

Makine öğrenmesi algoritmaları ve dalgacık dönüşümü ile EKG sinyalinden özellik çıkarımı

Hülya KODAL SEVİNDİR^{1,*}, Süleyman ÇETİNKAYA¹, Cüneyt YAZICI²

¹Kocaeli Üniversitesi, Fen Edebiyat Fakültesi, Matematik Bölümü, Kocaeli.

²Kocaeli Üniversitesi, Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitim Bölümü, Kocaeli.

Geliş Tarihi (Received Date): 24.07.2017

Kabul Tarihi (Accepted Date): 13.03.2018

Özet

Günümüzde biyomedikal sinyallerin analizinde dalgacık dönüşümünün kullanılması oldukça yaygın olup elde edilen sonuçlar etkileyicidir. Bu çalışmada, biyomedikal sinyallerden elektrokardiyogram (EKG) sinyallerinde QRS zirvesi belirleme hedeflenmiş ve daha iyi sonuçlar almak için öncelikle EKG sinyallerindeki zemin gezinme gürültüsünün giderilmesi ve yüksek frekanslı gürültünün temizlenmesi amacıyla dalgacık analizi kullanılmıştır. Daubechies 10 (db10) dalgacık dönüşümü uygulanan sinyalin 10. seviye yaklaşım katsayısı ve 10. seviye detay katsayısı çıkartılarak sinyaldeki zemin gezinmesi problemi giderilmiştir. Yüksek frekans gürültüsünün giderilmesi için ise zemin gezinmesi problemi giderilmiş olan sinyale dalgacık gürültü temizleme uygulanmıştır. Gürültüsü temizlenen sinyalde QRS zirvelerini belirlemek için sinyalin 1. türev ve 2. türev bilgileri ele alınarak Destek Vektör Makineleri ve Naive Bayes algoritmaları kullanılmıştır. QRS zirvelerinin bulunmasında, MIT-BIH aritmi veri tabanında verilen QRS zirvelerinin konum bilgileri kullanılmıştır. QRS zirvelerini doğru belirlemede Destek Vektör Makineleri algoritması Naive Bayes algoritmasından daha yavaş sonuç vermesine rağmen %99.46 duyarlılık, %100 seçicilik ve %0.54 hata değerlerine ulaşmıştır.

Anahtar kelimeler: EKG, Makine Öğrenmesi, Destek Vektör Makineleri, Naive Bayes, Gürültü Giderme, Dalgacık Dönüşümü, QRS zirve.

*Hülya KODAL SEVİNDİR, hkodal@kocaeli.edu.tr, <https://orcid.org/0000-0003-1460-0608>

Süleyman ÇETİNKAYA, suleyman.cetinkaya@kocaeli.edu.tr, <https://orcid.org/0000-0002-8214-5099>

Cüneyt YAZICI, cuneyt.yazici@kocaeli.edu.tr, <https://orcid.org/0000-0002-4535-510X>

ECG feature extraction via wavelet transform and machine learning algorithms

Abstract

Nowadays usage of wavelet transform on biomedical signals has been increased and effective results have been obtained. In this study wavelet transform is used to obtain better results on QRS complex detection; wavelets are used to get rid of baseline wandering and high frequency artifact on ECG. To do so, some ECG recordings on MIT-BIH arrhythmia database are used. To omit baseline wandering Daubechies wavelet up to level 10 is used and approximation and detail coefficients at level 10 are excluded from the ECG signal. To omit high frequency artifact wavelet denoising is applied to the ECG signals. Afterward, using the first and second derivative information of the signal, Support Vector Machine and Naive Bayes algorithms are applied separately. According to the study conducted, although SVM algorithm runs slower than Naive Bayes, results for SVM are much better for QRS detection; the results for SVM are %99.46 sensitivity, %100 positive sensitivity, and %0.54 error.

Keywords: ECG, Machine Learning, Support Vector Machine, Naive Bayes, Denoising, Wavelet Transform.

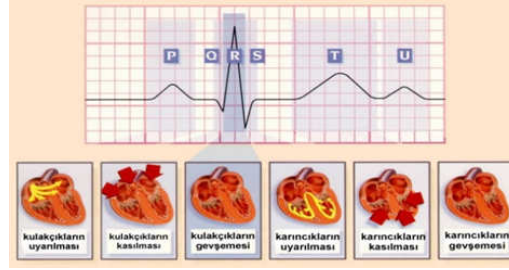
1. Giriş

Kalp insan hayatı için hayati öneme sahip bir organ olup kalp rahatsızlıklarının önceden tespit edilmesi oldukça önemlidir. Zira kalp hastalıkları dünyada hala bir numaralı ölüm sebebidir. Özellikle genç nüfustaki kalp krizi daha ölümcül olmaktadır [1]. Dolayısıyla olası kalp rahatsızlığının erken teşhisinin konulması büyük önem arz etmektedir. Bunun için en sık kullanılan yöntem ise kalbin Elektrokardiyogram (EKG) ile incelenmesidir. EKG, kalbin biyoelektriksel aktivitesini ölçer. EKG sinyali, kalp atımında kalp kaslarının depolarizasyonu sonucu vücut yüzeyinde oluşan çok düşük seviyedeki elektriksel voltajların sonucudur. Düşük maliyetli, vücuda herhangi bir zarar vermeyen, basit ve kalp rahatsızlıklarının teşhisinde etkili bir ölçüm aracıdır [2]. EKG'nin incelenmesi ile

- kalbin ritim ve iletim bozuklukları belirlenebilir,
 - koroner yetmezliği veya enfarktüs tanısı konulabilir,
 - kalp duvarlarında kalınlaşma ve kalp boşluklarında genişleme bulguları saptanabilir,
 - elektronik kalp pilinin işlevleri değerlendirilebilir,
 - bazı kalp ilaçlarının etkileri, elektrolit dengesizliği (özellikle serum potasyum eksikliği veya fazlalığı) ve kalp dışı hastalıkların kalbe etkileri araştırılabilir [2, 3, 4, 5].
- EKG dalga şekilleri miyokardiyal iskemi ve enfarktüs, ventriküler hipertrofi ve aktarım problemleri gibi kalp hastalıkları ve anormallikleri durumunda değişir. EKG analizi için kullanılan bilgisayarlı karar destek sistemleri teşhis sırasında kalp uzmanına bir yardımcı gibi davranarak tanıda yardımcı bilgi ve zaman tasarrufu sağlayabilir. Hasta yoğun bakım ünitesinde sürekli gözetim altındayken doktor normal kalp atış değişikliğinin aralığını belirleyebilir. Eğer nabız düzensizliği olursa doktoru ya da hemşireyi uyarmak için bir alarm sistemi kurulabilir. Bu ve benzeri uygulamalar için

EKG sinyalinin yapısındaki sivrilikler ve eğimler incelenir. Özellik çıkarımı kardiyolojik inceleme için de faydalıdır.

