

Beyin Bilgisayar Arayüzü Uygulamalarında Rasgele Orman Sınıflaması

Random Forest Classification for Brain Computer Interface Applications

Hatice OKUMUŞ ve Önder AYDEMİR

Elektrik-Elektronik Mühendisliği Bölümü

Karadeniz Teknik Üniversitesi

Trabzon, Türkiye

haticeokumus@ktu.edu.tr, onderaydemir@ktu.edu.tr

Özetçe—Beyin bilgisayar arayüzü uygulamaları insan beyni ve cihazlar arasında bir köprü oluşturmada önemli bir rol oynamaktadır. Bu alandaki çalışmalar, farklı öznelik çıkarma ve sınıflandırma yöntemlerinin kullanılmasıyla her geçen gün artmaktadır. Bu çalışmada BCI Competiton 2003 yarışmasında sunulmuş olan Data Set III veri kümesi kullanılarak, Rastgele Orman yöntemi ile sınıflandırma yapılmış ve güncel önerilerden olan Fast Walsh Hadamard dönüşümünün, Fourier ve Dalgacık Dönüşümü ile birleştirilmesinin sınıflandırma doğruluğunda etkili sonuçlar verdiği gösterilmiştir. Sıklıkla kullanılan Destek Vektör Makineleri, k- En Yakın Komşu ve Doğrusal Ayırıcı Analizi yöntemleri bu çalışmada da uygulanarak elde edilen sonuçlar literatürde aynı veri setini kullanan çalışmalar ile karşılaştırılmıştır. En yüksek sınıflandırma doğruluğu, belirtilen dönüşüm yöntemlerinin birlikte kullanılması ve sınıflandırma yönteminin Rastgele Orman olduğu durumda %89.06 olarak elde edilmiştir ve bu da tercih edilen sınıflandırma yöntemi ve dönüşüm yöntemlerinin literatürdeki diğer sınıflandırıcılara bakıldığında etkili olduğunu göstermektedir.

Anahtar Kelimeler — *beyin bilgisayar arayüzü; rastgele orman sınıflandırıcı; öznelik çıkarımı; dalgacık dönüşümü; fast walsh hadamard dönüşümü.*

Abstract— Brain computer interface applications have big importance in becoming a bridge between the human brain and devices. The studies in this area increase every day with the use of different feature extractions and classification methods. In this study, classification is done by Random Forest method using Data Set III presented in BCI Competiton 2003, and it has been shown that combining the Fast Walsh Hadamard Transformation with Fourier and Wavelet Transform is effective in the classification accuracy. The frequently used Support Vector Machines, k-Nearest Neighbor, and Linear Separator Analysis methods are also applied in this study and the results are compared with the literature results which have used the same dataset. The highest classification accuracy was obtained as 89.06% when the specified conversion methods were used together and the classification method was Random Forest, indicating that the preferred classification method and conversion methods are effective when compared to other classifiers in the literature.

Keywords — *brain computer interface; random forest classification; feature extraction; wavelet transform; fast walsh hadamard transform.*

I. GİRİŞ

Fiziki engellerin sebep olduğu durumları ortadan kaldırmayı hedefleyen Beyin Bilgisayar Arayüzü sistemleri, insan beyni ile cihazlar arasında bir köprü görevi görmektedir. Bu sistemlerde, elektrokortikogram, Fonksiyonel Manyetik Rezonans Görüntüleme, manyetoensefalografi gibi işaretlerin yanı sıra; daha kolay uygulanabilen, daha kararlı ve yüksek çözünürlüğe sahip olan elektroensefalogram (EEG) işareti tercih edilmektedir. Yapılan çalışmalar daha çok EEG işareti üzerine yapılmış olup, en iyi sınıflandırma doğruluğunu veren öznelik çıkarma yöntemleri ve sınıflandırıcılar üzerine yoğunlaşmaktadır [1].

Literatürde çeşitli hareket hayallerine dayalı EEG işaretleri ile; frekans domeninde bilgi veren Fourier Dönüşüm yöntemi [2], [3], zaman domeninde bilgi veren Dalgacık Dönüşüm yöntemi [4], [5], Hilbert-Huang Dönüşümü yöntemi [6], güç yoğunluğu parametreleri [7], Fast Walsh Hadamard Dönüşümü (FWHT) yöntemi [8] gibi öznelik çıkarma yöntemleri ve k- En Yakın Komşu (k-EYK), Destek Vektör Makineleri (DVM), Doğrusal Ayırıcı Analizi (DAA) gibi sınıflandırıcılar kullanılarak birçok çalışma yapılmıştır.

