

Elektrokardiyografi Sinyallerinde Deneysel Mod Ayrıştırma Ve Geliştirilmiş Karar Ağaçları Kullanarak Aritmi Tespiti

Özgür TOMAK^{1*}

¹Giresun Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Elektrik ve Enerji Bölümü, Giresun, Türkiye

*Sorumlu Yazar: ozgur.tomak@giresun.edu.tr

Geliş Tarihi: 29.03.2019

Kabul Tarihi: 23.05.2019

Öz

Günümüzde ölüme neden olan kalp hastalıkları yaygınlaşmıştır. Elektrokardiyografi bu hastalıkların teşhis aşamasında sıkça kullanılan biyomedikal bir sinyaldir. Bu çalışmada, EKG incelemesi sonucunda aritmiyi saptamada kullanılacak bir teknik önerilmiştir. Aritmiyi tespit için, Deneysel Mod Ayrıştırma ve de Tekil Değerlere Ayrıştırma kullanıldı. Deneysel Mod Ayrıştırma durağan, doğrusal olmayan serileri analiz için uygun bir tekniktir ve yerel düzeyindeki salınım sinyallerini kullanır. Sinyalleri, İç Mod Fonksiyonları adındaki salınım yapılarına ayırır. Tekil Değerlere Ayrıştırma ise karmaşık veri setlerinin boyutlarını küçültülmede kullanılan bir cebirsel yöntemdir ve gürültü etkilerini azaltmada kullanılmıştır. Gürültünün etkisinin azaltılmasından ve uygun öz niteliklerin elde edilmesinden sonra, Sınıflandırma, Geliştirilmiş Karar Ağaçları kullanılarak yapıldı. Sınıflandırmanın performansını değerlendirmede doğruluk, duyarlılık ve özgünlük değerleri hesaplandı.

Anahtar Kelimeler: Elektrokardiyografi (EKG), Deneysel Mod Ayrıştırma, Tekil Değerlere Ayrıştırma, Geliştirilmiş Karar Ağaçları, Aritmi.

Arrhythmia Detection Using Empirical Mode Decomposition and Boosted Trees in Electrocardiography Signals

Abstract

Nowadays, heart diseases that cause death have become widespread. Electrocardiography is a biomedical signal commonly used in the diagnosis of these diseases. In this study, a technique which can be used for detecting arrhythmia as a result of ECG examination is proposed. In order to detect arrhythmia, Empirical Mode Decomposition and Singular Value Decomposition were used. Empirical Mode Decomposition is an appropriate technique for analysis of the stationary, non-linear series and uses oscillation signals at the local levels. It separates the signals into oscillation structures called Intrinsic Mode Functions. Singular Value Decomposition is an algebraic method used to reduce the size of complex data sets and is used to reduce noise effects. After reducing the effect of noise and obtaining the appropriate features, the classification was made by using Boosted Trees. Accuracy, sensitivity, and specificity values were calculated to evaluate the performance of the classification.

Keywords: Electrocardiography (ECG), Empirical Mode Decomposition, Singular Value Decomposition, Boosted Trees, Arrhythmia.