Şekil 1, bir kalp atımı sırasında gerçekleşen P, QRS, T ve U dalgalarını gösterir [6].



Şekil 1. EKG sinyali örneği.

Diğer biyomedikal sinyaller gibi EKG de durağan olmayan bir yapıya sahiptir. Bu nedenle klasik sinyal işleme yöntemleri yeterince etkili olamayabilir. Ayrıca EKG kaydı sırasında herhangi bir kanalda görülen istenmeyen elektriksel aktivitelere, artefaktlar ya da gürültüler, yer alabilir. Artefaktlar fizyolojik ya da non-fizyolojik nedenlerden dolayı olabilirler. Fizyolojik artefakt, vücudun diğer organlarından kaynaklanan biyoelektriksel aktivitelere, hasta elini oynattığında veya öksürdüğünde, tremor veya titremeye bağlı hareketler vs. durumunda oluşabilir. Non-fizyolojik artefakt ise elektrokardiyogramın kendi iç devresi veya bağlı olduğu elektrik şebekesinden kaynaklanır. Cihaza bağlı metal, hareket ve birçok sebepten dolayı sinyalin artefakt içermesi yani bozuk-parazitli alınması mümkündür. EKG'nin doğru okunması ve yorumlanması büyük önem arz etmektedir. Birçok çalışmada bu artefaktların temizlenmesi için sinyal bazı önlemlere tabi tutulur. Bu çalışmada ise EKG sinyalindeki zemin gezinmesi ve yüksek frekans gürültüsünün giderilmesi için dalgacık analizi uygulanmıştır. Bu şekilde EKG özellik çıkarımının çok daha doğru bir şekilde yapılması hedeflenmiştir. Dalgacık metodu, zaman ve frekans bilgisini kaybetmeksizin durağan olmayan EKG sinyalinin analizini yapmaya olanak sağladığı için seçilmiştir. Sonrasında gürültüden arındırılmış sinyal için ayrı ayrı sinyalin 1. türev ve 2. türev bilgileri de ele alınarak Destek Vektör Makineleri ve Naive Bayes algoritmaları ile QRS zirveleri belirlenmesi yapılmıştır. Bölüm 2'de dalgacık dönüşümü ve uygulanacak 2 algoritma hakkında kısa bilgiler verilmiştir. Bölüm 3'te kullanılan veri ve yöntem hakkında bilgi verilmiş ve son bölümde sonuçlar paylaşılmıştır.

2. Dalgacık dönüşümü ve algoritmalar

2.1. Dalgacık dönüşümü

Dalgacık dönüşümleri zaman-frekans optimizasyonu sağlayan ve EKG işaretindeki geçici karakteristiklerin tespitinde kullanılabilir etkili bir yöntemdir. Dalgacık dönüşümü, $f(t)$ işaretin temel dalgacık fonksiyon kümesi ile integrali olarak tanımlanabilir. Temel dalgacık fonksiyon kümesi $\psi_{a,b}(t)$, temel dalgacığın ölçeklenmesi ve ötelenmesiyle elde edilir [7].

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a > 0, a, b \in \mathbb{R}. \quad (1)$$

Burada a ölçekleme b ise ötelemeyi temsil eden değişkenlerdir. $f(t)$ işaretinin dalgacık dönüşümü şu formülle verilmiştir:

$$W_t(a, b) = \int_{-\infty}^{\infty} f(t) \psi_{a,b}(t) dt \quad (2)$$

Sürekli Dalgacık Dönüşümü (SDD) ölçekleme ve öteleme değişkenlerinin sürekli değerler aldığı dönüşüm olarak tanımlanabilir. EKG işaretlerinin analizinde bu oldukça zaman alan ve hesapsal verimliliği düşüren bir yaklaşımdır. Ayrık Dalgacık Dönüşümünde (ADD) ise ölçekleme ve öteleme değişkenleri ayrık değerler alırlar [7]. Böylece işaret belirli ölçeklerde incelenir ve hesapsal verimlilik artırılmış olur. Bu yöntemle çoklu çözünürlük analizi adı verilir. Bu durumda ayrık dalgacık kümesi

$$\psi_{j,k}(x) = 2^{-j/2} \psi(2^{-j}x - k) \quad j, k \in \mathbb{Z} \quad (3)$$

olarak bulunabilir. Burada j ve k sırasıyla ölçekleme ve ötelemeyi gösteren tam sayılardır. Ayrık dalgacık fonksiyonları $\psi_{j,k}(x)$ ortogonal bir küme oluşturur. Ayrık dalgacık dönüşümü katsayıları ise

$$c_{j,k} = 2^{j/2} \int_{-\infty}^{\infty} f(x) \psi(2^j x - k) dx \quad (4)$$

ile bulunur. Elde edilen dalgacık katsayıları yardımıyla geri dönüşüm formülü ise aşağıdaki şekilde elde edilebilir:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(x). \quad (5)$$

Dalgacık dönüşümünde işaret, ayrıştırılırken ölçekleme ve dalgacık fonksiyonları cinsinden ifade edilebilir. Ölçekleme fonksiyonu $\phi(x)$ ve dalgacık fonksiyonu $\psi(x)$ aşağıdaki şekilde tanımlanabilir. Burada $g(k) = (-1)^k h(1-k)$ olmak üzere süzgeç katsayılarıdır.

$$\begin{aligned} \phi(x) &= \sqrt{2} \sum_k h(k) \phi(2x - k), \\ \psi(x) &= \sqrt{2} \sum_k g(k) \phi(2x - k). \end{aligned} \quad (6)$$

