player. In both cases, each player filled up a questionnaire (see Fig. 2) every time that he realized the test. This questionnaire was created and approved by the physical therapists who evaluated the players. The main goal is to evaluate how the players are perceived themselves physically and psychologically and to check which is their condition and to improve their performance. This evaluation is

carried out by comparing the responses obtained from the questionnaires filled in by the players and the results obtained through mDurance for each test. The complete procedure was explained to the subjects before performing the sessions, assuring the full understanding of their phases.

Name _____

Date _____

Have you played today?
a) No
b) Yes

Which is your condition today?
a) Ok
b) Fatigue
c) Discomfort
d) Injury

ID		YES	NO
Q1	Do you feel low back pain at this moment?		
Q2	Have you performed any activity that may have caused you low back fatigue?		
Q3	How many hours do you think you spent sitting today?		
Q4	Do you feel rested at this moment?		

Fig. 2. Part of the questionnaire used for the physical and mental player assessment.

2.4 Dataset Analysis

The dataset consists of the final results obtained through mDurance and the results obtained through the questionnaires. This dataset represents how the endurance and muscle fatigue affect the soccer players performance according to several features. They are divided into two blocks: Questionnaire and mDurance results.

Questionnaire results:

– *Player's condition.* This represents the health condition of each player in the exact moment of realizing the test. The condition can take the following values: "Ok", "Fatigue", "Discomfort", "Injury" depending on the player was in perfect condition, or he had done any diary activity which could cause muscle fatigue, or the player suffered some low back pain during the test or the player was injured in that date.

- *Has_Played*. This indicates if the players who were going to do the test had played or not. *Q1, Q2 and Q4* represent the different yes/no questions in the questionnaire. They describe both physical and mental status of the players for a particular date. *Q3* describes the number of hours that each player was sitting that day.

mDurance results:

- *Total time*. It indicates the endurance total time in second, which the player lasted in the STEET test.
- *Ratio*. It refers to a value calculated by the total time and an average value. It can be consulted in our previous work [10].
- *Ratio_label*. It has two levels: "Up" and "Down": The values are classified according to if the ratio is higher than 1 or lower than 1. In the first case, it means that the player is above the average and he is below in the opposite case.
- *RMS, ARV and MVC*. They describe the absolute muscle fatigue values measured in the end of the test by the application. They are expressed in mV.
- *RMS_per_second, ARV_per_second*. Another way to compare the muscle activity among the players is to create some new attributes, which allow to check that players had a major muscle activity. Thus, these variables are calculated dividing the final RMS and ARV values over the total time obtained in the test. They are expressed in mV/s.
- *Endurance label*. It describes the three possible endurance states that are registered by the mDurance application. They are: "bad shape", "good shape", "excellent shape" according to the time duration in the test.
- *Fatigue label*. One of the difficulties that this work has encountered is to get to categorize the results of muscle activity obtained by the application. We could not find any bibliography which allowed to label the muscle activity values. For this reason, we classify these results in three levels: "minimum fatigue", "medium fatigue" and "maximum fatigue". Each fatigue level is used to represent the minimum, medium and maximum value obtained in the three tests executed for each football player. For example, if a football player got 0.4, 0.5 and 0.6 mV for each RMS result obtained for each test, the first value would be classified as "Minimum fatigue", the second value as "Medium fatigue" and the third value as "Maximum fatigue".
- *Fatigue per second label*. Finally, it is also decided to create a new EMG range to categorize the results obtained by the RMS and ARV per second, in order to classify every player based on the same unit. The values are classified again in three levels: "minimum", "medium" and "maximum" fatigue per second.