Bu çalışmada BCI competiton 2003 yarışmasında kullanılan sağ/sol el hareketi hayaline dayalı EEG işaretleri, Rastgele Orman (RO) sınıflandırıcısı ile sınıflandırılmış ve öznelik çıkarma yöntemi olarak FWHT Dönüşümü, Fourier ve Dalgacık Dönüşümü ile beraber kullanılmıştır. Ayrıca, RO sınıflandırıcısına ek olarak performans karşılaştırması yapmak için k-EYK, DVM ve DAA sınıflandırıcıları ile de sınıflandırma doğruluğu hesaplanarak sonuçlar literatürde aynı veri setini kullanan çalışmalar ile karşılaştırılmıştır. RO sınıflandırıcısıyla elde edilen sonuçların diğerleriyle karşılaştırıldığında etkili sonuçlar verdiği görülmüştür. FWHT dönüşümünün diğer dönüşüm yöntemleri ile beraber kullanılmasının da sınıflandırma doğruluğunu pozitif yönde etkilediği görülmüştür.

II. MATERYAL VE YÖNTEM

A. Veriseti Tanıtımı

Yapılan çalışmada kullanılan veri kümesi BCI Competition 2003 yarışmasındaki Graz veri kümesi (database III) olup, 25 yaşında sağlıklı bir bayandan alınmıştır. Amaç, deneğe sağ/sol elini hareket ettirdiği hayalini kurdurarak ekrandaki çubuğun hareket etmesini sağlamaktır. Deney esnasındaki beyin aktiviteleri saç derisine yerleştirilmiş olan 3 elektrot (C3, Cz, C4) ile kaydedilmiştir. Her biri 9 sn süren toplam 280 deneme yapılmıştır. Bu denemelerin 140 tanesi eğitim kümesi için (eşit sayıda sağ ve sol el hareketi içermekte) diğer 140 tanesi de test kümesi (yine eşit sayıda sağ ve sol el hareketi içermekte) olarak iki gruba ayrılmıştır.

EEG işaretlerinin her biri 128 Hz ile örneklenmiş ve 0.5-30 Hz arasında filtreleme yapılmıştır. Denemelerin her birinde aynı adımlar tekrarlanmış olup, ilk 2 saniyede herhangi bir şey yapılmamış ve 2. saniyede denemenin başladığını belirten uyarı ile birlikte ekranda 1 saniye boyunca ‘+’ şekli gösterilmiştir. 3. saniyede ise ekranda bir çubuk belirmiş ve deneğin bu çubuğu hareket ettirmesi istenmiştir. (sağa veya sola). Verilerin hepsi aynı gün alınmış olup daha fazla bilgi için [9] ve [10] kaynaklarına başvurulabilir.

B. Fourier dönüşümü

İşaretleri frekans bileşenlerine ayırmakta kullanılan Fourier Dönüşümü, öznitelik çıkarmada sıklıkla kullanılan bir yöntemdir. Frekans bileşenlerine ayrılan işaretlerin farklı özellikleri ortaya çıkarılmış olmakta ve bu da EEG işaretlerinin sınıflandırılmasında kolaylık sağlamaktadır. Bu çalışmada, Hızlı Fourier Dönüşümü (FFT) alınmış işaretlere FWHT Dönüşümü uygulanmıştır.

C. Fast Walsh Hadamard Dönüşümü

FWHT Dönüşümü, Walsh Hadamard Dönüşümünün hesaplama karmaşıklığı ortadan kaldırmak için geliştirilen bir dönüşüm yöntemidir [11]. FWHT Dönüşüm matrisi (H_n), sadece +1 /-1 değerlerini alan ve giriş verisi uzunluğu 2'nin kuvveti şeklinde olan (1)'de belirtildiği gibi bir dönüşüm matrisidir.

$$H_n = \frac{1}{\sqrt{2}} \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix} \quad (1)$$

Yukarıdaki denklemde $\frac{1}{\sqrt{2}}$, normalizasyon faktörünü temsil etmektedir. Yeni veriler, orijinal verilerin dönüşüm matrisi ile çarpılmasıyla elde edilmektedir.

