Classification of sEMG Signals for Muscle Fatigue Detection Using Support Vector Machines

Fariba Biyouki*, Saeed Rahati**, Reza Boostani***, Katri Laimi****, Afsane Zadnia*****
*Dept. of Biomedical Engineering, Mashhad Branch, Islamic Azad University, Iran, FaribaBiyouki@mshdiau.ac.ir
**Dept. of Electrical Engineering, Mashhad Branch, Islamic Azad University, Iran, Rahati@mshdiau.ac.ir
***Assistant Professor of Neurology, Mashhad University of Medical Sciences, Iran, boostanir@mums.ac.ir
****The Rehabilitation Company of Social Insurance Institution, Turku, Finland, Katri.Laimi@tyks.fi
*****Dept. of Biomedical Engineering, Mashhad Branch, Islamic Azad University, Iran, AfsaneZadnia@mshdiau.ac.ir

Abstract: *Fatigue is a multidimensional and subjective concept and is a complex phenomenon including various causes, mechanisms and forms of manifestation. Thus, it is crucial to delineate the different levels and to quantify self- perceived fatigue. The aim of this study was to discriminate between fatigue and nonfatigue stages using support vector machine (SVM) approach. Thus, electromyographic (EMG) signals collected in the department of biomedical engineering of Islamic Azad university of Mashhad, were used. 10 features in time, frequency and time- scale domains were extracted from sEMG signals and the effect of different objective functions for dimensionality reduction and different σ values in RBF kernel SVM were evaluated for fatigue detection. The best accuracy (89.45%) was achieved through RBF kernel with σ=0.5 and ROC criterion while the best accuracy through linear SVM was 54.42%. These results suggest that the selected features contained some information that could be used by the nonlinear SVM with RBF kernel to best discriminate between fatigue and nonfatigue stages.*

Keywords: Surface Electromyography (sEMG), muscle fatigue, classification, Radial Basis Function (RBF) kernel, Support Vector Machines (SVM).

1. Introduction

Fatigue is a multidimensional and subjective concept and is a complex phenomenon including various causes, mechanisms and forms of manifestation, hence poses a complex problem for the physician [1,3,6]. Since fatigue has physiological and psychological dimensions, delineation of its different levels and quantification of self- perceived fatigue is crucial [1].

Fatigue definition is very complex, not unique and controversial [13] but in physiology, fatigue is usually defined as the loss of voluntary force- producing capacity during exercise. This can be due to both central and peripheral mechanisms [1]. Fatigue has mostly been studied at peripheral level, i.e. in the muscle tissue. During peripheral fatigue, the accumulation of lactate and extracellular potassium, together with a lowering of pH, affects membrane excitability [1]. Surface EMG signals provide useful information about the underlying mechanisms of fatigue [1,6]. In spite of the limitations of the application of sEMG method to muscles positioned directly below the skin and the problem of cross talk from neighbouring muscles, this method due to its non-invasiveness, applicability in situ, real- time monitoring of fatigue and correlation with biochemical and physiological changes of muscle during fatigue, is widely used to determine local muscle fatigue [6]. To reduce the difficulty of the problem and the number of factors affecting the EMG signal, most past researches focused on myoelectric manifestation of muscle fatigue during isometric, constant force conditions. It is clear that such easy to study conditions do not reflect the muscle function in daily life [13]. The myoelectric signal in dynamic conditions, in which the muscle force, length and position of body segments change, is a nonstationary signal [2,3,4]. Since fatigue itself is not a physical variable, its assessment requires the definition of indices based on physical variables that can be measured, such as force, power, or variables associated to the EMG signal, such as amplitude and spectral estimates [13]. The most widely used method for estimating the spectrum of the EMG signal is Fourier transform. Fourier methods suffer from several limitations. One of them is the stationarity assumption, otherwise information about spectral changes will be lost [3,4,5,6,13]. Therefore, the parameters commonly used as indicators of spectral changes (i.e. median and mean frequency) during dynamic contraction may not accurately reflect muscle fatigue [14]. Recent developments in time- scale analysis methods, have been suggested new EMG parameters to assess muscle fatigue and overcome the nonstationary condition [5,6].

During dynamic contraction, amplitude, timing and frequency of muscle events within EMG signals change. So, one should be able to use extracted features, specially time- scale features of sEMG signals as a vector to discriminate between fatigue and nonfatigue stages. A support vector machine (SVM) classifier is a powerful vector- based method that can be used for fatigue classification. In this study, the possibility of detecting two stages of muscle fatigue (nonfatigue and fatigue) during dynamic contraction was investigated. Thus, 10 features in time, frequency and time- scale domains were extracted from sEMG signals and then linear and nonlinear SVM were used to discriminate between different phases.