3 Results and Discussion

3.1 Muscle Activity Analysis

Most important attribute to detect the muscle fatigue is the RMS, so it is plotted with respect to the endurance and fatigue labels. Figure (see Fig. 3) shows that

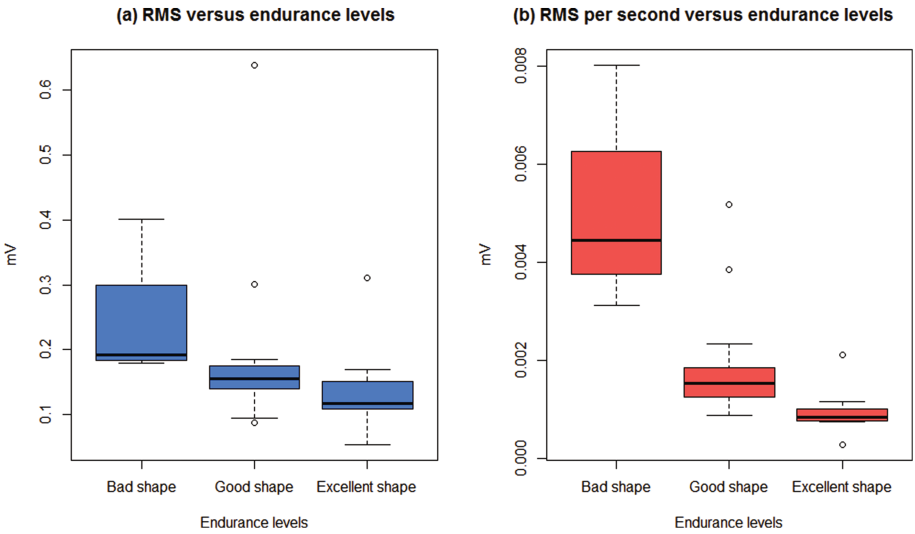

Fig. 3. Endurance levels with respect to: (left) RMS values and (right) RMS per second values.

RMS values are greater when the players endurance is worse. It indicates that many of the players who realized the tests and that lasted a short time obtained a higher RMS than the players who lasted more longer value. Thus, this is a clear indicator of the existence of muscle fatigue in these players. In addition, it can be observed that between the "good" and the "excellent" time classification, most of tests which had an "excellent" time classification obtained the lower RMS values. This supports the previous conclusion. In the figure (see Fig. 4) is compared how the RMS and RMS per second values are distributed for the three fatigue classifiers. It shows that the maximum, mean and minimum values of the "Maximum" group are bigger than these same values in the "Medium" and "Minimum" group. In addition, "Medium" values are bigger than these in the "Minimum" group. However, the "Medium" group takes values both of the "Maximum" group for the RMS values and the "Maximum" group for the RMS/s values. This can make difficult the supervised classification task.

3.2 Clustering Analysis

In the clustering theory, the first is to select an approximation functions. The choice of function determines the future results and the clustering quality. Moreover, it is important to select the attributes that are going to participate in the clustering process. There are two very differentiable variables sets: the quantitative attributes describing the final results obtained for each test and the qualitative attributes, which represent both the questionnaire results as the endurance and muscle activity. The most important variables are selected through Chi-squared methods [14] and the correlation feature selection method (CFS) [15]

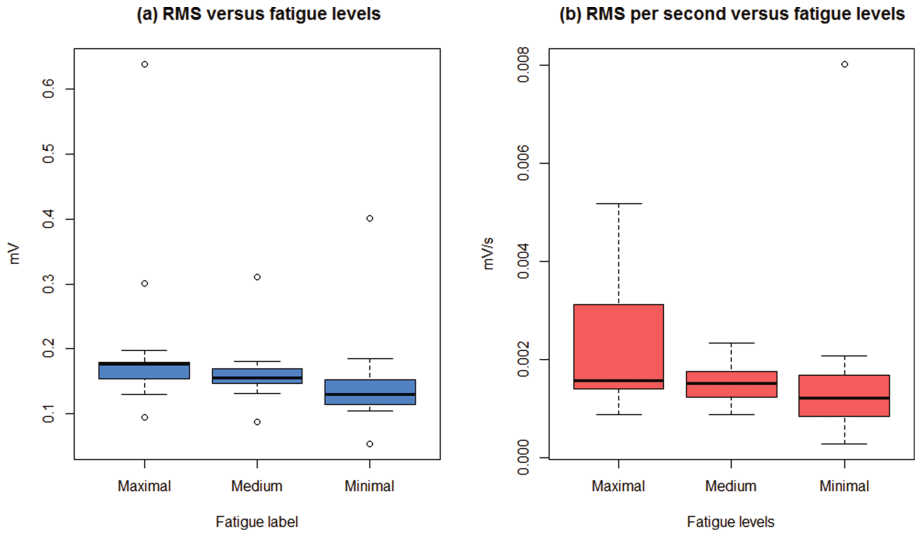

Fig. 4. Muscle activity levels with respect to: (left) RMS values and (right) RMS per second values.

were used. Finally, the selected attributes used in clustering were: RMS, RMS_sec representing to the quantitative variables and Condition, Q1, Q4, Ratio_label, Endurance_label and Fatigue_perSecond_label.