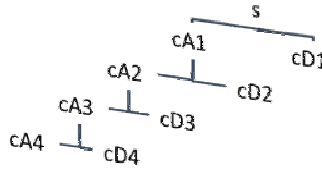
Çoklu çözünürlük analizinde J seviyeli ayrıştırma işleminde işaret $f(x)$ için aynı süzgeç katsayıları ile yinelemeli olarak dönüşüm katsayıları bulunur.

$$f(x) = \sum_k c_{j_0,k} \phi_{j_0,k}(k) + \sum_{j=j_0}^J \sum_k d_{j,k} \psi_{j,k}(x) \quad (7)$$

Burada $J \geq j \geq j_0$ olmak üzere j_0 keyfi bir başlangıç ölçüğüdür. j seviyesindeki $c_{j,k}$ (yaklaşıklık) ve $d_{j,k}$ (detay) katsayılarını elde etmek için, $j-1$ seviyesindeki $c_{j-1,k}$ katsayılarıyla alçak ve yüksek geçiren süzgeçler evriştirilmiştir. Örnek sayısı ise alt örnekleme ile iki kat azaltılmıştır.

$$\begin{aligned} c_{j,n} &= \sum_k c_{j-1,k} h(k-2n) \\ d_{j,n} &= \sum_k c_{j-1,k} g(k-2n) \end{aligned} \quad (8)$$

Dalgacık dönüşümünde elde edilen düşük boyutlu yaklaşıklık katsayıları işaretin sonlu elemanlı yaklaşıklığını verir. Dalgacık analizinde ise sadece alçak geçiren bantlar değil aynı zamanda yüksek geçiren bantlar da hesaba katılır. Böylece elde edilen tüm alt bantlarda (yaklaşıklık katsayılarında) yinelemeli olarak alçak ve yüksek geçiren süzgeçler yardımıyla dalgacık dönüşümü katsayıları elde edilir. Yaklaşım katsayıları haricinde detay katsayılarının da hesaba katılması ile işaretin daha ayrıntılı bir biçimde analizine olanak sağlanır [7]. 4-seviyeli bir dalgacık paket ağaç yapısı Şekil 1’de gösterilmiştir.



Şekil 2. Dört seviyeli ayrık dalgacık dönüşümünün ağaç yapısı.

Şekil 2’de cA_n , n . seviye yaklaşıklık katsayısını ve cD_n ise n . seviye detay katsayısını göstermektedir. Özetle dalgacık metodu kullanarak bir sinyalin ayrıştırılması ve tekrar oluşturulması genel olarak üç aşamadan oluşur [8].

- 1) Ayrık dalgacık dönüşümü kullanarak sinyalin bileşenlerine ayrılması,
- 2) Ortaya çıkan katsayıların eşikleme metoduyla yumuşatılması,
- 3) Eşikleme yapılmış ayrık dalgacık katsayılarından tekrar orijinal sinyalin oluşturulması [9].

2.2. Destek vektör makineleri

Destek Vektör Makineleri (DVM) öğrenme sisteminde yer alan bir modeldir. Vapnik tarafından geliştirilen DVM tekniği başlangıçta iki sınıfın sınıflandırma problemleri için önerildi [10]. DVM, eğitim verilerinden optimal ayırma aşırı düzlemini hesaplamak için geometrik özellikleri kullanır. Ayrıca, lineer olmayan ayrılabilir durumların için de metotlar sundular.

Temel olarak DVM, n boyutlu x girdi vektörünün bir K -boyutlu özellik uzayına $K > n$ lineer olmayan bir $\varphi(x)$ dönüşümü ile oluşturulan yüksek boyutlu özellik uzayında çalışan lineer bir işlemdir. $\varphi_0(x) = 1$, $w = [w_0, w_1, \dots, w_k]^T$ ve $\varphi(x) = [\varphi_0(x), \varphi_1(x), \dots, \varphi_k(x)]^T$ ağırlık fonksiyonu olmak üzere denklem (9), iki farklı sınıfı ayıran aşırı düzlem denklemini verir:

$$y(x) = \varphi_0(x)\varphi(x) = \sum_{j=1}^K w_j \varphi_j(x) + w_0 = 0 \quad (9)$$

$y(x) > 0$ koşulunu sağlayan bir sınıfa ait iken $y(x) < 0$ koşulunu sağlayan ise diğer sınıfa aittir.

DVM ile ilgili en belirgin gerçek, Lagrange çarpanları kullanılarak öğrenme aşamasının kuadratik programlamaya indirgenmesidir. DVM de yer alan öğrenme ve test aşamasındaki tüm işlemler, çekirdek fonksiyonları kullanılarak yapılır. Çekirdek fonksiyonu $K(x, x_i) = \varphi^T(x_i)\varphi(x)$ şeklinde tanımlanır.

DVM öğrenme aşamasında maksimal ayırma marjiniyle hedef değerlerin ya $d_i = 1$ ya da $d_i = -1$ diye iki sınıfa ayrılma problemi aşağıdaki gibi tanımlanan amaç fonksiyonunun dual maksimizasyon problemine indirgenmektedir. C kullanıcı tanımlı bir sabit ve $p, (x_i, d_i)$ öğrenme veri çiftlerinin sayısı olmak üzere

$$\sum_{i=1}^p \alpha_i d_i = 0 \quad (10)$$

$$0 \leq \alpha_i \leq C$$

kısıtlamasıyla

$$Q(\alpha) = \sum_{i=1}^p \alpha_i - \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j d_i d_j K(x_i, x_j)$$

dır. C regularizasyon parametresidir, marjinin maksimizasyonu ile sınıflandırma hatasının minimizasyonu arasındaki dengeyi belirler. Langrange çarpanlarına göre çözüm w_{opt} optimal ağırlık fonksiyonunu aşağıdaki gibi verir:

$$w_{opt} = \sum_{i=1}^{N_s} \alpha_{si} d_{si} \varphi(x_{si}) \quad (12)$$

(12) denklemde s, N_s destek vektörlerinin kümesini göstermektedir yani x_i öğrenme vektörleri

$$d_i \left(\sum_{j=1}^K w_j \varphi_j(x_i) + w_0 \right) \geq 1 - \mu_i \quad (13)$$

bağıntısı için eşitlik işareti ile yerine getirilir. μ_i skaler değişkenlerdir. Öğrenme aşamasından sonra denklem (14) de $y(x)$ çıkış sinyali çekirdeklerin fonksiyonu olarak belirlenir.

$$y(x) = \sum_{i=1}^{N_s} \alpha_{si} d_i K(x_{si}, x) + w_0 \quad (14)$$

ve $\varphi(x)$ lineer olmayan fonksiyonun açık formunun bilinmesine gerek yoktur. $y(x)$ 'in değeri sıfırdan büyükse +1 sınıfına ait iken sıfırdan küçükse -1 sınıfına aittir. Bu çalışmada öznitelik kullanılan çekirdek fonksiyonu denklem (15) deki gibidir:

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0 \quad (15)$$

Burada γ Gaussian radyal taban fonksiyonlarının serbest parametresidir.

