



# Aritmi Vuru Tespiti için Morfolojik Tabanlı Öznitelik Çıkarma

## Morphologic Based Feature Extraction for Arrhythmia Beat Detection

Merve DOGRUYOL BASAR<sup>1</sup>, Soner KOTAN<sup>1,2</sup>, Niyazi KILIÇ<sup>3</sup>, Aydın AKAN<sup>3,4</sup>

<sup>1</sup> Biyomedikal Mühendisliği Anabilim Dalı, İstanbul Üniversitesi, İstanbul, Türkiye  
mervedogruiolbasar@gmail.com

<sup>2</sup> Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü, Marmara Üniversitesi, İstanbul, Türkiye  
sonerkotan@marmara.edu.tr

<sup>3</sup> Elektrik – Elektronik Mühendisliği Bölümü, İstanbul Üniversitesi, İstanbul, Türkiye  
{niyazik, akan}@istanbul.edu.tr

<sup>4</sup> Biyomedikal Mühendisliği Bölümü, İzmir Katip Çelebi Üniversitesi, İzmir, Türkiye  
aydin.akan@ikc.edu.tr

**Özetçe**— Kalp hastalıkları günümüzde ölüm oranlarının yüksek olduğu hastalıkların başında gelmektedir. Kalp hastalıkları hakkında bilgi edinmede kalbin elektriksel aktivitesinin incelenmesi ve yorumlanması önemli yer tutmaktadır. Bu çalışmada hasta bir bireye ait elektrokardiyogram işaretleri incelenmiş olup, hastanın sağlıklı ve aritmili vuruları çıkarılmıştır. İncelenen ritimlerde vuruların RR, QRS, Çarpıklık ve Doğrusal Öngörü katsayıları göz önüne alınarak sınıflandırma yapılmıştır. Çalışmada K-NN, Rastgele alt uzaylar, Naive Bayes ve K-Star sınıflandırıcıları kullanılmıştır. K-NN algoritması %98.32 ile en yüksek doğruluk değerini vermiştir. İkinci aşama olarak K-NN algoritmasındaki 'k' parametresi değiştirilerek doğruluk değerleri yorumlanmıştır.

**Anahtar Kelimeler** – EKG, makine öğrenmesi, kalp hastalığı, öznitelik çıkarma

**Abstract**— Heart disease is one of the diseases which has highest mortality rate recently. Heart's electrical activity examination and interpretation are very important for the understanding of diseases. In this study, electrocardiogram signals are analyzed, then patient's healthy and arrhythmia beats are extracted. RR, QRS, Skewness and Linear Predictive Coding coefficients of the signals are considered for classification of the data. K-NN, Random SubSpaces, Naive Bayes and K-Star classifiers are used. The highest accuracy is obtained with the K-NN algorithm (98.32%). At the second stage of the K-NN algorithm, accuracy levels are examined by changing the 'k' parameter.

**Keywords** — ECG, machine learning, heart disease, feature extraction

### I. GİRİŞ

Kalp hastalıkları günümüzde en yüksek ölüm oranına sahip olan hastalıklardan biridir [1]. Her yıl yaklaşık 17 milyon ölüm kardiyovasküler hastalıklar (CVD) ve kalp krizleri sebebiyle olmaktadır. Dünya Sağlık örgütünden (WHO) alınan istatistikler incelendiğinde bu sayının yaklaşık 16.7 milyon kadarının CVD sebebiyle olduğu ortaya konmaktadır. Tahminlere göre 2020 yılı içinde yaklaşık 20 milyon, 2030 yılında ise yaklaşık 24 milyon bireyin CVD ve kalp yetersizliği sebepleriyle yaşamlarının sona ermesi tahmin edilmektedir [1-3].

Kalp hastalıklarına sebep olan etkenler arasında ailenin hastalığa olan yatkınlığı (genetik faktörü), yüksek kolesterol, yüksek tansiyon, diyabet ve metabolik sendromlar yer almaktadır [4]. Bu sebeple, erken teşhis her hastalıkta olduğu gibi kalp hastalıklarında da önemli bir yer tutmaktadır. Kalp hastalıklarının teşhisi hakkında en önemli teşhis cihazı elektrokardiyogram (EKG) cihazlarıdır. EKG teşhis cihazları insan vücudu üzerinden elektrotlarla algılanabilen ve alınan biyolojik işaretleri dönüştürücüler sayesinde kaydedip, doktorlara ve hastaya anında verilerin iletilmesini sağlayan medikal cihazlardır. Bu cihazların kullanımı son derece yaygın olup maliyeti yüksek olmayan cihazlardır [5-7].