1. Giriş

Aritmi tespitini genelde tecrübelerine dayanarak doktorlar yapmaktadır ancak bugünlerde akıllı saatler gibi donanımlar ve gelişmiş yazılımlar kullanarak doktora gitmeye gerek kalmaksızın bazı aritmi türlerini yüksek doğrulukla tespit edilebilmektedir. Bu alanda yapılan pek çok çalışma bulunmaktadır. Deneysel Mod Ayırıştırma (DMA) tekniği EKG sinyallerindeki gürültüyü kaldırmada kullanılabilir. Bu tarz çalışmalardan birisinde Weng ve arkadaşları DMA tabanlı bir yöntem ile yüksek frekanslı gürültüyü minimum sinyal bozulması ile kaldırmayı hedeflenmiştir (Weng ve ark., 2006). Yöntem, MIT-BIH veri tabanlarının kullanıldığı deneylerle doğrulanmıştır. Blanco-Velasco ve arkadaşları DMA yönteminden yararlanan yeni bir EKG geliştirme tekniği önerdiler (Blanco ve ark., 2008). Önerilen teknik, yüksek frekanslı gürültüyü ve taban sapmasını minimum sinyal bozulması ile kaldırabilmektedir. Teknik, MIT-BIH veri tabanlarının kullanıldığı deneylerle doğrulanmıştır. Slimane ve Naït-Ali, QRS bölgesinin tespiti amacıyla yeni bir DMA tabanlı bir algoritma sundular (Slimane ve Naït-Ali, 2010). Bu algoritma yüksek geçişli bir filtre, sinyalin DMA ayırıştırması, doğrusal olmayan bir dönüşüm, bir entegrasyon ve son olarak düşük geçişli bir filtre kullanır. Önerilen algoritmanın yüksek algılama performanslarına ulaştığı görülmüştür. Labate ve arkadaşları tek kanallı EKG'yi işleyerek solunum sinyalinin dalga biçimini yeniden yapılandırmışlardır (Labate ve ark., 2013). Bu hedeflere ulaşmak için, DMA ve dalgacık analizi gibi EKG sinyalinin uygun fonksiyon tabanlarına ayrıştırılması için iki teknik kullandılar. Sonuçlar ayrıca her iki algoritmanın da solunum dalgası biçimini yeniden yapılandırabildiğini göstermektedir, bununla birlikte DMA önceden seçilmiş bir fonksiyona ihtiyaç duymadan sinyali ayırıştırabilmektedir. Dalgacık analizinde ise bir ana dalgacık seçilmesi gerekmektedir. Pal ve Mitra, DMA tekniğini kullanan EKG sinyal geliştirme ayrıca QRS tespit algoritmasını geliştirmişler (Pal ve Mitra, 2017). DMA tekniği ayırıştırma çıkışı olarak İç Mod Fonksiyonlarını (İMF) üreten tam uyarlamalı bir sinyal ayırıştırma tekniğidir. QRS kompleksini ifade eden bir dizi İMF seçilir ve QRS geliştirmesi nedeniyle doğrusal olmayan bir dönüşüm yapılmasıyla, sürecin algılamanın doğruluğunu artırdığı tespit edildi. DMA genellikle EEG sinyallerinin sınıflandırılmasında kullanılmıştır. Bu çalışmalardan birisinde Bajaj ve Pachori İMF'lerden hesaplanan genlik modülasyon bant genişliği ve frekans modülasyon bant genişliği bilgisini destek vektör makine yöntemine giriş olarak verip nöbet geçiren ve geçirmeyen hastaları sınıflandırdılar (Bajaj ve Pachori, 2012). Uygun İMF seçimi ve Morlet dalgacığını kullanarak % 99.5-100 (minimum- maksimum) başarı elde ettiler.

2. Materyal ve Metot

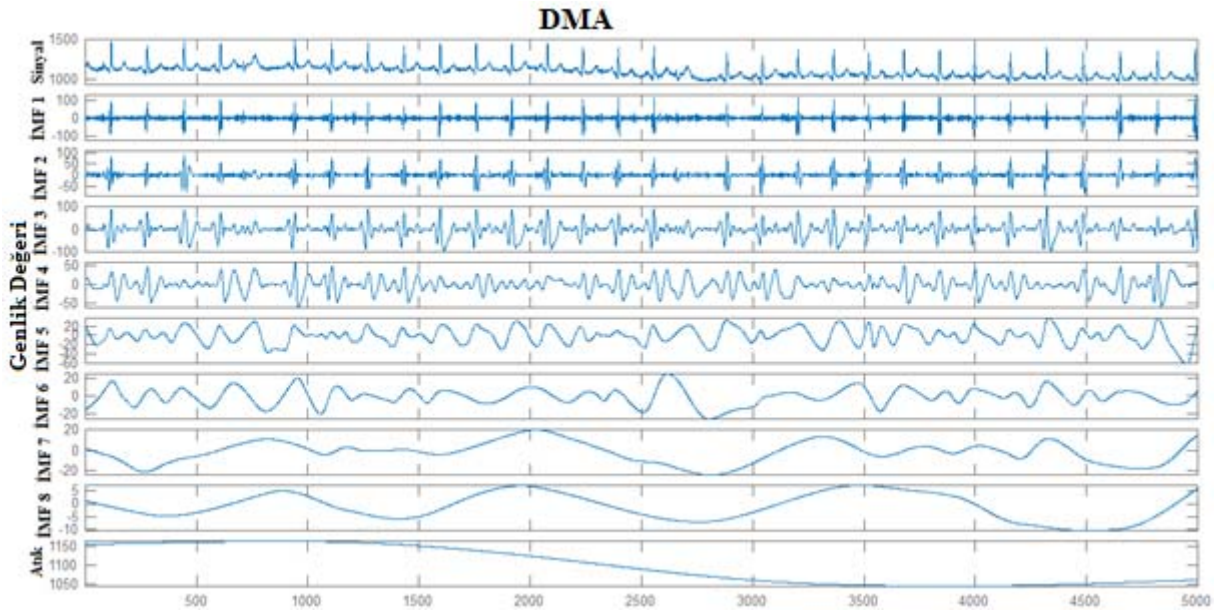
Yapılan çalışmada aritminin tespiti için DMA kullanılarak öznelikler elde edilmektedir. Tekil Değerlere Ayırıştırma (TDA) ile gürültü etkilerini azaltmaktadır. Sınıflandırma aşamasında ise Geliştirilmiş Karar Ağaçları kullanılmıştır. ST-Petersburg Enstitüsü Kardiyolojik veri tabanı analizler için tercih edilmiştir. Bu veri tabanı 32 adet Holter kaydı ve 75 açıklama içeren kayıt dosyasından meydana gelmektedir. Her kayıt 257 Hz'de örneklenmiştir ve 30 dakika süresindeki 12 kanal içermektedir (Goldberger ve ark., 2000).