2.2. Naive Bayes sınıflandırıcı

Naive Bayes sınıflandırıcısı Bayes teoremi üzerine kurulan basit Bayesian sınıflandırıcısı olarak adlandırılır [11]. Sınıflandırıcı tipik olarak basit bir Bayesian bağlantısıdır ve giriş vektörlerinin boyutu yüksek olduğu durumlarda kullanılması çok uygundur. Bayes kuralı, sınıflandırma değişkeninin posteriorunu hesaplamak için kullanılır. C sınıflandırma değişkeninin posterioru denklem (16) de verildiği gibi Z_1, Z_2, \dots, Z_n özellik değişkenlerine bağlıdır.

$$P(C|Z_1, Z_2, \dots, Z_n) = \frac{P(C) \prod_{i=1}^n P(Z_i|C)}{P(Z_1, Z_2, \dots, Z_n)} \quad (16)$$

Sınıf koşullarından bağımsız olduğu varsayılırsa, denklem (16), denklem (17) gibi yazılabilir.

$$P(Z_1, Z_2, \dots, Z_n|C) = P(Z_1|C)P(Z_2|C)\dots P(Z_n|C) \quad (17)$$

Denklem (17) den anlaşılacağı gibi, ortak koşullu olasılık, tüm marjinal koşullu olasılıkların çarpımı tarafından verilir. İkinci varsayım, Z_1, Z_2, \dots, Z_n özelliklerinin doğrudan C sınıflandırma değişkenine bağlı olduğu varsayımdır. Yukarıdaki iki varsayım Bayesian ağları sınıflandırıcısına bağlı düşünüldüğü zaman Naive Bayes sınıflandırıcısı denklem (18) olarak elde edilir.

$$P(C|Z_1, Z_2, \dots, Z_n) = \frac{P(C) \prod_{i=1}^n P(Z_i|C)}{P(Z_1, Z_2, \dots, Z_n)} \quad (18)$$

Bu nedenle, her örneğin sınıfını tahmin etmek için, model denklem (19) gibi yazılabilir.

$$C = \arg \max P(C) \prod_{i=1}^n P(Z_i|C) \quad (19)$$

Denklem (19), her bir örneğin sınıfını tahmin etmek için kullanılabilir.

3. Kullanılan EKG verileri ve yöntem

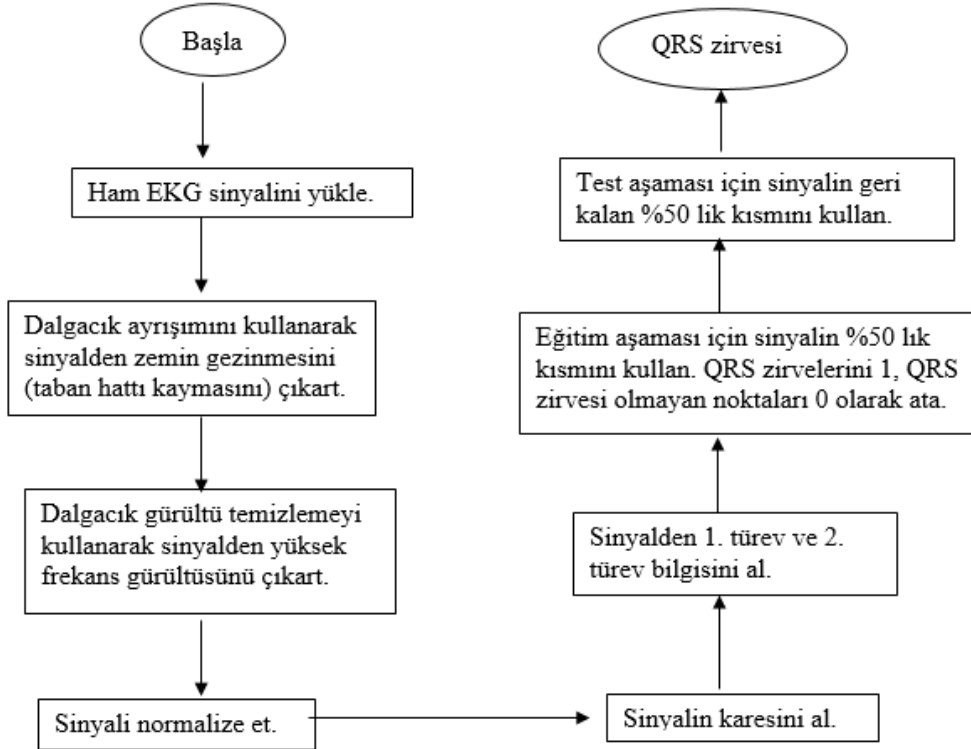
Bu bölümde bu çalışmada kullanılan medikal veriler ve yöntem açıklanmıştır.

3.1. Kullanılan medikal veri

Algoritmayı test etmek için kullanılan EKG sinyalleri Physionet MIT-BIH aritmi veri tabanından alınmıştır [12]. MIT-BIH aritmi veri tabanında 47 farklı kişiye ait her biri yaklaşık 30 dakika olarak çift kanallı kayıt edilmiş EKG sinyalleri mevcuttur. Her veri, bir başlık dosyası (.hea), bir binary dosyası (.dat) ve bir binary açıklama dosyasından (.atr) oluşmaktadır. Başlık dosyası örnekleme sayısı ve frekansı, EKG sinyalinin formatı, hastanın geçmişi ve detaylı klinik bilgisi vs. bilgilerini içermektedir. Bu çalışmada veri tabanından seçilen 15 verinin 10 saniyelik kısımları QRS zirvelerini belirlemek için kullanılmıştır.