## Sinyal İşleme 1

1. Gün / 27 Ekim 2016, Perşembe

Kalbin elektriksel aktivitelerini anlamlandırmada kalp dalgaları etkili olmaktadır. Bir kalp siklusundaki (PQRSTU) P, Q, R, S, T ve U dalgaları değerlendirilmektedir. Değerlendirmelerde çeşitli aralıklar ele alınır. Örneğin, ventriküler kasılmalar için QRS kompleksi göz önüne alınmaktayken, atriyum depolarizasyonu için P dalgası, ventrikül repolarizasyonu için T dalgası incelenir. Aynı zamanda bir kalp siklusunda PR, ST segmentleri ve PR, QT intervalleri de ele alınmaktadır [6-8].

Yapılan çalışmalar incelendiğinde EKG verilerini değerlendirmek amacıyla birden fazla sınıflandırma yöntemleri kullanılmıştır. Kullanılan yöntemlere; Srinivas ve diğerlerinin kullandığı Karar ağaçları ve Naive Bayes sınıflandırıcıları [9], Masethe ve diğerleri tarafından kullanılan J48, REPTree karar ağaçları ve Bayes, Naive Bayes sınıflandırıcıları [10], Soni ve diğerlerinin kullandığı k-NN, Naive Bayes algoritmaları ve karar ağaçları [11], Jabbar ve diğerlerinin [12] ve Mert ve diğerlerinin [13] kullandığı k-NN sınıflandırıcıları örnek verilebilmektedir. Literatürdeki sınıflandırmaya dair çalışmalar genel olarak incelendiğinde en çok kullanılan sınıflandırma metodları J48, k-NN, Random SubSpace, Adaboost, Bagging karar ağaçları ve Bayes, Naive Bayes algoritmalarıdır.

Bu çalışmada, sınıflandırma doğruluğunun performansının artırılması için kalp ritimlerinden önemli bilgilerin çıkarılması ve bu bilgilerin sınıflandırıcılara sokulması gerekmektedir. Bunun için kalp ritimlerinden morfolojik tabanlı öznitelikler çıkarılarak aritmi vuru tespiti yapılmaya çalışılmıştır. Tespit için k-NN, Random Subspace, Naive Bayes ve K-Star sınıflandırıcıları kullanılmıştır. k-NN sınıflandırıcısı için çeşitli k değerleri atanarak doğruluk, kappa, ortalama standart hata (MAE) ve karesel ortalama hata (RMSE) oranları hesaplanmıştır.

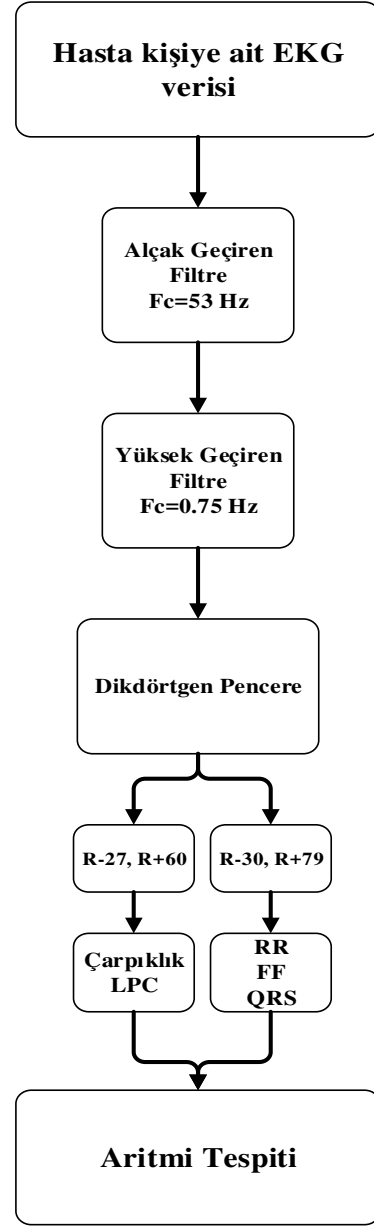
## II. YÖNTEM

Çalışma öznitelik çıkarımı ve çıkarılan veriyi işleme olarak iki kısımdan oluşmuştur. İlk olarak öznitelik çıkarımı için kullanılan veri ön işleme yapılarak hastaya ait sinyal işlenecek hale getirilmiştir. Ön işleme sırasında EKG sinyallerinden 5 öz nitelik çıkarılmıştır. Bu öznitelikler RR, QRS, FF (form faktörü), çarpıklık (Skewness) ve LPC (Doğrusal Öngörü Kodlaması) katsayılarıdır. Elde edilen öznitelikler sınıflandırma işlemi için kullanılmıştır.