D. Dalgacık Dönüşümü

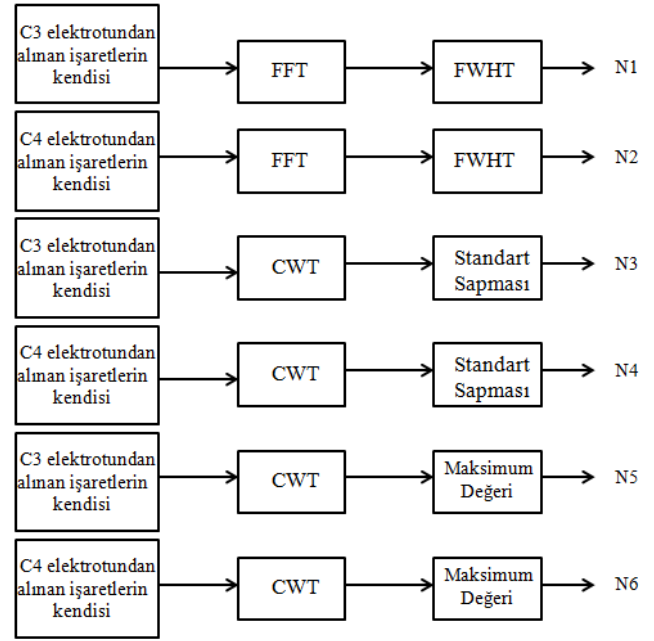
İşaretin hem frekans hem de zaman bileşenlerini içeren Dalgacık Dönüşümü, zaman içerisinde farklı frekans bileşenlerine sahip olan EEG işaretlerinde öznitelik çıkarma yöntemi olarak sıkça kullanılmaktadır. Çalışmada kullanılan sürekli dalgacık dönüşüm (CWT) katsayıları Denklem (2)'de verilen şekilde hesaplanmaktadır.

$$SDD(x, y) = x^{-1/2} \int_{-\infty}^{+\infty} Y(t) \psi\left(\frac{t-x}{y}\right) dt \quad (2)$$

Burada; $Y(t)$ işaretin kendisini, $\psi(t)$ dalgacık fonksiyonunu, x ölçeği ve y adım büyüklüğünü temsil etmektedir. Çalışmada kullanılan dalgacık modeli ve diğer parametreler işaretin kendisine bakılarak ve sınıflandırma doğruluğuna pozitif etki edecek şekilde belirlenmiştir.

E. Öznitelik Çıkarma

Bu çalışmada FWHT ve Dalgacık Dönüşüm yöntemleri kullanılarak C3 ve C4 elektrotlarından alınan EEG işaretlerinin öznitelikleri çıkarılmıştır. Öznitelik çıkarmasında sınıflandırma doğruluğuna artırıcı etki gösterdikleri için C3 ve C4 elektrotlarından alınan işaretler kullanılmıştır. Çıkarılan her bir özniteliğin (N1, N2, N3, N4, N5, N6) çıkarım aşamaları Şekil 1'de verilen blok diyagramları ile açıklanmıştır.



Şekil 1. Her bir özniteliğin çıkarım aşamaları

N1 ve N2 öznitelik çıkarımlarında görüldüğü üzere FFT'si alınmış olan işarete FWHT uygulanmıştır. İleriki aşamalarda bu yöntemin sağ/sol el hareketi işaretlerinin sınıflandırılmasında etkili olduğu görülecektir.

F. Sınıflandırma

Çalışmada RO sınıflandırıcısının yanında k-EYK, DVM ve DAA sınıflandırıcıları ile de performans incelenmesi yapılmıştır. 10 kat çapraz onaylama ile kullanılan özniteliklerin başarısı incelenmiştir. Çapraz onaylamada öncelikle eğitim verileri rastgele 10 eşit parçaya bölünmüş ve bu 10 parçadan bir tanesi onaylama verileri olarak kullanılmış geri kalan veriler ise eğitim verileri olarak kullanılmıştır. Her bir parça onaylama kümesi olacak şekilde parçalar taranmış ve her seferinde en yüksek sınıflandırma doğruluğunu verecek parametre elde edilmeye çalışılmıştır. Bu adımlar 100 kere tekrarlanarak optimum parametre elde edilmiş ve test verilerinin sınıflandırılmasında kullanılmıştır.