3.2. Kullanılan yöntem

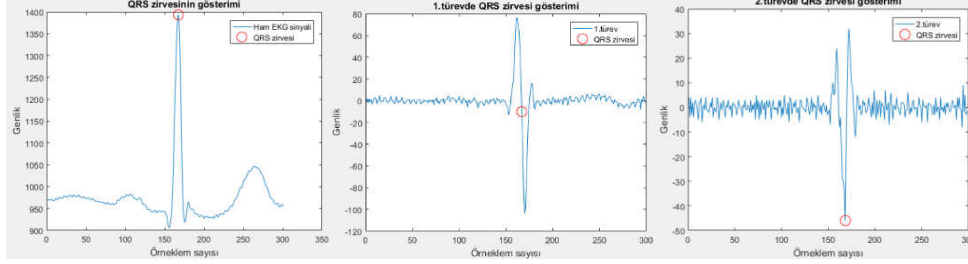
Kalp sinyalinin izo-elektrik temeli çeşitli anormallikler için değişebilir ve çeşitli gürültüler sinyale karışabilir [13]. Bu çalışmada EKG sinyalinde zemin gezinmesi ve gürültü temizleme problemlerinin giderilmesi için dalgacık dönüşümü kullanılmıştır. Gürültüsü temizlenmiş sinyal için, 1. ve 2. türev bilgilerinin de dahil edilmesi ile, ayrı ayrı Destek Vektör Makineleri ve Naive Bayes algoritmaları ile QRS zirveleri belirlenmesi yapılmıştır. Yapılan çalışmada kullanılan algoritmaya ilişkin akış diyagramı Şekil 3' de görülmektedir.



Şekil 3. Önerilen algoritmaya ilişkin akış diyagramı.

Bir sinyalden gürültüleri çıkarma için dalgacıkları kullanmak, gürültü içeren bileşen ya da bileşenleri belirlemeyi ve o bileşenler olmaksızın sinyali yeniden oluşturmayı gerektirir. Bir sinyal, ayrık dalgacık dönüşümü kullanılarak ayrıştırıldığı zaman, sinyali yüksek frekans ve alçak frekans bileşenlerine ayıran detay ve yaklaşıklık sinyalleri elde

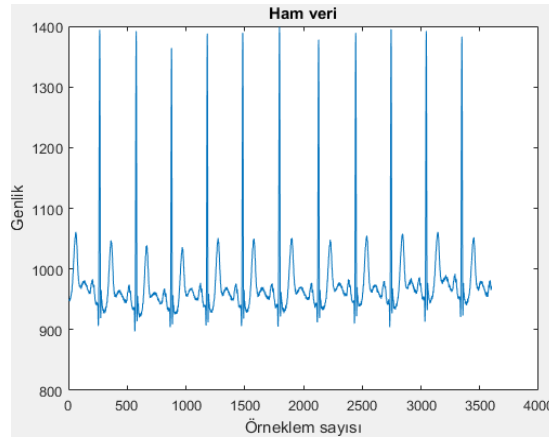
edilir. Dalgacık ağaç dönüşümünde sinyalin alçak frekans bilgisini içeren yaklaşık sinyal tekrar aynı şekilde ayrıştırılır. Bu şekilde elde edilen yaklaşık sinyalde yüksek frekans bileşenleri olmadığı için yüksek frekans gürültüsü giderilir. Yüksek frekans bilgisinin dışarı çıkartılması aşamasında orijinal sinyallerin belirgin özellikleri de kaybedilebilmektedir. Optimal gürültü temizleme, eşikleme olarak adlandırılan bir yaklaşımla yapılabilir. Eşikleme sadece belirli limiti aşan detayların kısımlarını göz ardı etmekten oluşur [2].



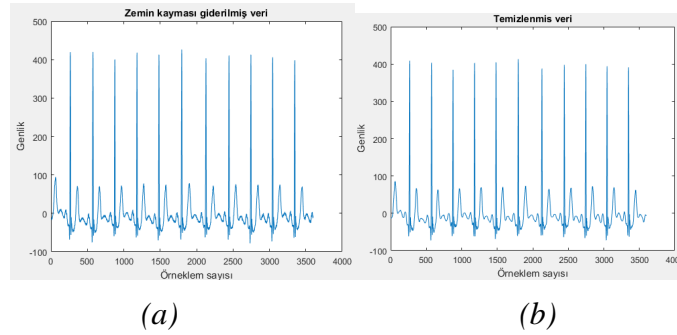
Şekil 4. Physionet MIT-BIH aritmi veri tabanında yer alan 103 no.lu EKG verisinden bir bölüm.

Şekil 4’te soldan sağa ham veri üzerinde QRS zirvesinin, ham verinin 1. türevi üzerinde QRS zirvesinin ve son olarak ham verinin 2. türevi üzerinde QRS zirvesinin hangi konumda yer aldığını göstermek için çizdirilmiştir. Şekiller incelendiği zaman QRS zirvesinin matematiksel olarak sinyalde yer alan diğer noktalardan farklı olduğu düşünülmüştür. Dikkat edilirse 1. türev üzerinde QRS zirvesinin yer aldığı noktanın genliği sıfıra yakındır ve 2. türev üzerinde QRS zirvesinin olduğu noktanın genliği ise negatif yönde sinyalde yer alan diğer noktalardan gözle görülür şekilde ayrılmaktadır. Bu gözlemden sonra Destek Vektör Makineleri ve Naive Bayes algoritmalarında kullanılmak üzere 1. türev ve 2. türev bilgilerinin QRS zirvesi belirlenmesinde bize avantaj sağlayacağı düşünülmüştür. Önerilen algoritmada önceki çalışmalardan farklı olarak 1. türev yanında 2. türev bilgisi de kullanılmıştır.

İncelenen verilerden 103 no.lu ham verinin 10 saniyelik kısmı Şekil 5’de gösterilmektedir. Zemin gezinmesi düşük frekans bandında kendini göstermektedir. Bu düşük frekans bandındaki bileşeni gidermek için dalgacık ayrışım paketindeki analiz yöntemi yardımıyla 10. seviyeye ayrıştırılmış sinyalden 10. seviye yaklaşım katsayısı ve 10. seviye detay katsayısı çıkartılmıştır (Şekil 6a). Yüksek frekans gürültüsünün giderilmesi için ise zemin gezinmesi problemi giderilmiş olan sinyale dalgacık gürültü temizleme uygulanmıştır. (Şekil 6b).

