

Amyotrofik Lateral Skleroz (ALS) Hastalığının Destek Vektör Makinesi ile Teşhisi

Diagnosis of Amyotrophic Lateral Sclerosis (ALS) with Support Vector Machine

Hanife Küçük¹, Cengiz Tepe², İlyas Eminoğlu³

^{1,2,3}Elektrik-Elektronik Mühendisliği, Ondokuz Mayıs Üniversitesi

¹husta@omu.edu.tr, ²ctepe@omu.edu.tr, ³ilyaseminoglu@hotmail.com

Özet

Elektromiyogram (EMG) işaretindeki Motor Ünite Aksiyon Potansiyelleri (MÜAP) kas ve sinir hastalıkları tanısı için önemli bilgi sunarlar. Bu çalışmada ALS hastalığı ve sağlıklı verilerinin sınıflandırılması için her bir EMG verisine ait MÜAP' ların zaman uzayı ve frekans uzayında toplam 10 tane öznelik vektörü ve bunların çeşitli şekilde birleştirilmesi ile elde edilen hibrit öznelik vektörleri hesaplanmıştır. Seçilen öznelik vektörlerinin Destek Vektör Makinesi (DVM) sınıflandırıcısı üzerindeki etkileri incelenmiştir. Sonuç olarak, hibrit öznelik vektörlerinin kullanılması durumunda DVM sınıflandırıcısı, ALS ve sağlıklı EMG işaretlerini % 98.36 doğrulukla sınıflandırmıştır.

Abstract

The Motor Unit Action Potentials (MUAPs) in an electromyogram (EMG) signal provide an important source of information for the diagnosis of neuromuscular disorders. In this study, MUAPs' in an EMG data set (that contains both healthy and ALS (Amyotrophic Lateral Sclerosis) disease subjects) are represented in time domain and frequency domain with a total of 10 feature vectors. The selected feature vectors and their hybridized versions are processed by using Support Vector Machine (SVM) classifier. An accuracy of 98.36% is obtained with hybrid feature vectors.

1. Giriş

EMG, kasın kasılmasıyla meydana gelen elektrokimyasal olaylar sonucu ortaya çıkan biyopotansiyel işaretlerdir. Elektromiyografi ise, kasların kasılmasını sağlayan elektriksel aktivitenin izlendiği ve yorumlandığı bir kas incelemesidir. Tıpta birçok uygulama alanında kendine yer bulan EMG işaretleri özellikle kas ve sinir davranışlarının incelenmesinde aktif rol oynamaktadır. EMG işareti ne periyodik ne de deterministik bir işarettir [1]. Başka bir deyişle ayrı ayrı zamanlarda, keyfi olarak seçilecek EMG işaretlerinin istatistiksel davranışı tam olarak aynı değildir.

Belirli zaman aralıklarında EMG işaretleri kendilerini tekrar etmezler ve tek bir matematiksel ifade kayıt süresinin tamamı boyunca elde edilen EMG işaretini temsil edemez [2].

İşaretlerin değerlendirilmesi zaman düzleminde ve frekans düzleminde yapılabilir. EMG işaretleri durağan işaret değildirler. Buna ilaveten, her bir motor birimlerin ateşleme frekansı, kasılma miktarıyla beraber artmakta ve yeni bir motor ünite ateşleme yapmaya kadar bu davranışını korumaktadır. Dolayısıyla motor birimlerin ve bunun sonucu olarak EMG işaretlerinin belirli bir frekans davranışından söz edilmesi zordur [3].

EMG verisinde hastalık tanısı MÜAP dalga şeklinin bütüne bakılarak verilmelidir. Zamansal parametreler olan MÜAP' ın genlik, süre, faz ve dönüş sayısı değerleri tanı koymada hesaplanan değerlerdir. Ancak bu zamansal değerler tanı için yeterli değildir. Bu nedenle EMG verisinin tanısız değerini arttırmak için yüksek dereceli istatistikler, spektral ve dalgacık analizi, doğrusal ayırım analizi, sayısal filtreleme, modelleme analizleri ve yapay sinir ağları gibi yeni işaret işleme tekniklerine ihtiyaç duyulmaktadır [4].