Yapılan çalışmanın akış diyagramı Şekil 1.'de gösterilmektedir.

### A. Materyal

Hastaya ait EKG verisi Rangaraj M. Rangayyan sitesi veri tabanından elde edilmiştir [14]. Veri tabanında hastaya ait sinyalleri verileri bulunmaktadır. Hasta verisinde 397 kalp siklusu bulunmaktadır.



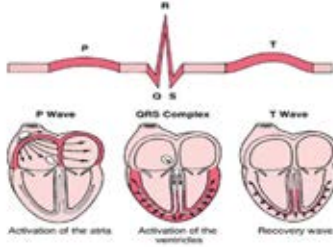
Şekil 1. Çalışma aşamalarını içeren akış diyagramı

Bu kalp sikluslarından 305 tanesi Matlab yazılımı sayesinde hastanın normal siklusları olarak tanımlanırken, 92 tanesi ise hastanın hastalığını içeren siklusları içermektedir. Hastalıklı verilere “hasta”, normal verilere ise “sağlıklı” adı verilmiştir. Hastalık hakkında bilgi veren özniteliklerin hastalıkla bağlantılı olduğu bilinmektedir. Özniteliklerin çıkarılması “Waikato Environment for Knowledge Analysis (WEKA)” makine öğrenmesi yazılımı ile elde edilmiştir.

Normal bir bireye ait kalp siklusu ve bileşenleri Şekil 2. deki gibidir [15].

## Sinyal İşleme 1

1. Gün / 27 Ekim 2016, Perşembe



Şekil 2. Normal bir bireyin EKG' sinin dalga formu

### B. Ön İşleme

Ön işleme aşamasında, MIT-BIH aritmi veri tabanından alınmış ECG sinyali kullanılmıştır. Sinyal, içerisinde ventriküler prematüre atımların da bulunduğu 397 adet kalp atışına sahiptir. Bu atımlar, EKG üzerinde normal atımlara göre daha geniş ve daha yüksek genlikli olarak belirmektedir.

Sinyal, öncelikle kesim frekansı 53 Hz olan alçak geçiren ve kesim frekansı 0.75 Hz olan yüksek geçiren filtreler kullanılarak filtrelenmiştir. Filtrelenmiş sinyalde R noktaları tespit edilerek RR aralıkları hesaplanmıştır. R noktalarından R-27, R+60 olacak şekilde pencereler seçilerek çarpıklık ve LPC; R-30, R+79 olacak şekilde pencereler seçilerek ise FF değerleri hesaplanmıştır.

Sinyale ait her bir QRS dizisi, bir uzman eşliğinde "sağlıklı" ve "hasta" olmak üzere iki sınıfa ayrılmıştır. Elde edilen öznitelikler, yapılan sınıflandırmayla eşleştirilerek eğitilmesi için uygun formatta hazırlanmıştır.

### C. Sınıflandırma

Çalışmada k-NN, Random SubSpace, Naïve Bayes ve K-Star algoritmaları kullanılmıştır. EKG verilerini değerlendirme aşamasında RR aralığı, QRS genişliği, FF aralığı, Çarpıklık ve LPC denklemleri kullanılmıştır.

K-NN (K Nearest Neighborhood-En Yakın K Komşu): Gözetimli bir öğrenme yöntemidir. Sınıflandırılmak istenen verinin en yakın olduğu düşünülen k verisiyle uzaklık ortalamasına ve belirlenen en yakın uzaklıklara göre sınıflara atamaları yapılır. Yöntemin performansını k en yakın komşu sayısı ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir. Atamaları yapmak için çeşitli uzaklık fonksiyonları kullanılmaktadır. Bunlar; Manhattan, Minkowski ve Öklid uzaklık fonksiyonlarıdır. Literatürdeki çalışmalarda genellikle Öklid uzaklık fonksiyonuna göre atama yapılmaktadır [16].

Random SubSpace (Rastgele Alt Uzaylar): Bu algoritma, veri setindeki her veriyi aynı eğitim seti ile eğitmek yerine aynı veri setinin bazı özelliklerini dışarıda tutup yeni veri setleri oluşturarak veriyi rastgele eğitme amacıyla olan bir sınıflandırıcıdır. Doğrusal sınıflandırma işlemlerinde sıklıkla kullanılmaktadır [17].