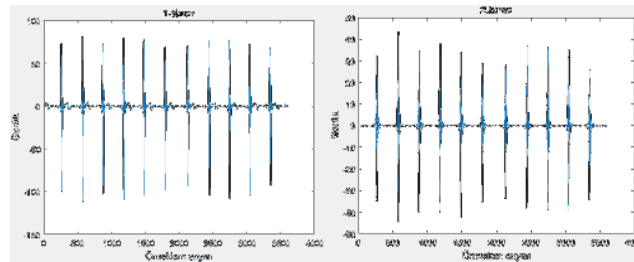


Şekil 5. Physionet MIT-BIH aritmi veri tabanında yer alan 103 no.lu EKG verisinden bir bölüm.



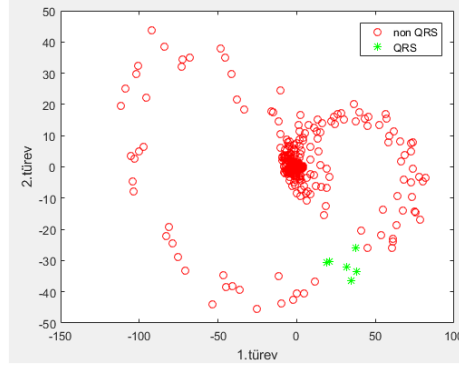
Şekil 6. 103 no.lu ham veri için (a) zemin gezinmesinin giderilmiş durumu ve (b) yüksek frekans gürültüsünün giderilmiş durumu.

Gürültü temizleme sonrası elde edilen sinyal normalize edildikten sonra QRS zirvelerini sinyal içerisinde daha belirgin hale getirmek için sinyalin karesi alınmıştır ve böylece son sinyal QRS zirvelerini gösteren yüksek genlikli sivri uçları içermektedir. Son sinyalin 1.türev ve 2. türev görüntüleri Şekil 7’de görülmektedir.



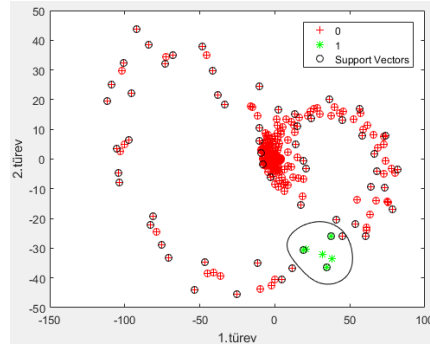
Şekil 7. Gürültüden temizlenmiş 103 no.lu sinyalin 1. ve 2. türev görüntüleri.

Destek vektör makineleri ve Naive Bayes algoritmalarında kullanılmak üzere 1. türev ve 2. türev bilgileri sinyalden alınmaktadır. Şekil 8’de yeşil renkte olan noktalar QRS zirvelerini, kırmızı renkte olan noktalar ise QRS zirvesi olmayan noktaları göstermektedir.



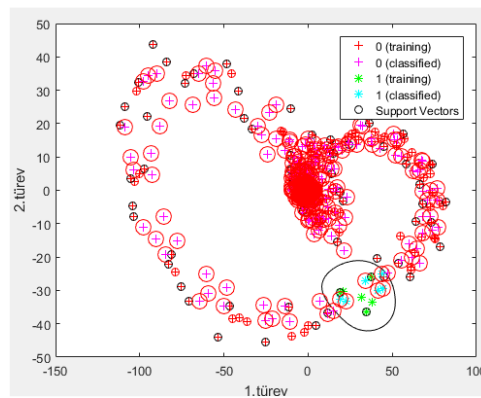
Şekil 8. 1. türev ve 2. türev bilgilerine bakılarak QRS zirvelerinin gösterimi.

Ele aldığımız verinin %50 lik kısmı Destek Vektör makineleri ve Naive Bayes algoritmalarında eğitim aşaması için kullanılmıştır. Eğitim aşamasında QRS zirvesi olan noktalar 1, QRS zirvesi olmayan noktalar 0 olarak atanmıştır. Eğitim aşaması sonrası elde edilen görüntü Şekil 9’da görülmektedir.



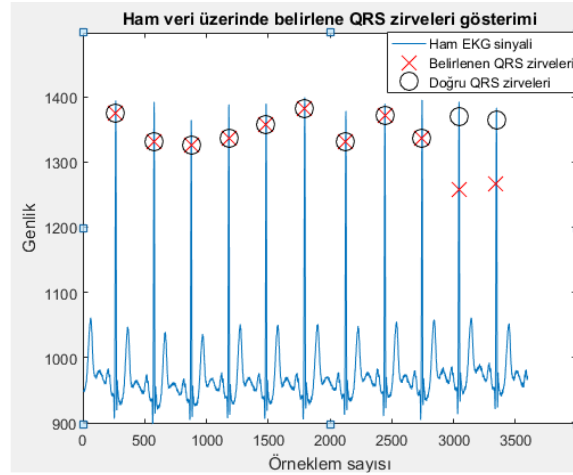
Şekil 9. Destek Vektör Makineleri algoritması eğitim aşaması sonrası elde edilen görüntü.

Eğitim aşaması tamamlandıktan sonra ele alınan verinin geri kalan %50 lik kısmı bu sefer test aşamasında kullanılmıştır. Test aşaması sonrası elde edilen görüntü Şekil 10’da gösterilmektedir.



Şekil 10. Destek Vektör Makineleri algoritması test aşaması sonrası elde edilen görüntü.

Test aşaması sonrası elde edilen sonuçlarda bazı QRS zirvesi olmayan noktalar da 1 olarak (QRS zirvesi olarak) tahmin edilmiştir. Bu durum QRS zirvesine yakın bulunan noktaların hemen hemen QRS zirvesiyle aynı özelliklere sahip olmasından kaynaklanmaktadır. Bu durumu çözmek için eğitim verisinde yer alan QRS zirveleri gerçekleşme süresi ortalaması belirlenmiştir. Test verisinde de yaklaşık bu ortalama sürede gerçekleşmeyen QRS zirveleri elenmiştir. Böylece kontrole hazır QRS zirveleri vektörü oluşturulmuştur. Algoritma uygulandıktan sonra elde edilen QRS zirveleri Şekil 11’de gösterilmiştir.



Şekil 11. Destek Vektör Makineleri algoritması 103 no.lu veriye uygulandıktan sonra elde edilen QRS zirvelerinin ham veri üzerinde gösterimi.