Günümüzde EMG işaretleri üzerine yapılan çalışmaların çoğunda EMG işaretlerinin karmaşık ve zamanla değişken olması gibi özelliklerine uymayan zaman veya frekans düzlemleri teknikler kullanılmaktadır. Araştırmacılar kas ve sinir (nöromusküler) sisteminin davranışını, o bölgeden alınan EMG işaretlerinden anlayabilmek için tümleştirme (integrasyon), lineer yüzey örtüsü, kareler ortalamasının kare kökü, sıfır geçişlerini sayma gibi bir çok teknikler uygulamışlardır [5]. Son yıllarda geliştirilen bu yöntemlerin bazıları işaretlerin istatistiksel analizinin yapılmasında, diğerleri ise özel dalga şekillerinin belirlenip, ayrıştırılıp incelenmesinde kullanılmaktadır. Bu yöntemlere örnek olarak; zaman analizi, frekans analizi, zaman-frekans analizi yöntemi ve dalga şekillerinin ayrıştırılması ve sınıflandırılması sayılabilir [6].

Son zamanlarda yapılan çalışmalara baktığımızda işaret işleme teknikleri ve yapay zeka tekniklerinin gelişmesine paralel olarak bilgisayar tabanlı EMG analiz algoritmaları gelişmiştir.

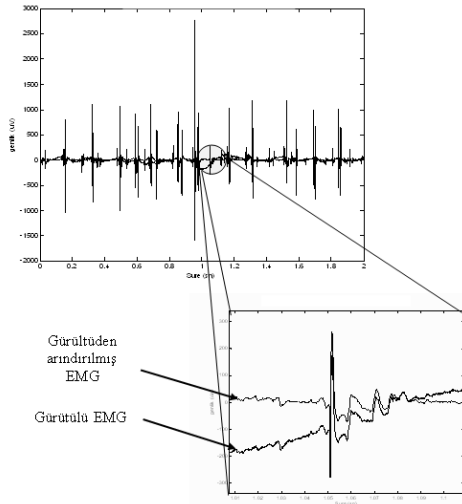
Bu çalışmada EMG işaretlerinden gerekli olan bilgilerin çıkarılması ve sınıflandırılması beş aşamada gerçekleştirildi: i) ön işlem aşamasında, EMG işaretindeki gürültü ve izoelektrik çizgideki kayma/dalgalanmayı yok etmek için bant geçiren süzgeç kullanıldı. ii) Bölütleme aşamasında, MÜAP' ların başlangıç ve bitiş noktaları bulunarak bölütleme yöntemi ile EMG işaretindeki MÜAP' lar

belirlendi. iii) Benzer olan MÜAP' ların kümelenmesi hiyerarşik kümeleme yöntemi ile sağlandı. iv) Kümelenmiş MÜAP' ların zaman uzayı ve frekans uzayında öznelik vektörleri hesaplandı v) Son olarak çıkarılan öznelik vektörleri Destek Vektör Makinesine verilerek sınıflandırıldı.

2. Materyal ve Yöntem

2.1. Veri Edinme Ve Önışlem

Bu çalışmada kullanılan veriler EMGLAB [17]' deki internet sitesinden alınmıştır. Veriler 8 saniye boyunca brachial biceps, vastus medialis kaslarından iğne elektrot ile alınmıştır. Bu EMG işaretleri 10'u sağlıklı, 7 ALS hastalığı olan kişilerden toplam 371 veri alınmıştır. EMG ölçümü sırasında elektrotların vücutla etkileşimi ve hareketleri nedeniyle EMG işaretinin izoelektrik çizgide kayma/dalgalanmalar (Baseline drift/Baseline wander) meydana gelmektedir. Bu kaymalar EMG' den elde edilen bazı önemli bilgilerin bozulmasına neden olur. EMG işaretini doğru yorumlamak için bu kaymaları yok etmek gerekmektedir. EMG işaretindeki yüksek frekans bileşenli gürültüyü ve düşük frekanslı kaymaları gidermek için frekans aralığı 10-1500 Hz olan "cheybshev" band geçiren süzgeç kullanıldı. İzoelektrik çizgideki dalgalanmayı gidermek için yüksek geçiren süzgeç kullanılması durumunda elde edilen EMG işareti Şekil 1' de verilmiştir.



Şekil 1: İzoelektrik çizgideki dalgalanmanın giderilmesi

2.2. MÜAP' ların Başlangıç Ve Bitiş Noktaları Bulunarak Bölütleme

Bu bölümde, EMG işaretindeki MÜAP' ların bölütlenmesi 3 adımda gerçekleştirildi.