The hierarchical clustering approach is a method of group analysis which seeks to build a hierarchy of groups. Most algorithms are of the agglomerative type, i.e., partitions start with as many groups as items and at each step are united together. This approach could give a global idea about how many groups should be considered. The Ward method was selected [16], because it offered the best results. Thus, the distance matrix of the data was calculated, because these algorithms are based on making transformations about this matrix decreasing its size. To select the best group, it is necessary to use the silhouette coefficient of the possible partitions. The silhouette coefficient [17] is a measure that combines the cohesion and separation. It compares the distance means among instances of a same group and the distance means among instances of different groups. The results obtained are showed in Table 1.

When two groups are only considered (K = 2), the cohesion and the separation are the largest with respect to three (K = 3) and four (K = 4) groups. However, this happens because the second group has few units with respect to the first group. Thus, this group is not a good option. If the groups are compared with three and four groups respectively, it is observed that the Silhouette coefficients are very similar (0.3418, 0.3456). However, it is not worth to take four groups because one of these groups just has one item. Thus, it was decided to take the three groups option.

Table 1. Units per group and Silhouette coefficient for all the possible groups in all the possible pruning (k = 2, 3 and 4).

	Size	Ind. Silhouette width	Mean Silhouette width
K = 2			
Group 1	37	0.6022	0.5825
Group 2	4	0.3999	
K = 3			
Group 1	16	0.2648	0.3418
Group 2	21	0.4036	
Group 3	4	0.3251	
K = 4			
Group 1	16	0.2648	0.3456
Group 2	21	0.4023	
Group 3	1	0.0000	
Group 4	3	0.4932	

Figure 5 shows that players of the second group that have a major RMS value per second will have a major probability of being classified into "Maximum" or "Medium" muscle activity. In addition, all players labeled with the worst endurance needed more muscle activity to maintain the same load as others labeled with better endurance. The 71.42% the players classified with the best endurance needed the minimum muscle activity. Another aspect to note is the clustering of these instances according to the endurance labels. In Fig. 5, it is also demonstrated the accuracy of the classification used by mDurance to label the RMS value in function of the player endurance. It is seen that the worst endurance results obtained in the tests correspond with the instances of the third group, labeled as "Maximum muscle activity" instances. This aspect is extended to the second and first group, where most of the instances are grouped as good and excellent shape and "Minimum muscle activity".

3.3 Classification Supervised Analysis

Once it has been able to determine the number of groups where to classify any new RMS value obtained in the execution of a new STEET test, it is time to build a classifier and evaluate its performance. The supervised classification allows classifying a new instance in a certain groups of the considered ones.

Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. Leave-p-out cross-validation (LpO CV) involves using p observations as the validation set and the remaining observations as the training set [18]. This is repeated to split the original sample into a validation set of p observations and a training set. In this work, it is used Leave-one-out cross-validation (LOOCV) because there are not a