4. Sonuçlar ve tartışma

Bu çalışmada kullanılan EKG verileri MIT-BIH aritmi veri tabanından elde edilmiştir. Verilerin ön işlenmesi (ham veriyi gürültülerden arındırma) ve makine öğrenmesi algoritmaları için MATLAB 2016b paket programı kullanılmıştır. Bulunan QRS zirvesi zamanlaması ile veri tabanındaki etiketlenme anı arasındaki zaman farkı 0,07 sn’ den az ise QRS zirvesi doğru tespit edilmiş olarak değerlendirilmiştir. TP sınıflandırıcının doğru olarak sınıflandırdığı QRS zirveleri sayısı, FP sınıflandırıcının yanlış olarak sınıflandırdığı QRS zirveleri sayısı, FN ise sınıflandırıcının belirleyemediği QRS zirveleri sayısıdır. Destek Vektör makineleri ve Naive Bayes algoritmalarının performansı aşağıdaki başarımlar ölçütlerine göre belirlenmiştir ve sırasıyla sonuçlar Tablo 1 ve Tablo 2 de gösterilmiştir ayrıca tablolarda algoritmaların gerçekleşme süreleri de yer almaktadır.

$$\text{Duyarlılık} = 100 * \frac{TP}{TP + FN}$$

$$\text{Seçicilik} = 100 * \frac{TP}{TP + FP}$$

$$\text{Hata} = \frac{FP + FN}{TP + FN}$$

MIT-BIH aritmi veri tabanından alınarak incelenen 15 kayıt, toplam 185 vuru içermektedir. Sinyalin 1. ve 2. turev bilgilerinin dahil edilmesi durumunda Tablo 1’de

Destek Vektör Makineleri algoritması ve Tablo 2'de ise Naive Bayes algoritması kullanıldığı zaman elde edilen QRS zirveleri belirleme oranları gösterilmektedir.

Tablo 1. MIT-BIH veri tabanından alınan verilerde Destek Vektör Makineleri algoritması uygulandığı zaman elde edilen başarı oranı.

Veri no	Etiketlenmiş Atım Sayısı	TP	FN	FP	Duyarlılık (%)	Seçicilik (%)	Hata (%)	Süre (sn)
101	11	11	0	0	100	100	0	5.11
102	12	12	0	0	100	100	0	5.09
103	11	11	0	0	100	100	0	5.02
105	14	14	0	0	100	100	0	5.30
113	9	9	0	0	100	100	0	5.08
116	13	13	0	0	100	100	0	5.19
118	12	12	0	0	100	100	0	5.39
123	8	8	0	0	100	100	0	4.85
124	8	8	0	0	100	100	0	4.57
205	15	15	0	0	100	100	0	5.86
212	15	15	0	0	100	100	0	5.42
213	18	18	0	0	100	100	0	5.86
222	13	13	0	0	100	100	0	5.64
231	11	10	1	0	90.91	100	9.09	5.14
234	15	15	0	0	100	100	0	5.53
Toplam	185	184	1	0	99.46	100	0.54	5.27

Tablo 2. MIT-BIH veri tabanından alınan verilerde Naive Bayes algoritması uygulandığı zaman elde edilen başarı oranı.

Veri no	Etiketlenmiş Atım Sayısı	TP	FN	FP	Duyarlılık (%)	Seçicilik (%)	Hata (%)	Süre (sn)
101	11	10	1	0	90.91	100	9.09	2.79
102	12	12	0	0	100	100	0	2.85
103	11	11	0	0	100	100	0	2.71
105	14	14	0	0	100	100	0	2.98
113	9	9	0	0	100	100	0	2.45
116	13	13	0	1	100	92.86	7.69	2.95
118	12	12	0	0	100	100	0	2.55
123	8	8	0	0	100	100	0	2.15
124	8	8	0	0	100	100	0	2.24
205	15	15	0	0	100	100	0	2.78
212	15	15	0	0	100	100	0	2.78
213	18	18	0	1	100	94.74	5.56	3.29
222	13	9	4	0	69.23	100	30.76	2.88
231	11	10	1	0	90.91	100	9.09	2.36
234	15	15	0	0	100	100	0	3.01
Toplam	185	179	6	2	96.76	98.90	4.32	2.71

Tablolarda yer alan sonuçların bütünü göz önüne alındığı zaman Naive Bayes algoritmasından daha hızlı sonuç alınmasına rağmen, %99.46 duyarlılık, %100 seçicilik ve %0.54 hata değerleriyle destek vektör makineleri algoritması daha iyi sonuç vermiştir.

2008 yılında Mehta ve Lingayat, CSE veri tabanından alınan 12 kanallı EKG verisini destek vektör makinelerini kullanarak incelemişler; %99.3 gibi bir QRS zirvesi belirleme oranı elde etmişlerdir [14]. 2010 yılında Sasikala ve Wahidabanu tarafından yapılan diğer bir çalışmada ise %99,89 QRS zirvesi belirleme oranı beyan edilmiştir [15]. 2001 yılında ise Dinh ve arkadaşları EKG verilerini incelemek için en iyi dalgacığın %0,75 hata oranı ile kübik B-spline olduğunu beyan etmişlerdir [16]. Bu çalışmada önerilen algoritmada literatürde belirlenen önceki çalışmalardan farklı olarak 1. türev yanında 2. türev bilgisi de kullanılmıştır. Ayrıca destek vektör makineleri ve Naive-Bayes algoritması ile verilere yapılan uygulama sonuçları, literatürdeki bazı çalışmalar ile de kıyaslanmıştır (Tablo 3). Bu çalışmadaki sunduğumuz sonuçlara göre ele aldığımız veri kümesi üzerinde QRS zirvesi belirleme oranı literatürde daha önce bulunan sonuçlara yakındır. Bu çalışmada en az hesap yükü ile en yüksek başarımın elde edilmesi için hesaplanması kolay 2 öznitelik kullanan sınıflandırma şeması önerilmiştir. QRS zirvesi belirleme işlemi sırasında sinyale ait 1.türev ve 2.türev bilgisi kullanılması, bunlara ek olarak herhangi bir matematiksel öznitelik hesaplama yapılmaması, öznitelik boyutunun literatürde yapılan tüm çalışmaların çoğundan daha az olması ve kullanılan sınıflandırma şemasının basitliği sayesinde önerilen yöntem geliştirilecek olan tele tıp (örneğin cep telefonu veya gömülü sistemler gibi) EKG kayıt ve analiz uygulamaları ile bütünleşik çalışabilir. Dolayısıyla sunulan algoritma esnek ve pratikte uygulanabilir.