2. Methods

2.1 Data

Surface EMG data used in this study was taken from [7]. These signals were collected in the department of biomedical engineering of Islamic Azad university of Mashhad. In [7], sEMG signals from flexor digitorum profondus and trapezius muscles were recorded during typing with Power Lab 4/25T system (ADInstruments Pty Ltd. Australia) under the supervision of a physiotherapist. The subjects were 9 healthy girl students that all of them type with both hands and none of them feel weakness or fatigue in mentioned muscles before recording. In order to cover EMG frequency range and eliminate the low- frequency and high- frequency noise, the Power Lab filters were adjusted to 500 Hz and 8 Hz, respectively. Furthermore, in order to remove power line noise, notch filter centred at 50 Hz frequency was used. The raw EMG signals were passed through a 16- bit AD converter with a sampling frequency of 1000 Hz. This sampling rate is chosen regarding the largest frequency of EMG frequency content and Nyquist rate to avoid interference. Because all subjects report their fatigue in trapezius muscle, analyses for quantification of fatigue performed in this muscle [7].

2.2 Feature Extraction

In order to quantify muscle fatigue and discriminate between fatigue and nonfatigue classes using linear and nonlinear SVM, first sEMG signal was split into one second epochs and then 10 features in time, frequency and time- scale domains were extracted from each epoch. Then features were normalized to come into a suitable scale. Since the recorded signals were too long (minimum recording duration was 15 min), after extracting features from each epoch, the average value of features in each five epoch, considered as one sample. It should be noted that all analyses of sEMG data were performed with MATLAB R2010a software (Mathworks Inc, USA).

2.2.1 Time Domain Features

Root Mean Square (RMS) and Zero Crossing Rate (ZCR) are the features that can be extracted from EMG signal in time domain.

RMS: RMS of sEMG signal is indicative of firing frequency, duration and velocity of the myoelectric signal. The increment of this feature shows the recruitment of extra motor units to produce constant force and is an index of fatigue development.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \qquad (1)$$

Where x_i is the ith sample of a signal and N is the number of samples in each epoch [6].

ZCR: The Zero Crossing Rate (ZCR) or zero crossing frequency f_z of the signal x(t) is defined as half the number of zero crossings of x(t) per second. This feature indicates the number of baseline crossings of EMG signal. When the muscle activity increases, the more action potential will produce. But when fatigue started, this feature decreases because of the decrease in muscle fiber conduction.

$$ZCR = 2\left[\frac{\int_0^{\frac{f_s}{2}} f^2 X(f)df}{\int_0^{\frac{f_s}{2}} X(f)df}\right] \qquad (2)$$

Where X(f) is the power spectral density (PSD) of signal and f_s is the sampling frequency [6].

2.2.2 Frequency Domain Features

Median frequency (MDF), mean frequency (MNF) and dominant frequency (DF) are the features that extracted from sEMG signal in frequency domain.

MDF and MNF: These two features are the most commonly used frequency variables in EMG studies and have been extensively used to provide basic information about changes of power spectrum during time. MNF is the average frequency of the power spectrum and is defined as its first- order moment. MDF is the frequency at which the spectrum is divided into two parts of equal power and is defined as zero- order moments of PSD. These two variables, described by the following equations:

$$\mathrm{MNF} = \frac{\int_0^{\infty} \omega P(\omega)d\omega}{\int_0^{\infty} P(\omega)d\omega} \qquad (3)$$

$$\int_0^{MDF} P(\omega)d\omega = \int_{MDF}^{\infty} P(\omega)d\omega = \frac{1}{2}\int_0^{\infty} P(\omega)d\omega \qquad (4)$$

In both “Equation (3)” and “Equation (4)”, P(ω) is the PSD of EMG signal and ω is the frequency variable [8].

DF: This feature finds dominant frequency within a band of the selected frequency band (usually estimated using Welch’s method). For this feature a frequency band of 15 Hz to 45 Hz has been selected (since they repeated most frequently) [9].