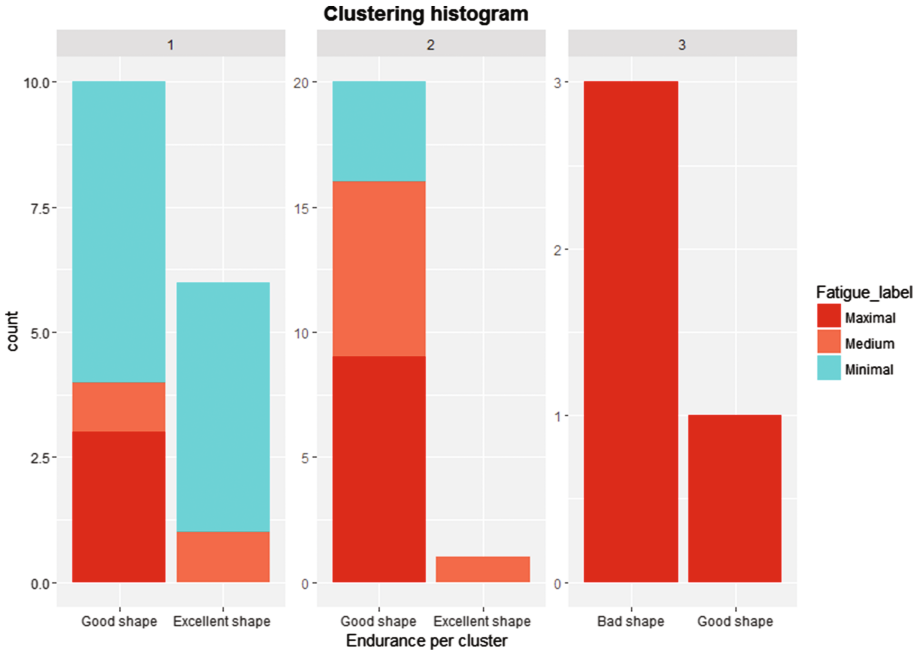

Fig. 5. Representation of the three groups obtained by Ward methods in histograms. The groups are represented by the results obtained through mDurance and the questionnaires.

large number of samples. This is a particular case of leave-p-out cross-validation with p = 1. Classification algorithms can be grouped into two large families attending the goodness of the results. *Individual classifiers* are learning a single model, which can be of various kinds: KNN [19], SVM Linear [20] and SVM Radial [21]. *Ensemble classifiers* are algorithms that combine several individual classifiers, decision trees normally. For this study was used classification trees (CTree) [22].

In Table 2, these groups are compared on equal terms. The same variables used for clustering are used in this section.

Table 2. Comparison between the main classifiers mentioned above.

Method	Accuracy	Mean bal. accuracy	Mean sensitivity	Mean specificity
KNN (k = 9)	0.671	0.699	0.652	0.826
SVM linear	0.658	0.733	0.640	0.825
SVM radial	**0.683**	**0.756**	**0.674**	**0.838**
CTree	0.610	0.694	0.586	0.801

It may be concluded that the individual classifiers clearly outweigh the ensembles classifiers, and SVM Radial seems to be the most appropriate. In this paper, with 68.3% accuracy, it is shown that a classifier could be constructed taking into account the results obtained through mDurance and the answers obtained through a simple questionnaire. In addition, this could increase until a 75.6% if it is measured the mean of all the accuracies obtained for each fold in LOOCV. It should be noted that the mean specificity, which identifies the percentage of non-fatigued people who are correctly identified is the best result obtained in this study with a 83.8%. On the contrary, the mean sensitivity, which measures the percentage of fatigued people who are correctly identified as having is the lowest result.

4 Conclusion

It has been shown a study about the behavior of the muscles of the lumbar area on a professional football team. This work intends to be the starting point for the development of a new classification system of muscle activity in the low back muscle, by applying clustering and data mining in the new data collected through mDurance results and the answers to the questionnaire (see Fig. 2). This questionnaire aims to evaluate how the players are perceived themselves physically and mentally, and to check what are their conditions to can improve their performance. The first groups of what could be a new unsupervised classification system of the lumbar muscle activity have been grouped. The results claim that groups correspond to the values of minimum-medium and medium-maximum muscle activity, further demonstrating that muscle activity shown by the players was in most cases consistent with their perception. Once it was possible to group instances into three groups, they were applied the most famous supervised classification techniques, both individual and ensembles, in order to test the accuracy of the classification of new instances. It must be concluded that the highest accuracy, specificity and sensibility were obtained with a single classifier, such as SVM Radial, which obtained a 68.3%, 67.4% and 83.8% in precision and sensitivity and specificity respectively.

The design of this new classification system can be applied in several investigations: functional evaluation and verification of the rehabilitation therapies and the prevention of muscular fatigue in activities of human occupation.

Acknowledgements. This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) Projects TIN2015-71873-R and TIN2015-67020-P together with the European Fund for Regional Development (FEDER). The authors want to especially thank all of the volunteers who participated in the experiments and Cristian Rivera Peregrina for your help.