2.1. Deneysel Mod Ayırıştırma

DMA durağan ve doğrusal olmayan serileri analiz etmede kullanılan uygun bir tekniktir ve yerel salınım düzeyindeki salınım sinyalleri kullanılmaktadır. Sinyaller, İMF'leri bazı salınım yapılarına ayrıştırılır (McDonald ve ark., 2007). DMA algoritmasında ilk olarak giriş sinyalinde yerel tepe noktaları bulunur. Yerel maksimum noktasındaki 3. derece bir eğriyi alarak üst sarma eğrisi oluşturulur. Yerel minimum noktasındaki 3. derece bir eğriyi alarak alt sarma eğrisi oluşturulur. Üst ve alt sarma eğrilerine ait ortalamalar hesaplanır. Bu ortalamalar giriş sinyalinden çıkartılarak sinyale ait düşük frekanslı bileşen ortadan kaldırılmaktadır. Elde ettiğimiz sinyal İMF ise işlem sonlandırılır. Eğer İMF değilse işlemler yeni sinyal için tekrarlanır. Sinyalin İMF olup olmadığını anlamak için İMF koşullarına bakarız. İMF koşullarının ilkinde sinyale ait sıfır geçişleri tepe sayısına eşit veya bir fazladır. İkinci koşulda yerel minimum ve de maksimum tarafından belirlenmiş sarma eğrileri simetrik olur. Bu tekniğin önemli avantajlarından birisi de, üretilen İMF'ler Hilbert dönüşümüyle geçici uzay verilerinden zaman frekans alanına dönüştürülebilir olmasıdır. Bir $x(t)$ fonksiyona ait Hilbert dönüşümü, $y(t)$, aşağıda denklem 1'deki gibi yapılabilir (McDonald ve ark., 2007).

$$y(t) = H[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \quad (1)$$

Hilbert spektrum grafiği için, sinyalin İMF'lerine ihtiyacımız vardır. İMF'lerini ve sinyalin atıklarını hesaplamada DMA kullanıldı. İnterpolasyon yöntemi olarak pchip tercih edildi. Örnek ayırıştırma Şekil 1'de verilmektedir.



Şekil 1. Deneysel Mod Ayrıştırma ve İç Mod Fonksiyonları

2.2. Tekil Değerlere Ayrıştırma

TDA, karmaşık veri setimizin boyutunda küçültülme için uygulanmış bir cebirsel tekniktir. TDA'nın asıl uygulamaları matrisin temel alt uzaylarını bulmak, normu hesaplamak, sözde ters matris hesaplaması, düşük derece yaklaşık yapının elde edilmesidir (Lagerlund ve ark., 1997). Bu teknik sayesinde eğitim ve sınıflandırmada kullanılan süre kısalmaktadır. Her $M * N$ boyutlu A matrisi denklem 2'deki gibi ayrıştırılabilir.

$$A = UDV^T \quad (2)$$

D , köşegen, gerçek ve negatif olmayan bir $N \times N$ matris, U , AA^T özvektörleri tarafından hesaplanan ortogonal bir $M \times M$ matris ve V , AA^T özvektörleri tarafından hesaplanan ortogonal bir $N \times N$ matrisdir.