Kullanılan RO sınıflandırıcısı hakkında ayrıntılı bilgi verilmiş olup diğer sınıflandırma yöntemlerinden de kısaca bahsedilmiştir.

1) Rastgele Orman

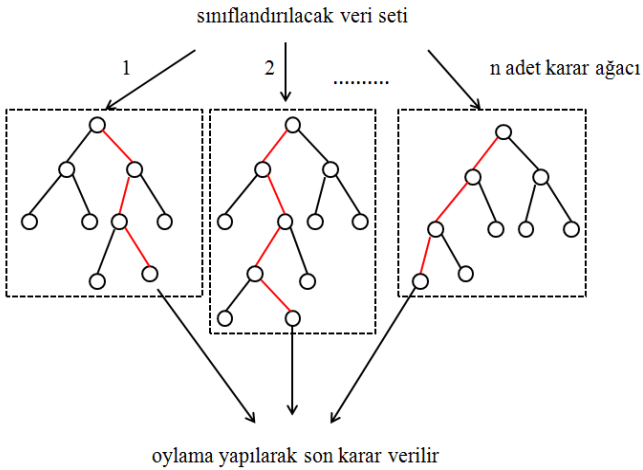
RO sınıflandırıcısı Breiman [12] tarafından ortaya atılan, karar ağaçları ve düğümlerden oluşan bir sınıflandırıcıdır. Bu sınıflandırıcı, düğümleri dallarına ayırırken her düğümde rastgele olarak seçilmiş değişkenlerden en iyi olanını kullanır.

Oluşturulan karar ağaçlarının (Out of Bag, OOB) iç hatalarına göre her bir ağaca belirli bir ağırlık verilir. En düşük hataya sahip karar ağacı en yüksek, en yüksek hataya sahip karar ağacı ise en düşük ağırlığa sahip olur. Bu ağırlıklara göre yapılan sınıf tahmininde oy verme işlemi gerçekleştirilir. Sonrasında oylar toplanarak en son karar verilir.

Rastgele orman sınıflandırıcısının adımları aşağıda verilmiştir:

- Veri setinin öznelikleri kullanılarak oluşturulacak karar ağacı adedi (n) belirlenir.
- Karar ağaçları içerisindeki her düğümde rastgele m adet değişken seçilir ve en iyi dal belirlenir. (gini indeksi ile hesaplama yapılarak [13])
- Belirlenen en iyi dal tekrar iki alt dala ayrılır ve gini indeksi sıfıra ulaşmaya kadar yani her bir yaprak düğümde bir sınıf kaldığında ağaç dallanma işlemi sonlanır [14].
- n karar ağacının ayrı ayrı yaptığı tahminler arasında en çok oyu alan sınıf son karar tahmini olarak seçilir.

Şekil 2’de RO sınıflandırıcısının genel yapısı verilmiştir.



Şekil 2. RO Sınıflandırıcısının Genel Yapısı

Bu sınıflandırıcıda en önemli seçim, her düğümde kullanılacak değişken sayısı (m) ve geliştirilecek ağaç sayısı (n) parametrelerinin seçimidir. Breiman [15]’e göre, m parametresi değişken sayısının kareköküne eşit alındığında en kararlı sonuç elde edilmektedir.

2) *k- En Yakın Komşu*: Makine öğrenme yöntemleri arasında uygulaması en basit yöntemlerden biri olan k-EYK yönteminde, sınıflandırılacak verinin en yakın k komşusu belirlenir ve bu komşuların ait oldukları sınıflara göre veri sınıflandırılır [6]. Çalışmada komşu uzaklıkları hesaplanırken Öklid uzaklığı kullanılmıştır.

3) *Destek vektör Makinaları*: DVM yönteminde amaç sınıflar arasındaki uzaklığı maksimize edecek doğruyu bulmaktır. Bu doğru doğrusal veya doğrusal olmayan fonksiyon olabilmektedir. Çalışmada kullanılan fonksiyon ve parametreler yüksek sınıflandırma doğruluğu verecek şekilde seçilmiştir.