2.2.3 Time- Scale Domain Features

During isometric contraction, EMG signal can be assumed stationary for short time intervals (0.5- 2 s). With this assumption, spectral analysis based on Fourier transform can be used. But for dynamic contractions, like the situation of the present study, this assumption is not true. Recently, time- scale methods (Wavelet transform) were proposed to overcome the limitations of the time-frequency methods. The time- scale methods do not require any stationarity assumption [3]. In this study, 5 different indices were calculated using discrete wavelet transform (DWT).

(a) Wavelet index of ratios between moment -1 at scale 5 and moment 5 at scale 1 (w_1) that can be described by "Equation (5)":

$$W_1 = \frac{\int_{f_1}^{f_2} f^{-1} D_5(f).df}{\int_{f_1}^{f_2} f^{5} D_1(f).df} \tag{5}$$

Where $D_5(f)$ and $D_1(f)$ are the power spectra calculated using Fourier transform of the fifth and first scales of the DWT using the wavelet sym5, respectively, and f_1= 8 Hz and f_2= 500 Hz [5].

(b) Wavelet index of ratio between moment -1 at maximum energy scale and moment 5 at scale 1 (w_2) that can be described by "Equation (6)":

$$W_2 = \frac{\int_{f_1}^{f_2} f^{-1} D_{\max}(f).df}{\int_{f_1}^{f_2} f^{5} D_1(f).df} \tag{6}$$

Where $D_{max}(f)$ and $D_1(f)$ are the power spectra calculated using Fourier transform of the maximum energy and first scales of the DWT using the wavelet sym5, respectively, and f_1= 8 Hz and f_2= 500 Hz [5].

(c) Wavelet index ratio between moment -1 at scale 5 and moment 2 at scale 2 (w_3) that can be described by "Equation (7)":

$$W_3 = \frac{\int_{f_1}^{f_2} f^{-1} D_5(f).df}{\int_{f_1}^{f_2} f^{2} D_2(f).df} \tag{7}$$

Where $D_5(f)$ and $D_2(f)$ are the power spectra calculated using Fourier transform of the fifth and second scales, respectively of the DWT using the wavelet db5, and f_1= 8 Hz and f_2= 500 Hz [5].

(d) Wavelet index of ratios of energies at scale 5 and 1 (w_4) that can be described by "Equation (8)":

$$W_4 = \frac{\sum_{i=1}^{N} D_5^2[n]}{\sum_{i=1}^{N} D_1^2[n]} \tag{8}$$

Where $D_5(f)$ and $D_1(f)$ are the details at scales five and one, respectively of the DWT calculated using wavelet sym5 [5].

(e) Wavelet index ratio between square waveform length at different scales (w_5):

$$W_5 = \frac{\sum_{i=2}^{N} \left| D_5[i] - D_5[i-1] \right|^2}{\sum_{i=2}^{N} \left| D_1[i] - D_1[i-1] \right|^2} \tag{9}$$

Where $D_5(f)$ and $D_1(f)$ are the details at scales five and one, respectively of the DWT calculated using wavelet sym5 [5].

2.3 Classification Using SVM

Application of support vector machine (SVM) for classification problems is a new approach which became popular in recent years. The SVM approach is that in the training phase, it tries to select the decision boundary so that its minimum distance from each of the classes become maximum. This selection caused decision to bear the noisy situations in practice and to have a good response. This boundary selection method is based on the points called support vectors [10].

The linear SVM problem is a two- class classification problem using linear models of the form g(x)= w^T. x + b. In practice, however, the class conditional distributions may overlap, in which case exact separation of the training data can lead to poor generalization [11]. Thus there is a need to modify SVM because the solution for linearly separable data is not applicable for nonlinearly separable case. One solution is to allow some of the training points to be misclassified. To do this, the slack variables, $\xi_i \geq 0$, are assigned to each training data point. Another solution is the use of nonlinear SVM. Nonlinear SVM operates in two stages: (1) nonlinear mapping of the feature vector onto a high dimensional space and (2) construct an optimal separating hyperplane in the high dimensional space. Since SVM is a supervised learning method, so it is necessary to label each epoch before the processing started. Labelling epochs into (1) fatigue and (2) nonfatigue classes performed based on self- report of subjects. In this study, linear and nonlinear SVM is used for EMG signal classification. To implement nonlinear SVM, the kernel function should be introduced. In this paper, radial basis function (RBF) kernel is used. RBF kernel can be defined as "Equation (10)":

$$\kappa(X_i, X_j) = \exp\left(-\frac{\left\| x_i - x_j \right\|^2}{2\sigma^2} \right) \tag{10}$$

3. Results

Since the dimensionality were high (10 features), first, it was tried to reduce the dimensions using feature selection method with filter strategy using different objective functions included t- test, entropy, receiver operating characteristic (ROC) and wilcoxon. TABLE I, shows the key features that selected through each criterion.