Naïve Bayes: Uygulanabilirliği yüksek ve kolay bir sınıflandırıcıdır. İstatistiksel yöntemler sayesinde

sınıflandırma yapabilmektedir. Uygulama aşamasında bir takım kabuller yapılmaktadır. En önemli kabulü niteliklerin birbirinden bağımsız olması durumudur. Niteliklerin birbirini etkilemesi durumunda olasılık hesabını yapabilmek zor olmaktadır. Bu sınıflandırıcıda tüm nitelikler eşit derecede önemli kabul edilmektedir. Verilerin mutlaka bir sınıfı bulunmalıdır. Naive Bayes algoritması bit ağırlıklandırma yöntemi ile numerik ve nominal veriler kolaylıkla sınıflandırılabilir [18].

K-Star: Bazı benzerlik fonksiyonlarıyla belirlenebilen örnek tabanlı bir sınıflandırıcıdır. Eğitilen verilerle aynı sınıfa dayandırılır ve test edilen verileri sınıflandırır. Diğer örnek tabanlı verilerden eğiten sınıflandırıcılardan fonksiyonları kullanması yönüyle farklıdır [19].

RR: Bir kalp siklusunun tamamını veren değerdir.

QRS kompleksi: Ventriküllerin depolarizasyonu hakkında bilgi verir. Q dalgası P dalgasından sonraki ilk negatif dalgayı, R dalgası ilk pozitif dalgayı, S dalgası ise R'dan sonraki negatif dalgayı ifade eder.

FF: Sinyalin türevinin mobilitesinin, orijinal sinyalin mobilitesine bölünmesiyle elde edilir (Denklem 1, Denklem 2).

$$FF = \frac{M_{x'}}{M_x} \quad (1)$$

$$M_x = \sqrt{\frac{\sigma^2_{x'}}{\sigma^2_x}} \quad (2)$$

Çarpıklık: Gerçek değerli bir rasgele değişkenin olasılık dağılım fonksiyonunun bir yana yatık oluşu, yani simetrik olmayışı şeklinde tanımlanır (Denklem 3).

$$s = E \left[ \left( \frac{x - \mu}{\sigma} \right)^3 \right] \quad (3)$$

LPC: Orijinal sinyalin önceki değerlerinin doğrusal kombinasyonundan yararlanılarak, sonraki değerlerinin tahmin edilmesi işlemidir.

$$\tilde{x}(n) = \sum_{i=1}^p \alpha_i x(n-i) \quad (4)$$

### III. DENEYSEL SONUÇLAR

Çalışmada veri setinde ön işleme sonucunda elde edilen 5 öznitelik kullanılmıştır. Bu özniteliklere sınıflandırıcıda 10 katlı çapraz doğrulama yapılmış olup, veri seti %70 eğitim ve %30 test aşamasından geçirilmişlerdir.

Karşılaştırma yapılırken doğruluk oranı, kappa değeri, ortalama karesel hata (MAE), karesel ortalama hata (RMSE) değerleri göz önünde bulundurulmuştur.



## Sinyal İşleme 1

1. Gün / 27 Ekim 2016, Perşembe

Öznitelikler çeşitli sınıflandırıcılara sokulmuştur. Elde edilen sonuçlar Tablo 1. deki gibidir.

SINIFLANDIRICILAR		Değerler			
		Doğruluk(%)	Kappa	MAE	RMSE
	IBk (k-NN)	98.3193	0.9478	0.0203	0.1292
	Random SubSpace	97.2292	0.9188	0.1802	0.2231
	Naïve Bayes	96.6387	0.8957	0.055	0.1735
	K-Star	93.7028	0.8211	0.0798	0.2222

**Tablo 1.** Veri setindeki özniteliklerden alınan sınıflandırıcı değerlendirmeleri

Tablo 1. incelendiğinde en yüksek doğruluk değerlerine sahip olan k-NN (IBk) sınıflandırıcısının olduğu görülmüştür.

K-NN sınıflandırıcısı “k” değerleri değiştirilerek sınıflandırıcıdan elde edilen yeni değerler Tablo 2. deki gibidir.

SINIFLANDIRICILAR		k-NN = “k” Değerleri			
		k=1	k=32	k=128	k=256
	Doğruluk(%)	98.3193	97.479	80.672	78.992
	Kappa	0.9478	0.9205	0.1208	0
	MAE	0.0203	0.0768	0.185	0.3333
	RMSE	0.1292	0.1631	0.2843	0.3947

**Tablo 2.** “k” değerlerinin değişiminin incelenmesi

Bu çalışmada k-NN değerlerini hesaplarken Öklid uzaklıkları ele alınmıştır. Tablo 2. yi yorumlarsak; “k” değeri arttıkça ele alınan veri çevresindeki daha fazla veri ile komşuluk kurmak durumunda kalmıştır. Bu sebeple Tablo 2. den de çıkarılacağı gibi verinin doğruluk oranında sayı arttıkça azalma meydana gelmektedir.