4) *Doğrusal Analiz Ayırıcı*: Bu yöntemde ise sınıflar arasındaki dağılım maksimum olurken sınıf içi varyansı minimum yapacak doğru bulunmaya çalışılır. Verilerin normal dağılıma sahip olduğu varsayılmaktadır.

III. SONUÇLAR VE TARTIŞMA

Şekil 1’de belirtilen öznelilik kombinasyonları için öncelikle RO sınıflandırıcısının performansı incelenmiştir. RO sınıflandırıcı parametreleri, sınıflandırma doğruluğuna pozitif etki edecek şekilde belirlenmiştir. Her bir kombinasyonda test verilerinin tamamı kullanılarak elde edilen maksimum, minimum ve ortalama RO sınıflandırma doğruluğu değerleri Tablo 1’de verilmiştir. Tabloda görüldüğü üzere bütün özneliklerin beraber kullanılmasıyla maksimum %90.71 sınıflandırma doğruluğu elde edilmiştir. Farklı öznelilik çıkarma yöntemlerinin beraber kullanılmasının RO sınıflandırıcısı sonuçlarını olumlu yönde etkilediği görülmüştür.

TABLO I. RO SINIFLANDIRICI SONUÇLARI

Öznelikler	Sınıflandırma Doğruluğu (%)		
	Maksimum Doğruluk	Minimum Doğruluk	Ortalama Doğruluk
N1+N2	83.57	78.57	81.31
N3+N4	86.43	82.14	84.46
N5+N6	86.43	81.43	83.89
N1+N3+N5	74.29	69.29	71.64
N2+N4+N6	77.14	72.14	74.82
N1+N2+N3+N4	90.00	84.29	88.11
N1+N2+N5+N6	90.71	85.00	87.84
N1+N2+N3+N4+N5+N6	90.71	86.43	89.06

Yapılan çalışmada sadece RO sınıflandırıcı kullanılmayıp k-EYK, DVM ve DAA sınıflandırıcılarının da performansı incelenmiştir ve sonuçlar Tablo 2’de verilmiştir. Her bir sınıflandırma yöntemi için en yüksek sınıflandırma doğruluğunun bütün öznelilik çıkarma yöntemleri birlikte kullanıldığında sırasıyla RO, k-EYK, DVM ve DAA için %89.06, %84.47, %83.57 ve %84.28 olarak elde edildiği

görülmüştür. Bu da FWHT dönüşüm yönteminin diğer dönüşüm yöntemleriyle birlikte kullanılmasının bu veri seti üzerinde daha etkili sonuçlar verdiğini göstermektedir.

TABLO II. FARKLI SINIFLANDIRICILAR İLE ELDE EDİLEN SONUÇLAR

Öznitelikler	Sınıflandırma Doğruluğu (%)			
	DVM	DAA	k-EYK	RO
N1+N2	67.85	77.85	82.2	81.31
N3+N4	81.42	82.14	81.96	84.46
N5+N6	81.42	82.14	81.8	83.89
N1+N3+N5	63.57	70.71	63.32	71.64
N2+N4+N6	71.42	75.71	68.35	74.82
N1+N2+N3+N4	80.71	84.28	84.1	88.11
N1+N2+N5+N6	81.42	84.28	83.07	87.84
N1+N2+N3+N4+N5+N6	83.57	84.28	84.47	89.06

Daha önce aynı veri setiyle yapılan çalışmalardan elde edilen sonuçlar Tablo 3'te verilmiştir. Elde edilen sonuçlara bakıldığında, EEG işaretlerinin RO yöntemi ile sınıflandırılmasının ve dönüşüm yöntemlerinin birleştirilmesiyle elde edilen öznitelik çıkarma yönteminin en yüksek sınıflandırma doğruluğuna yakın değerlerde etkili sonuç verdiği görülmüştür. İleriki çalışmalarda öznitelik eleme yöntemleri kullanılarak daha etkili öznitelikler ortaya çıkartılması ile daha iyi sonuçlar elde edilmesi amaçlanmaktadır.