TABLE I: Key Features Selected by Different Objective Functions.

Criterion	t-test	entropy	ROC	Wilcoxon
Key features	DF W2	DF ZCR	RMS W2	W2 W4

After training SVM classifier using training dataset, the performance of the classifier is evaluated for test set and different criteria. TABLE II, shows the accuracy achieved by linear SVM and different criteria and Fig. 1 indicates the result of separation using entropy criterion and linear SVM.

TABLE II: The Accuracy Achieved by Linear SVM.

Criterion	t- test	entropy	ROC	wilcoxon
Accuracy	50.14%	54.42%	50.76%	51.41%

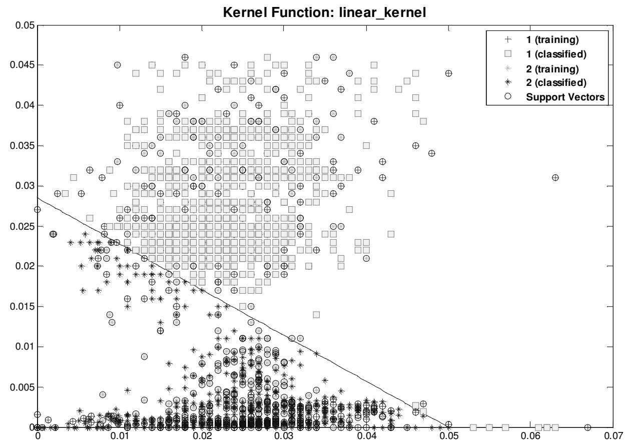

Fig. 1: Result of classification using linear kernel and entropy criterion.

TABLE III, indicates the accuracy achieved by RBF kernel with different criteria and different σ values and Fig. 2 shows the results of discrimination via RBF kernel and ROC criterion and σ= 0.5. As can be seen in TABLE II and TABLE III, the best accuracy for linear SVM was 54.42% with entropy criterion and the best accuracy for nonlinear SVM with RBF kernel was 89.45% with ROC criterion and σ= 0.5. Fig. 3 shows the effect of σ in decision surface and boundary and number of support vectors.

TABLE III: Accuracy Achieved by RBF Kernel for Different σ Values And Different Criteria.

Sigma / criteria	0.2	0.5	1	1.5
t- test	59.46%	58.53%	56.54%	55.81%
Entropy	71.71%	72.78%	71.85%	71.26%
ROC	-	89.45%	72.64%	66.65%
wilcoxon	70.89%	68.38%	66.20%	61.49%

4. Conclusion

Fatigue is a subjective concept and its definition is very complex, not unique and controversial. The detection and classification of muscle fatigue, provides

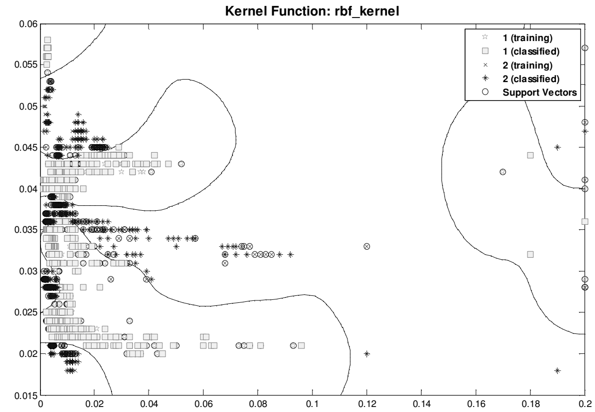

Fig. 2: Result of classification using RBF kernel and ROC criterion and σ= 0.5.

useful information in many research areas. For example, in the branch of ergonomics which deals with musculoskeletal disorders, muscle fatigue may be considered as a major risk factor. Fatigue detection and classification through biofeedback system, may contribute to an awareness of a sustained muscle activation patterns. Fatigue classification also can be applied to the fields of human- computer interactions, sport injuries and performance.