## IV. SONUÇLAR

Çalışmada, sınıflandırma doğruluğunun performansını arttırılmak için kalp ritimlerinden bilgiler çıkarılmış olup bu bilgilerin verileri sınıflandırıcılara sokulmuştur. Sınıflandırıcıların doğruluk kıyaslamaları yapıldığında en yüksek analiz değeri %98.31 ile k-NN sınıflandırıcısı olmuştur.

Çalışmanın devamında k-NN sınıflandırıcısındaki ‘k’ değerleri arttırılarak doğruluk yüzdesine etkisi incelenmiş olup doğruluk başarısı k değerleri arttıkça azalmıştır. Doğruluk değerleri ‘k’ nın 4 değeri için %98.32’ den %78,9’ a kadar azalmıştır. Bu durum bize verinin en yakın komşusuna olan uzaklığının ve veri miktarının arttıkça elde

edilen doğruluk ve kappa değerlerinin azaldığını göstermektedir.

## KAYNAKÇA

- [1] Acampora, G., Lee, C.S., Vitiello, A., Wang, M.H., “Evaluating cardiac health through semantic soft computing techniques”, *Soft Comput*, 16:1165–1181 DOI 10.1007/s00500-011-0792-2, 2012.
- [2] Chairperson, J.J.V. et al., “ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure”, *European Journal of Heart Failure*, Volume 14, Issue 8, pages 803–869, 2012.
- [3] “International Statistical Classification of Diseases and Related Health Problems; Heart Diseases”, *World Health Organization (WHO)*, ISBN 92 4 154653, Volume 2.
- [4] Mozaffarian, D. et al., “Heart Disease and Stroke Statistics, *American Heart Association, Circulation*, 133:000-000, 2016.
- [5] Acton, A., “Heart Diseases-Advances in Research and Treatment”, ScholarlyEditions, GA, 2013.
- [6] Rangayyan, R.M., “Biomedical Signal Analysis, A Case-Study Approach”, *IEEE Press*, ISBN 0-471-2081 1-6, 2002.
- [7] Rautaharju, P.M., “ST-T Waveform Features, QT and Mortality Risk”, *The Female Electrocardiogram* Chapter 9, pp 87-107, 2015.
- [8] Hall, J.E., Guyton, A.C., “Guyton and Hall Textbook of Medical Physiology”, *Elsevier*, 13th Edition, 2016.
- [9] Srinivas, K. et. al., Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks, *International Journal on Computer Science and Engineering*, Vol. 02, No. 02, 250-255, 2010.
- [10] Masethe, H.D. et al., “Prediction of Heart Disease using Classification Algorithms”, *Proceedings of the World Congress on Engineering and Computer Science*, Vol II, USA, 2014.
- [11] Soni, J. et. al., “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, *International Journal of Computer Applications*, (0975 – 8887) Volume 17– No.8, 2011.
- [12] Jabbar, M.A. et. al., “Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm”, *First International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA)*, Volume 10, Pages 85–94, 2013.
- [13] Mert A., Kiliç N., Akan A., “Geliştirilmiş Morfolojik Öznitelikler Kullanarak Ekg Sinyal Analizi”, *Akıllı sistemlerde Yenilikler ve Uygulamaları Sempozyumu*, Türkiye, ss.234-237, 2012.
- [14] [https://people.ece.cornell.edu/land/courses/ece5030/labs/f2009/EC\\_G\\_analysis.html](https://people.ece.cornell.edu/land/courses/ece5030/labs/f2009/EC_G_analysis.html), erişim tarihi: 09.05.16
- [15] [http://www.homepages.ucl.ac.uk/~zchab6a/v2/ssm\\_yr1\\_1/diagnosi\\_s\\_ecg.htm](http://www.homepages.ucl.ac.uk/~zchab6a/v2/ssm_yr1_1/diagnosi_s_ecg.htm), erişim tarihi: 09.05.16
- [16] Garcia, V. et. al., “Fast k Nearest Neighbor Search using GPU”, *Computer Vision and Pattern Recognition, CVGPU*, 2008.
- [17] Ho, T.K., “The random subspace method for constructing decision forests”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume:20 , Issue: 8, Page(s): 832 – 844, 1998.
- [18] Wang, Z. et. al., “An EEG workload classifier for multiple subjects”, *Conf Proc IEEE Eng Med Biol Soc.*, doi: 10.1109/IEMBS.2011.6091612, 2011.
- [19] Inbarani, H.H. et. al., “Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis”, *Computer Methods and Programs in Biomedicine*, Volume 113, Issue 1, Pages 175–185, 2014.