TABLO III. ÖNCEKİ ÇALIŞMALARDAN ELDE EDİLEN SINIFLANDIRMA DOĞRULUKLARI

Yapılan Çalışmalar	Sınıflandırma Yöntemi	Sınıflandırma Doğruluğu (%)
Darvishi, S. vd. [16]	Bulanık Mantık	82.1
Bhattacharyya S. vd. [7]	DVM	82.14
Baig, M. Z. vd. [4]	SOM Neural Network	83.45
Tavakolian vd. [17]	Çok Katmanlı Algılayıcı	84.29
Khalid, M.B. vd. [18]	Yapay Sinir Ağları	87
Saka, K. vd. [8]	Yapay Sinir Ağları	88.87
Liu, Y.H. vd. [6]	DVM	90.7
Bashar, S.K. vd. [5]	k-EYK	91.07
Aydemir, Ö. vd. [2]	DAA	91.4
Önerilen Yöntem	RO	89.09

KAYNAKLAR

- [1] Anderson R.A., Musallam S., Pesaran B., "Selecting the signals for a brain-machine interface", *Curr Opin Neurobiol* 14(6): 720-726, 2004
- [2] Aydemir, Ö. and Kayıkçıoğlu T., "Classification of EEG signals recorded during right/left hand movement imagery using Fourier Transform based features", *IEEE 19th Conference on Signal Processing and Communications Applications (SIU)*, 415-418, 2011.
- [3] Elbaz, A. M. vd., "Motor imagery based brain computer interface using transform domain features", *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 6421-6424, 2016.
- [4] Baig, M. Z. vd., "Classification of left/right hand movement from EEG signal by intelligent algorithms", *IEEE Symposium on Computer Applications and Industrial Electronics (ISCAIE)*, 163-168, 2014.
- [5] Bashar, S. K., Hassan A. R. and Bhuiyan M. I. H., "Identification of motor imagery movements from eeg signals using dual tree complex wavelet transform", *International Conference on. Advances in Computing, Communications and Informatics (ICACCI)*, 290-296, 2015.
- [6] Liu, Y. H., Cheng C. A. and Huang H. P., "Novel feature of the EEG based motor imagery BCI system: Degree of imagery" *International Conference on System Science and Engineering (ICSSE)*, 515-520, 2011.
- [7] Bhattacharyya, S. vd., "Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms", *IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, 1-8, 2011.
- [8] Saka, K., Aydemir Ö. and Öztürk M., "Classification of EEG signals recorded during right/left hand movement imagery using Fast Walsh Hadamard Transform based features", *39th International Conference on Telecommunications and Signal Processing (TSP)*, 413-416, 2016.
- [9] Young, M., *The Technical Writer's Handbook*, Mill Valley, CA: University Science, 1989.
- [10] Lin, S. C., I. Chang Y. C. and Yang W. N., "Meta-learning for imbalanced data and classification ensemble in binary classification", *Neurocomputing*, 73(1): 484-494, 2009.
- [11] Shanks, J. L. "Computation of the fast Walsh-Fourier transform", *IEEE Transactions on Computers*, 100(5): 457-459, 1969.
- [12] Breiman, L. "Random forests," *Machine learning*, 45(1):5-32, 2001
- [13] Pal M., "Random forest classifier for remote sensing classification", *International Journal Of Remote Sensing*, 26(1): 217-222, 2005.
- [14] Watts J. D., Powell S. L., Lawrence R. L., Hilker T., "Improved classification of conservation tillage adoption using high temporal and synthetic satellite imagery", *Remote Sensing of Environment*, 115(1): 66-75, 2011.
- [15] Breiman, Leo. "Manual on setting up, using, and understanding random forests v3. 1." *Statistics Department University of California Berkeley*, CA, USA, 2002.
- [16] Darvishi, S. and Al-Ani, A., "Brain-Computer interface analysis using continuous wavelet transform and adaptive Neuro-Fuzzy classifier", *2007 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1(16):3220-3223, 2007.
- [17] Tavakolian, K. vd., "Mental task classification for brain computer interface applications", *Canadian Student Conference on Biomedical Computing*. 2006.
- [18] Khalid, M.B., Rao, N.I., Rizwan-i-Haque, I., vd., "A Brain Computer Interface (BCI) using Fractional Fourier Transform with Time Domain Normalization and Heuristic Weight Adjustment", *9th International Conference On Signal Processing*, 1(5):2731-2734, 2008.