This study used linear and nonlinear SVM with RBF kernel to classify sEMG signal into fatigue and non fatigue classes. The kernel function or nonlinear mapping results in different kinds of support vector classifiers (SVCs) with different performance levels. But the choice of the appropriate kernel for a specific application is often a difficult task. If the data is known to be nonlinearly separable, we would expect that a nonlinear kernel based SVC would perform better than the one based on a linear kernel. Since the data in this study was nonlinearly separable, so it is reasonable to expect that the nonlinear SVM with RBF kernel have better performance than the linear one. The results of the classification accuracy shown in TABLE II and III, approve this claim (54.42% for linear SVM versus 89.45% for RBF kernel). In general, the RBF kernel is a reasonable first choice. Because, first, this kernel nonlinearly maps samples into a higher dimensional space so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of RBF, because the linear kernel with a penalty parameter C, has the same performance as the RBF kernel with some parameters (C,γ). The second reason is the less number of hyperparameters which reduces the complexity of model selection. Finally, the RBF kernel has fewer numerical difficulties. However, this kernel is difficult to design, in the sense that it is difficult to arrive an optimum σ. The fact that certain σ value make the SVM highly sensitive to training data also contributes to the error rate of the RBF- based SVM. A larger value of σ will give a smoother decision surface and more regular decision boundary. This is because an

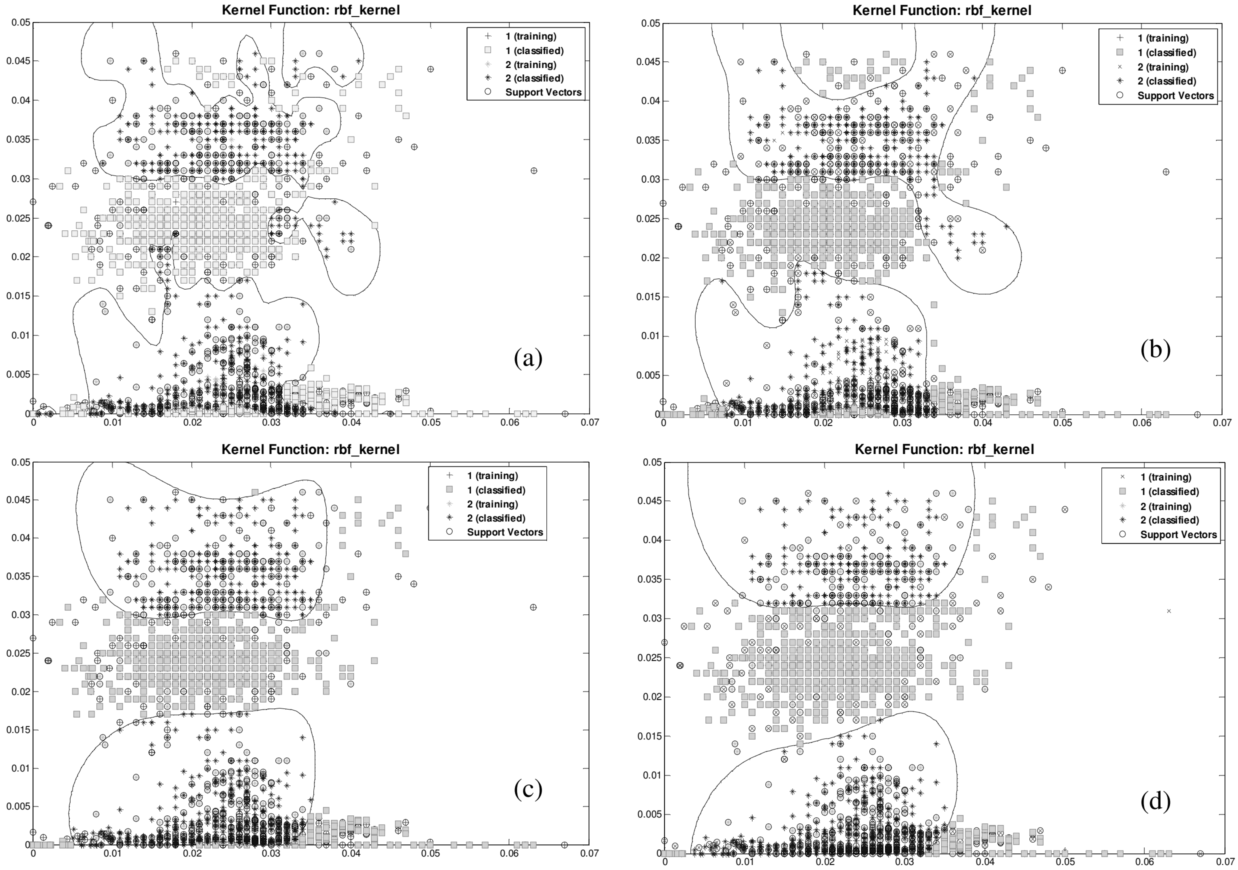

Fig. 3: The result of classification with RBF kernel and entropy criterion. (a) σ=0.2, (b) σ=0.5, (c) σ=1, (d) σ=1.5.