2.3. Sınıflandırma

Geliştirilmiş Karar Ağaçları kullanırken tercih ettiğimiz topluluk yöntemi AdaBoostM1'dir. AdaBoost'un eğitim sürecinde, modelde performansı artıran özellikler seçilerek yürütmenin süresini geliştirebilir (Freund ve Schapire, 1997). Algoritma, sırayla öğrenmektedir. İndeks t ile her öğrenci için, AdaBoostM1 ağırlıklı sınıflandırma hatasını hesaplar ve denklem 3 ile verilmiştir.

$$\varepsilon_t = \sum_{n=1}^N d_n^{(t)} I(y_n \neq h_t(x_n)) \quad (3)$$

x_n , n gözlem için belirleyici değerlerin bir vektörüdür.

y_n , gerçek sınıf etiketidir.

h_t , indeks t ile öğrencinin (hipotez) tahminidir.

I , gösterge fonksiyonudur.

$d_n^{(t)}$, t adımında n gözleminin ağırlığıdır.

AdaBoostM1, daha sonra, öğrenci t tarafından yanlış sınıflandırılan gözlemler için ağırlıkları artırır ve öğrenci t tarafından doğru şekilde sınıflandırılan gözlemler için ağırlıkları azaltır. Bir sonraki öğrenci t + 1 daha sonra güncellenmiş ağırlıklar ile veriler üzerinde eğitilir.

Eğitim bittikten sonra, AdaBoostM1 kullanarak yeni veri için sınıf tahmini yapar. Bu süreç denklem 4 ve 5’de verilmiştir.

$$f(x) = \sum_{t=1}^T \alpha_t h_t(x) \quad (4)$$

$$\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t} \quad (5)$$

AdaBoostM1 tarafından yapılan eğitim, üstel kaydın kademeli olarak en aza indirilmesi olarak görülebilir ve indirilmesi gereken fonksiyon denklem 6’da verilmiştir.

$$\sum_{n=1}^N w_n \exp(-y_n f(x_n)) \quad (6)$$

Burada y_n gerçek sınıf etiketidir.

w_n normalize edilmiş gözlem ağırlıklarıdır.

$f(x_n)$ tahmini sınıflandırma skorudur.

Çoğu zaman AdaBoostM1 sığ ağaçlarla kullanılır. Öğrenme oranını daha düşük bir sayıya ayarlarsanız, topluluk daha yavaş bir hızda öğrenir, ancak daha iyi bir çözüme yaklaşabilir. 0.1,

öğrenme oranı için popüler bir seçimdir. 1'den az bir oranda öğrenmek genellikle büzülme olarak adlandırılır.

Sahte kayıpları yorumlamak, sınıflandırma hatasından daha zordur, ancak fikir aynıdır. Sözde kayıp, bir topluluktaki herhangi bir öğrencinin sınıflandırma doğruluğu ölçüsü olarak kullanılabilir. Sözde kayıp genellikle AdaBoostM1 için ağırlıklı sınıflandırma hatasıyla aynı davranışı gösterir. Güçlendirilmiş bir topluluktaki ilk birkaç öğrenci düşük yalancı kayıp değerleri verir. İlk birkaç eğitim aşamasından sonra, topluluk daha yavaş bir tempoda öğrenmeye başlar ve kayıp değeri 0,5'ten aşağıya doğru yaklaşır (Freund ve Schapire, 1997).

Uygulanabilirlik, hafızanın az kullanımı, artan tahmin hızından dolayı tercih edilmektedir. Fakat etkili bir tahmin için, fazla miktarda topluluk üyelerine ihtiyaç duyabilir. Öğrenme şekli olarak karar ağacı kullanıldı. Karar ağacı, hızlı sınıflandırma için tercih edilebilecek en iyi seçeneklerden birisidir. Yorumlanması kolaydır. Maksimum bölünme sayısını kullanarak ağacın derinliğini belirleyebiliriz. Model esnekliği öğrenen sayısının artmasıyla artar. Fakat öğrenen sayısını artarsa modelin sınıflandırmakta kullandığı zaman ihtiyacı artacaktır.