Adım 1: EMG işaretinde eşiği aşan tepeler belirlenir:

Literatürde yapılan çalışmalarda MÜAP genliği için standart bir eşik seviyesi kullanılmamaktadır. Dorfman ve McGill [7]' deki çalışmalarında eşik genliğini 20–50 uV, Stalberg ve ark., 1995' deki çalışmalarında eşik genliğini, 50 uV olarak seçmiştir [8]. Bu çalışma için eşik değeri 50 uV olarak seçilmiştir.

Adım 2: MÜAP' ların başlangıç ve bitiş noktalarını bulmak için EMG işareti üzerinde kayan, 1 ms süren pencerenin alt ve üst eşik değerinin belirlenmesi:

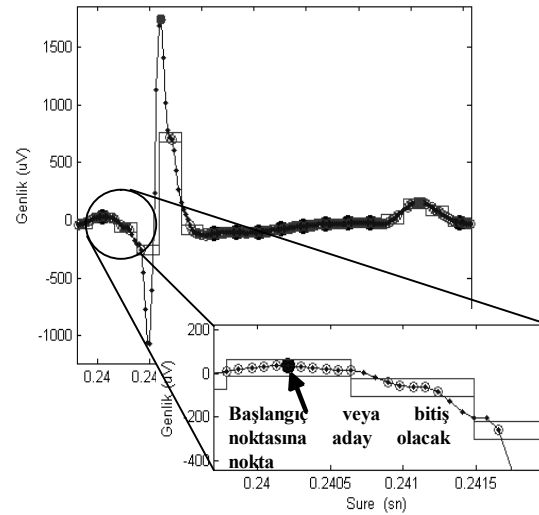
İşaret üzerinde kayan pencerenin dalga biçimine göre tolerans aralığı bu çalışmada 1ms süren pencere içerisinde kalan noktalardan genlik olarak ortanca genliğin ± 40 uV değeri alt ve üst eşik değeri olarak belirlenmiştir [9].

Adım 3: Başlangıç ve bitiş noktasına aday olacak noktaların belirlenmesi (Şekil 2):

eğer Adım 2' de belirlenen pencere içerisinde alt ve üst eşiği aşan tepe yoksa bu noktalardan ortancası başlangıç veya bitiş noktası olmaya adaydır.

değilse (pencere içerisinde en az bir nokta eşik değerini aşarsa) pencere içerisindeki noktalar dikkate alınmaz.

Adım 1' de EMG işaretinde eşiği aşan her bir MÜAP tepesinin solunda kalan Adım 2' de bulunan ilk aday nokta MÜAP' ın başlangıç noktası olurken; sağında kalan ilk aday nokta MÜAP' ın bitiş noktasıdır

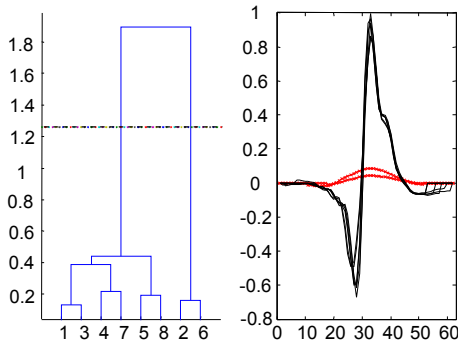


Şekil 2: Başlangıç ve bitiş noktasına aday olacak noktalar

2.3. Aşamalı (Hiyerarşik) Kümeleme Tekniği ile MÜAP Kümeleme

Hiyerarşik kümeleme metotları, veri setinin birimlerinin birbirlerine olan uzaklık değerlerini kullanarak, veri setindeki birimlerin hiyerarşik ayrıştırmasını yapar. Hiyerarşik ayrıştırma sırasında, "ağaç veri yapısı" olarak da bilinen dendogram kullanılır. Dendogram, hiyerarşik kümeleme tekniğiyle elde edilen kümelerin görselleştirilmesini sağlar.

Şekil 3' te Tek Bağlantılı Kümeleme Yöntemi ile kümeleme kullanılarak elde edilmiş ağaç veri yapısı ve belirli eşik değerine karşılık elde edilen kümeleme sonuçları verilmiştir. Bu uygulamada eşik değeri her bir eşleştirme sonucunda elde edilen Öklid mesafesinin en büyük değeri belirli bir katsayı ile çarpılarak elde edilmiştir. Denemeler sonucunda en iyi sonucu veren katsayı 0.66 olarak belirlenmiştir.