RBF with large σ will allow a support vector to have a strong influence over a larger area. A larger σ value also increases the α value (the Lagrange multiplier) for the classifier. In this study, the best accuracy achieved with σ= 0.5. One of the advantages of the RBF kernel is that given the kernel, the α_i (the Lagrange multipliers), the number of support vectors and the support vectors are all automatically obtained as a part of the training procedure, i.e. they need not be specified by the training mechanism. At the end, we can summarize the advantages of SVM and the reasons for using it as a classifier as follow: (1) There are no problems with local minima, because the solution is a quadratic programming (QP) problem. (2) There are few model parameters to select.

(3) The final results are stable and repeatable.

(4) SVM represents a general methodology for many pattern recognition (PR) problems: classification, regression, feature extraction, clustering, ...

(5) SVM is a minimum memory space approach.

(6) SVM provides a method to control complexity independently of dimensionality.

(7) SVM have been shown (theoretically and empirically) to have excellent generalization capability.

References

[1] M. J. Zwarts, G. Bleijenberg, and B. G. M. v. Engelen, "Clinical neurophysiology of fatigue," *Clinical Neurophysiology*, vol. 119, pp. 2-10, 2008.

[2] S. Karlsson, J. Yu, and M. Akay, "Enhancement of Spectral Analysis of Myoelectric Signals During Static Contractions Using Wavelet Methods," *IEEE Trans. Biomed. Eng.,* vol. 46, pp. 670-685, 1999.

[3] M. A. Oskoei, H. Hu, and J. Q. Gan, "Manifestation of Fatigue in Myoelectric Signals of Dynamic Contractions Produced During Playing PC Games," presented at the 30th Annual Int. Conf. IEEE EMBS, Vancouver, British Columbia, Canada, 2008.

[4] G. T. Allison and T. Fujiwara, "The relationship between EMG median frequency and low frequency band amplitude changes at different levels of muscle capacity," *Clinical Biomechanics*, vol. 17, pp. 464-469, 2002.

[5] G. I. M, R. C. I, M. G. F, M. A, and I. M, "New Wavelet Indices to Assess Muscle Fatigue during Dynamic Contractions," World Academy of Science, Engineering and Technology, vol. 5 , pp. 480-485, 2009.

[6] M. Cifrek, V. Medved, S. Tonkovic, and S. Ostojic, "Surface EMG based muscle fatigue evaluation in biomechanics," *Clinical Biomechanics*, vol. 24, pp. 327-340, 2009.

[7] E. Ghoochani, S. R. Ghoochani, M. Ravari, H. A. Hoseyni, "Detecting and Predicting Muscle Fatigue during Typing by SEMG Signal Processing and Artificial Neural Networks," *Iranian Journal of Medical Physics*, vol. 3, pp. 31-40, 2011.

[8] A. Georgakis, L. K. Stergioulas, and G. Giakas, "Fatigue Analysis of the Surface EMG Signal in Isometric Constant Force Contractions Using the Averaged Instantaneous Frequency," *IEEE Trans. Biomed. Eng.,*vol. 50, pp. 262-265, 2003.

[9] F. Al-Mulla, "Statistical Class Separation using sEMG Features Towards Automated Muscle Fatigue Detection and Prediction ", IEEE pp. 1-5, 2009.
[10] C. J.C.Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, pp. 121-167, 1998.
[11] L. M. Stirling, V. v. Tscharner, P. F. Kugler, and B. M. Nigg, "Classification of muscle activity based on effort level during constant pace running," *Journal of Electromyography and Kinesiology*, 2011.
[12] C. M. Bishop, "Sparse Kernel Machines," in *Pattern Recognition and Machine Learning*: Speringer, 2006, pp. 325-357.
[13] R. Merletti and P. Parker, *Electromygraphy, Physiology, Engineering, and Noninvasive Applications*. New Jersey: John Wiley & Sons, Inc., 2004.
[14] C. R. Hendrix, T. J. Housh, G. O. Johnson, M. Mielke, C. L. Camic, J. M. Zuniga, and R. J. Schmidt, "A new EMG frequency-based fatigue threshold test," *Neuroscience Methods,* vol. 181, pp. 45-51, 2009.