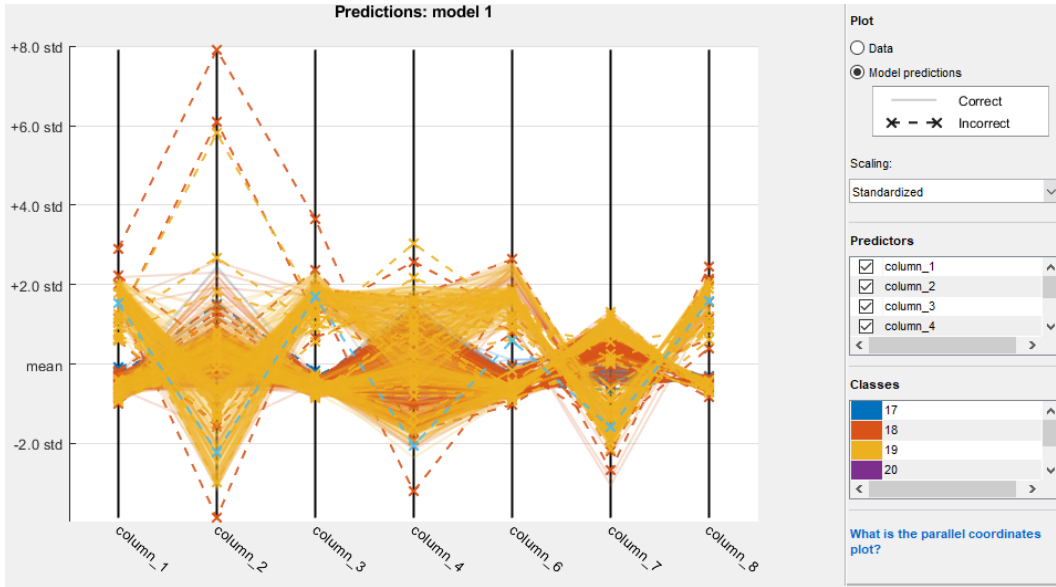
Sınıflandırma yönteminin performansını değerlendirmede doğruluk, duyarlılık ve özgünlük, kullanılmıştır. Doğruluk(D), özgünlük(Ö) ve hassaslık(H) matematiksel ifadeleri Denklem 7-9'da verilmektedir. Burada TP gerçek pozitif değeri, FP yanlış pozitif değeri, TN gerçek negatif değeri ve FN yanlış negatif değeri ifade etmektedir(Tomak ve Kayıkçıoğlu, 2018).

$$H = \frac{TP}{TP+FN} \quad (7)$$

$$\text{Ö} = \frac{TN}{TN+FP} \quad (8)$$

$$D = \frac{TP+TN}{TP+FP+FN+TN} \quad (9)$$

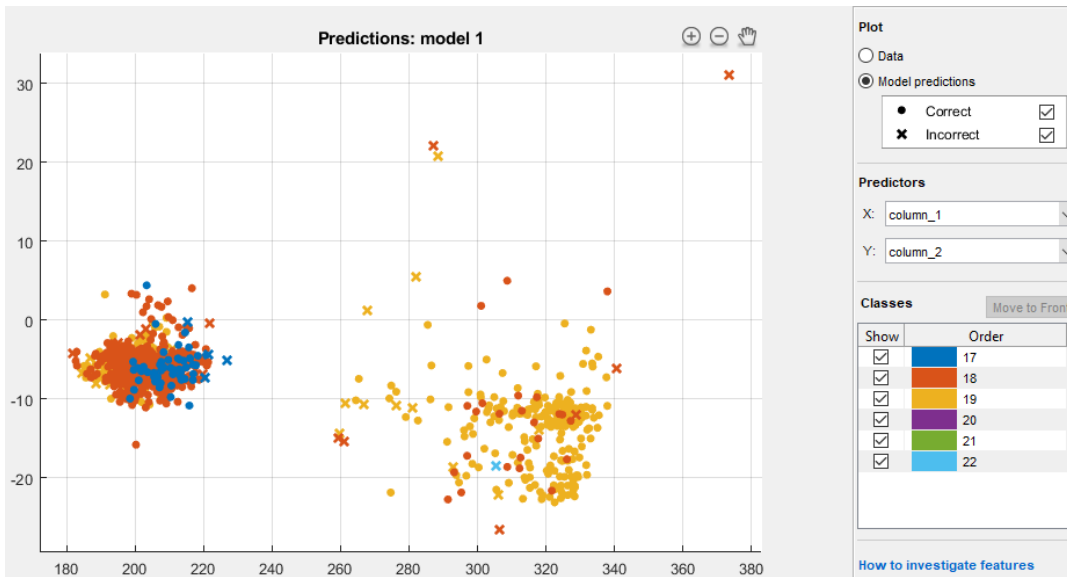
Paralel Koordinat Çizimleri Şekil 2'de verilmiştir. Böylece birçok değişkeni birlikte karşılaştırabiliriz ve aralarındaki ilişkileri görebiliriz. Paralel Koordinat Çizimlerinde, her bir değişkene kendi eksenini veririz ve tüm eksenleri birbirine paralel olarak yerleştiririz. Her bir eksen farklı bir ölçü biriminde çalıştığından her eksen farklı bir ölçek alabilir veya tüm ölçekleri eşit tutmak için tüm eksenler normalleştirilebiliriz. Değerler, tüm eksenlere bağlanan bir dizi çizgi olarak çizilir.