Şekil 3: Ağaç Veri Yapısı ve Belirli Eşik Değerlerine Karşılık Elde Edilen Kümeleme Sonuçları

2.4. Öznitelik vektörlerinin çıkarılması

Bu çalışmada Çizelge 1'de verilen zaman uzayı ve frekans uzayında toplam 10 tane öznitelik vektörü ve bunların çeşitli şekilde birleştirilmesi ile elde edilen hibrit öznitelik vektörleri her bir EMG verisi için hesaplanmıştır. Zor örüntü tanıma problemlerinde ham veriden elde edilen çoklu öznitelik setleri çeşitli şekilde birleştirilip sınıflandırıcıya giriş olarak verilmesi halinde genel sınıflandırma performansını artırmak mümkündür [10]. Bu çalışmada sınıflandırma işleminin verimini arttırmak amacıyla, EMG verisinden elde edilen MÜAP' den çıkartılan özniteliklerin çeşitli şekilde birleştirilmesi sınıflandırıcıya giriş olarak verilmesi ilkesine dayanmaktadır. Bu çalışmada her bir EMG verisine ait MÜAP' ların öznitelikleri hesaplandıktan sonra ortalamaları ve standart sapma değerleri sınıflandırıcıya giriş olarak verilmiştir.

Çizelge 1: Zaman ve Frekans Uzayında Öznitelik Vektörleri

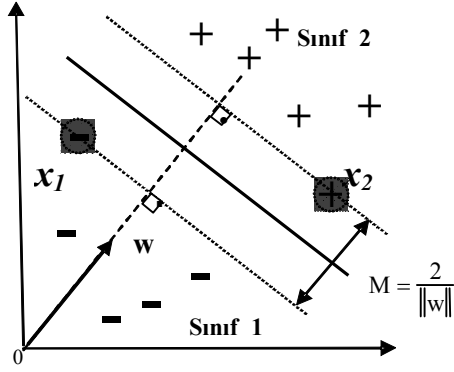
<p><i>Tümleşik EMG</i></p> $TEMG = \sum_{i=1}^N x_i $	<p><i>Ortalama Frekans</i></p> $ORTF = \frac{\sum_{i=1}^M f_i P_i}{\sum_{i=1}^M P_i}$
<p><i>Ortalama Mutlak Değer</i></p> $OMD = \frac{1}{N} \sum_{i=1}^N x_i $	<p><i>Ortanca Frekans (OCF)</i></p> $\sum_{i=1}^{OCF} P_i = \sum_{i=OCF}^M P_i = \frac{1}{2} \sum_{i=1}^M P_i$
<p><i>Emg Varyans</i></p> $EVAR = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$	<p><i>Ortalama Güç</i></p> $OG = \frac{\sum_{i=1}^{OCF} P_i}{M}$
<p><i>Etkin Değer</i></p> $ED = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$	<p><i>Toplam Güç</i></p> $TG = \sum_{i=1}^M P_i$
<p><i>Willison Genliği</i></p> $WG = \sum_{i=1}^{N-1} [f(x_n - x_{n+1})]$ $f(x) = \begin{cases} 1, & x \geq \text{esik} \\ 0, & \text{aksi durumda} \end{cases}$	<p><i>Sıfır Geçiş Sayısı (SGS)</i></p> $\{x_i > 0 \text{ ve } x_{i+1} < 0\} \text{ veya } \{x_i < 0 \text{ ve } x_{i+1} > 0\} \text{ ve } x_i - x_{i+1} \geq \text{esik}$

2.5. Destek Vektör Makinesi (DVM) İle Sınıflandırma

Destek vektör makineleri (DVM), Vladimir N. Vapnik tarafından ortaya konulan istatistiksel öğrenme teorisine dayalı öğrenme algoritmasıdır [12]. Genellikle DVM' de ikili sınıflandırmada etiketlenmiş eğitim verisi (1) ifadesindeki gibi belirtilir. Burada l ; örnek sayısı x_i ; örnekler, y_i ; bu örneklere ait etiketleri temsil etmektedir.

$$\{x_i, y_i\}_{i=1}^l \quad x_i \in R, \quad y_i \in \{-1, +1\} \quad (1)$$

İkili sınıflandırma probleminde doğrusal olarak ayrılabilen bir veri setinin olduğu düşünülürse, bu veri setini ayırabilen sonsuz sayıda hiper-düzlem vardır. DVM karar yüzeyini oluştururken iki sınıfa olan uzaklığı maksimum (en fazla) yapmaya çalışır. Bu düzlemler arasında maksimum sınıra sahip sadece bir hiper-düzlem bulunmaktadır. Sınır genişliğini sınırlandıran noktalara ise destek vektörleri adı verilir. Şekil 4' de M maksimum aralık, x_1, x_2 ; sınır genişliğini sınırlandıran destek vektörleridir. Destek vektör algoritması en büyük sınır genişliğine sahip ayırıcı hiper-düzlem ile sınıflandırma yaparak eğitim hatasını minimize etmeye çalışır.