References

1. Woolf, A.D., Pfleger, B.: Burden of major musculoskeletal conditions. Bull. World Health Organ. **81**(9), 646–656 (2003)

2. Kim, G., Ahad, M.A., Ferdjallah, M., Harris, G.F.: Correlation of muscle fatigue indices between intramuscular and surface EMG signals. In: Proceedings SouthEastCon, pp. 378–382, IEEE (2007)
3. Fernández, J.M., Acevedo, R.C., Tabernig, C.B.; Influencia de la fatiga muscular en la señal electromiográfica de músculos estimulados eléctricamente. Revista EIA (7), 111–119 (2007)
4. Phinyomark, A., Phukpattaranont, P., Limsakul, C.: Feature reduction and selection for EMG signal classification. Expert Syst. Appl. **39**(8), 7420–7431 (2012)
5. Phinyomark, A., Thongpanja, S., Hu, H., Phukpattaranont, P., Limsakul, C.: The usefulness of mean and median frequencies in electromyography analysis, pp. 195. INTECH Open Access Publisher (2012)
6. Siemienski, A., Kebel, A., Klajner, P.: Fatigue independent amplitude-frequency correlations in EMG signals. arXiv preprint physics/0611120 (2006)
7. Al Zaman, A., Sharmin, T., Mohammad Ashraf Ali Khan, and Ferdjallah, M.: Muscle fatigue analysis in young adults at different MVC levels using EMG metrics. In: Proceedings SoutheastCon, pp. 390–394. IEEE (2007)
8. Pantelopoulos, A., Bourbakis, N.G.: A survey on wearable sensor-based systems for health monitoring, prognosis. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **40**(1), 1–12 (2010)
9. Türker, H., Sözen, H.: Surface electromyography in sports and exercise. In: Electrodiagnosis in New Frontiers Clinical Research, p. 181 (2013)
10. Banos, O., Moral-Munoz, J.A., Diaz-Reyes, I., Arroyo-Morales, M., Damas, M., Herrera-Viedma, E., Hong, C.S., Lee, S., Pomares, H., Rojas, I., et al.: mDurance: a novel mobile health system to support trunk endurance assessment. Sensors **15**(6), 13159–13183 (2015)
11. Banos, O., Bang, J.H., Kang, D.U., Hong, C.S., Lee, S., Munoz, J.A.M., Damas, M., Reyes, I.D., Pomares, H., Rojas, I., Villalonga, C., Morales, M.A.: Facilitating trunk endurance assessment by means of mobile health technologies. In; ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp, pp. 149–152. ACM (2015)
12. Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado-Terriza, J.A., Lee, S., Pomares, H., Rojas, I.: Design, implementation and validation of a novel open framework for agile development of mobile health applications. Biomed. Eng. Online **14**(2), S6 (2015)
13. Biering-Sorensen, F.: Physical measurements as risk indicators for low-back trouble over a one-year period. Spine **9**(2), 106–119 (1984)
14. Higgins, J.E., Koch, G.G.: Variable selection and generalized chi-square analysis of categorical data applied to a large cross-sectional occupational health survey. Int. Stat. Rev./Revue Internationale de Statistique **45**, 38–51 (1977)
15. Hall, M.A.: Correlation-based feature selection for discrete and numeric class machine learning. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 359–366. Morgan Kaufmann Publishers Inc. (2000)
16. Batagelj, V.: Generalized ward and related clustering problems. In: Classification and Related Methods of Data Analysis, pp. 67–74 (1988)
17. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
18. Arlot, S., Celisse, A., et al.: A survey of cross-validation procedures for model selection. Stat. Surv. **4**, 40–79 (2010)
19. Murugappan, M.: Electromyogram signal based human emotion classification using KNN and LDA. In: IEEE International Conference on System Engineering and Technology (ICSET), pp. 106–110 (2011)

20. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)
21. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge (2000)
22. Kaur, A., Sharma, S.: Human activity recognition using ensemble modelling. In: Unal, A., Nayak, M., Mishra, D.K., Singh, D., Joshi, A. (eds.) SmartCom 2016. CCIS, vol. 628, pp. 292–298. Springer, Singapore (2016). doi:10.1007/978-981-10-3433-6_35