Şekil 2. Verilere ait Paralel Koordinat Çizimleri.

3. Bulgular ve Tartışma

ST-Petersburg Enstitüsü Kardiyolojik veri tabanından alınan 175497 kalp atışını içeren sinyaller rastgele ve eşit sayıda eğitim ve test kümelerine ayrılır ve sonra önışleme öznitelik çıkarımı ve sınıflandırma yapılmaktadır. Eğitim kümesindeki verilerimiz aritmi içeren ve de içermeyen veriler şeklinde etiketlendi. Öznitelikler elde edilirken DMA, gürültüyü bastırmada TDA, sınıflandırmada ise Geliştirilmiş Karar Ağaçları kullanıldı. MATLAB kullanılarak yapılan simülasyonlar sonucunda % 92.47'lık doğruluk % 95.71 duyarlılık ve % 88.62 özgünlük elde edildi. Bu verilerin dağılımı ve sınıflandırmasına ait dağılım grafiği Şekil 3'de verilmiştir.



Şekil 3. Verilerin dağılımı ve sınıflandırmasına ait dağılım grafiği

4. Sonuçlar ve Öneriler

DMA ile öznelikleri oluşturup Geliştirilmiş Karar Ağaçları ile sınıflandırma yaptığımız sisteme ait sonuçlar başarılıdır ve geliştirilecek aritmi tespit sistemlerinde kullanılabilir. TDA kullanımı gürültü etkilerini azaltmanın yanında eğitim ve sınıflandırmada kullanılan süreyi kısaltmaktadır. Benzer çalışmalarda DMA özellik olarak genelde EEG sinyallerini sınıflandırmada kullanılmıştır ve başarılı sonuçlar elde edilmiştir mesela Bajaj ve Pachori 2012’de uygun parametre seçimiyle %100’e ulaşan başarı elde etmişlerdir ancak bu tekniğin EKG sinyallerinin sınıflandırmasında kullanıldığı bir çalışmaya rastlanılmamıştır. EKG sinyallerini sınıflandırmada kullanılan diğer yöntemlere yakın başarı seviyelerine sahiptir ve TDA kullanımı ile başarıyı ve süre konusunda iyileşme sağlanmıştır. Algoritmanın optimizasyonu süreyi daha da kısaltabilecektir böylece gerçek zamanlı analizler için uygun hale getirilebilir

Kaynaklar

- Bajaj, V., & Pachori, R. B. (2012). Classification of seizure and nonseizure EEG signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine*, 16(6), 1135-1142.
- Blanco-Velasco, M., Weng, B., & Barner, K. E. (2008). ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in biology and medicine*, 38(1), 1-13.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).
- Labate, D., La Foresta, F., Occhiuto, G., Morabito, F. C., Lay-Ekuakille, A., & Vergallo, P. (2013). Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison. *IEEE Sensors Journal*, 13(7), 2666-2674.
- Lagerlund, T. D., Sharbrough, F. W., & Busacker, N. E. (1997). Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of clinical neurophysiology*, 14(1), 73-82.
- McDonald, A. J., Baumgaertner, A. J. G., Fraser, G. J., George, S. E., & Marsh, S. (2007, March). Empirical Mode Decomposition of the atmospheric wave field. In *Annales Geophysicae* (Vol. 25, No. 2, pp. 375-384).
- Pal, S., & Mitra, M. (2012). Empirical mode decomposition based ECG enhancement and QRS detection. *Computers in biology and medicine*, 42(1), 83-92.
- Tomak, Ö., & Kayıkçıoğlu, T. (2018). Bagged tree classification of arrhythmia using wavelets for denoising, compression, and feature extraction. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(3), 1555-1571.
- Weng, B., Blanco-Velasco, M., & Barner, K. E. (2006, August). ECG denoising based on the empirical mode decomposition. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 1-4). IEEE. (Lagerlund ve ark., 1997). (Tomak ve Kayıkçıoğlu, 2018)