Şekil 4: İkili sınıflandırmada karar düzlemi.

Sınırı maksimum yapmak; denklem (2)'deki amaç fonksiyonunu, (3)'deki kısıtlar altında minimize etmeye denktir. (3) ifadesindeki w : ağırlık parametreleri vektörü, b ise: öteleme değeridir.

$$\frac{1}{2} \|w\|^2 \quad (2)$$

$$y_i (w^T x_i + b) \geq 1 \quad (3)$$

Eğer veriler oluşturulan sınırın içinde veya hatalı tarafında ise bu veri noktaları için genelleme yapılır ve kısıtları gevşetmek için yapay/aylak değişkenler tanımlanır [13]. (2) ve (3) denklemlerine yapay/aylak değişkenler eklenmesi durumunda (4) ve (5) denklemleri elde edilir.

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \zeta_i \quad (4)$$

$$y_i (w^T x_i + b) \geq 1 - \zeta_i \quad i = 1, 2, \dots, N \text{ ve } \zeta_i \geq 0 \quad (5)$$

(4) denklemini minimize etmek için; Lagrange çarpanları yöntemi kullanılmaktadır. Böylece problem, denklem (6)'daki ikili (dual) biçime dönüşür. Denklem (6)'daki α_i 'ler Lagrange çarpanlarıdır. Karush-Kuhn-Tucker şartına göre [14]; herhangi bir veri için $\alpha_i > 0$ ise bu veriler destek vektörleri olarak adlandırılır, $\alpha_i = 0$ ise bu veri destek vektörü değildir ve yeni verilerin sınıflandırılmasında hiçbir rol oynamamaktadır. Denklem (7)'deki α_i 'nin üst sınırı olan C kullanımcı girişlidir.

$$\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (6)$$

$$0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, N \quad (7)$$

Diğer bir çözüm yolu olarak giriş uzayında doğrusal olarak ayrılamayan veriler haritalama fonksiyonu, $\varphi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^{d^p}$ ile yüksek boyutlu bir uzaya haritalanarak ayrılabilir bir duruma gelmektedir. Özellik uzayına aktarılan iki vektör çarpılır bu çarpıma kernel fonksiyonu denir: $K(x, z) = \varphi(x)^T \varphi(z)$. Böylece denklem (6)'daki ikili problem, denklem (8)'e dönüşür.

$$\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (8)$$

DVM, Mercer koşullarını sağlayan kernel fonksiyonlarını kullanır. Yaygın olarak kullanılan kernel fonksiyonlarından biri denklem (9)'da verilen radyal tabanlı fonksiyon (RBF) kerneldir [15]. Denklem (10)'daki σ değeri RBF fonksiyonunun genişliğidir.

$$K(x_i, x_j) = \exp\left(-\frac{1}{\gamma} \|x_i - x_j\|^2\right) \quad (9)$$

$$\gamma = 1 / 2 \sigma^2 \quad (10)$$

3. Sonuçlar

Bu çalışmada EMG analizi için toplam 17 kişiye (10'u sağlıklı, 7 ALS hastası) ait veriler Matlab'ın Bioinformatics Toolbox™ da yer alan ikili destek vektör makinesi kullanılarak sınıflandırıldı [16]. Sınıflandırıcının performansını belirlemek için çapraz geçerlilik (Cross Validation- CV) yöntemi kullanılmıştır. Çapraz geçerlilikte k değeri 10 seçilmiştir. Kernel fonksiyonu olarak RBF seçilmiştir. Deneysel çalışmalarda değişen Gauss genişliği: γ ve ceza faktörü: C değerleri için denemeler yapılmış; en iyi sonuç $\gamma=0.85$, $C=100$ için elde edilmiştir. Bu çalışmada zaman uzayı ve frekans uzayında toplam 10 tane öznelik vektörü ve bunların çeşitli şekilde birleştirilmesi ile elde edilen hibrit öznelik vektörleri ile sınıflandırma sonuçları Çizelge 2'de verilmiştir. Sonuç olarak ALS ve sağlıklı EMG işaretleri hibrit TEMG+SGS+WG vektörünün ortalama olarak % 98.36 doğrulukla sınıflandırmıştır. Çizelge 2'deki sınıflandırma sonuçlarına baktığımızda, zaman düzleminde çıkartılan öznelik vektörlerinin sınıflandırma başarımları, frekans düzlemi öznelik vektörlerine göre daha yüksek olduğu gözlemlenmiştir. Bundan sonraki çalışmada, kullandığımız veri seti literatürde geçen farklı öznelik vektörleri ve sınıflandırma algoritmaları ile denenecektir.