Tablo 3. Literatürde yer alan bazı çalışmalar ile kıyaslama.

Referans	Metod	Veri tabanı	Duyarlılık	Seçicilik
Önerilen algoritma	Destek vektör makineleri	MIT-BIH	99.46	100
Önerilen algoritma	Naive Bayes	MIT-BIH	96.76	98.90
Kodal Sevindir ve Ark. [2]	Eşikleme ve dalgacık dönüşümü	MIT-BIH	99.76	-
Mehta & Lingayat [14]	Destek vektör makineleri	CSE	99.30	-
Xia ve Ark. [17]	Dalgacık dönüşümü ve K-ortalama	MIT-BIH	99.72	99.80
Gritzali [18]	Uzunluk ve enerji dönüşümü	CSE	99.60	-
Mehta ve Ark. [19]	K-ortalama	CSE	98.63	-
Chouhan & Mehta [20]	Uyarlanabilir eşikleme	CSE	98.56	-
Mehta & Lingayat [21]	Destek vektör makineleri	CSE	99.93	-
Gayake & Shete [22]	Destek vektör makineleri	MIT-BIH	98.66	99.66
Singh ve Khosla [23]	K en yakın komşuluk	MIT-BIH	99.81	99.86

Kaynaklar

- [1] <https://www.medikalakademi.com.tr/kalp-hastaliklari-neden-genclerde-daha-olumcul/>, (20.06.2017).
- [2] Kodal Sevindir, H., Cetinkaya, S. ve Sayli, O., Wavelet transform based noise removal from ECG signal for accurate heart rate detection using ECG, **Medical Technologies National Conference (TIPTEKNO)**, Muğla, (2015)
- [3] Yanık, H. ve Değirmenci, E., Detection of ECG characteristic points using multiresolution analysis, **Sinyal işleme ve iletişim uygulamaları (SIU)**, Malatya, 383-386, (2015).
- [4] Jiang, X. ve Zhang, L., ECG arrhythmias recognition system based on independent component analysis feature extraction, **IEEE Region 10 Conference**, Hong Kong, (2006).
- [5] http://tinaztepehastanesi.com.tr/saglik_kosesi/kardiyoloji/ekg-nedir, (18 Mart 2015).
- [6] <http://www.slideshare.net/husam685/ekg-ritim-ve-pace-maker>, (25 Kasım 2016)
- [7] Uslu, E. ve Bilgin, G., Dalgacık ve birleşik dalgacık paket dönüşümü kullanarak kalp aritmilerinin sınıflandırılması, **IEEE 16. Sinyal İşleme ve Uygulamaları Kurultayı**, Antalya, (2008).
- [8] Smith, M.J., ve Barnwell, T.P., A procedure for designing exact reconstruction filter banks for tree structured sub-band coders, **In Proc. IEEE Int.Conf. Acoust., Speech, and Signal Proc.**, San Diego, (1984).
- [9] Turan, M.D., EKG Sinyalindeki gürültülerin IIR filtreler ile matlabda filtrelenmesi, Bitirme Ödevi, Süleyman Demirel Üniversitesi, (2005).
- [10] Vapnik, V., **Statistical learning theory**, Wiley Press, New York, (1998).
- [11] Domingos, P. ve Pazzani, M., On the optimality of the simple Bayesian classifier under zero-one loss, **Machine Learning**, 29, 103–130, (1997).
- [12] Goldberger, A.L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C, Mark, R. G., Mietus, J.E., Moody, G. B., Peng, C. K. ve Stanley, H. E., PhysioBank, PhysioToolkit, PhysioNet, components of a new research resource for complex physiologic signals, **Circulation**, 101, 23, 215–220, (2000).
- [13] Sahambi, J.S., Tandon S.N. ve Bhatt, R. K. P., Using wavelet transforms for ECG characterization, **IEEE Engineering in Medicine and Biology**, 97, 77–83, (1997).
- [14] Mehta, S.S. ve Lingayat, N.S., SVM-based algorithm for recognition of QRS complexes in electrocardiogram, **Elsevier Masson IRBM**, 29, 310–317, (2008).
- [15] Sasikala, P. ve Wahidabanu, R.S.D., Robust QRS Peak and QRS detection in Electrocardiogram using Wavelet Transform, **International Journal of Advanced Computer Science and Applications**, 1, 6, 48–53, (2010).
- [16] Dinh, H.A.N., Kumar, D.K., Pah, N.D. ve Burton, P., Wavelets for QRS Detection, **Proceedings of the 23rd Annual EMBS International Conference**, 1, 1883–1887, İstanbul, (2001).
- [17] Xia, Y., Han, J. ve Wang, K., Quick detection of QRS complexes and R-waves using a wavelet transform and K-means clustering, **Bio-Medical Materials and Engineering**, 26, 1059–1065, (2015).
- [18] Gritzali, F., Towards a Generalized Scheme For QRS Detection in ECG Waveforms, **Signal Processing**, 15, 183–192, (1988).

- [19] Mehta, S.S., Shete, D.A., Lingayat, N.S. ve Chouhan, V.S., K-means algorithm for the detection and delineation of QRS-complexes in electrocardiogram, **Elsevier Masson IRBM**, 31, 48–54, (2010).
- [20] Chouhan, V.S. ve Mehta, S.S., Detection of QRS complex in 12-lead ECG using adaptive quantized threshold, **International Journal of Computer Science and Network Security**, 8, 155–163, (2008).
- [21] Mehta, S.S. ve Lingayat, N.S., Combined entropy based method for detection of QRS complexes in 12-lead electrocardiogram using SVM, **Computers in Biology and Medicine**, 38, 138–145, (2008).
- [22] Gayake, M.A. ve Shete, V.V., ECG QRS-Complex Detection using SVM, **ITSI Transactions on Electrical and Electronics Engineering**, 2, 5-8, (2014).
- [23] Singh, D. ve Khosla, A., QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases, **Journal of Advanced Research**, 4, 331–344, (2013).