Çizelge 2: Sınıflandırma Sonuçları

Öznelik Vektörü	Test Başarım (%)
TEMG+SGS+WG	98.36
OMD+SGS+WG	97.90
OMD	97.09
TEMG	96.98
EVAR	96.67
OMD+OG	95.17
OG	94.55
ED	93.74
WG	92.01
TG	73.67
OCF	71.72
ORTF	66.03
SGS	50.83

Teşekkür

Bu çalışma PYO. MUH.1906.10.001-BAL-LAB numaralı Ondokuz Mayıs Üniversitesi- BAP projesi tarafından desteklenmiştir.

4. Kaynakça

- [1] Rodriguez, I., Malanda, A., Gila, L., Filter design for cancellation of baseline – fluctuation in needle EMG recordings, *Comput. Meth. and Prog. in Biomed.*, 81, 79-93, 2006.
- [2] Okkesim, Ş., Ortodontik Anormallikler Bulunan Hastalarda Kullanılan Pre Ortodontik Trainer Aparentinin Çene Ve Ağız Çevresindeki Kaslara Olan Etkisinin Trainer Aparentinin Çene Ve Ağız Çevresindeki Kaslara Olan Etkisinin EMG Kayıtlarıyla Değerlendirilmesi, Yüksek Lisans Tezi, Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 2006.
- [3] Özkaya, U., Yüzey EMG Sinyallerinin Analizi ve Motor Ünite Aksiyon Potansiyellerinin Belirlenmesi, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Isparta, 2005.
- [4] Dobrowolski, A. P., Wierzbowski, M., Tomczykiewicz, K., Multiresolution MÜAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders, *Computer Methods and Programs in Biomedicine*, Elsevier, 2010.
- [5] Pagniotacopulos, N. D., Lae, J. S., Pope, M. H., Evaluation of EMG signals from rehabilitated patients with low back pain using wavelets. *J. Electromyography and Kinesiology* 8, 269 – 278, 1998.
- [6] Bozkurt, M. R., EMG İşaretlerinin Modern Yöntemlerle Önişlemesi ve Sınıflandırılması, Doktora Tezi, Sakarya Üni. , Fen Bilimleri Enst., Sakarya, 2007.
- [7] Dorfman, L. J., McGill, K. C., AAEE minimonograph 29: automatic quantitative electromyography, *Muscle & nerve* 11 (8), 804-818, 1988.
- [8] Stalberg, E., Falck, B., Sonoo, M., Stalberg, S., and Astrom, M., Multi-MUP EMG analysis- a two year experience in daily clinical work. *Electroencealography and Clinical Neurophysiology* 97. Amsterdam, the Netherlands: Elsevier Science, 145-154, 1995.
- [9] Gurmanik, K., Shatru, A., Jain, V. K., Multi-Class Support Vector Machine Classifier in EMG Diagnosis, *WSEAS Transactions on Signal Processing*. 5 (12), 379-389, 2009.
- [10] Christodoulou, C. I., Pattichis, C. S., Middleton, L. T. Multi-Feature, Multi-Classifer Analysis in EMG Diagnosis. *Proceedings of the VIII Mediterranean Conference on Medical and Biological Engineering and Computing*, 1998.
- [12] Burges, J.C., A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2, 121-167,1998.
- [13] Tan, Y. ve Wang, J. A Support Vector Machine with a Hybrid Kernel and Minimal Vapnik-Chervonenkis Dimension. *IEEE Transactions on Knowledge and Data Engineering*, v.16 n.4, pp. 385-395, 2004.
- [14] Abe, S. Support Vector Machines for Pattern Classification. (*Advances in Pattern Recognition*) Springer; 1 edition, 2005.
- [15] Vapnik, V.N. An Overview of Statistical Learning Theory, *IEEE transactions on Neural Networks*, Vol.10, No.5, pp. 988-999., 1999
- [16] MathWorks, *Matlab Toolbox of Bioinformatics: User's Guide*. Mathworks Inc, 2007.
- [17] 10/07/12 tarihinde aktif adres: <http://www.EMGlab.net>