

**T.C.  
YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BIYOMEDİKAL İŞARET VE GÖRÜNTÜLERDE GÖRGÜL KİP AYRIŞIMI**

**ÖMER FARUK KARAASLAN**

**YÜKSEK LİSANS TEZİ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI  
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI**

**DANIŞMAN  
YRD. DOÇ. DR. GÖKHAN BİLGİN**

**İSTANBUL, 2015**

**T.C.**  
**YILDIZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİYOMEDİKAL İŞARET VE GÖRÜNTÜLERDE GÖRGÜL KİP AYRIŞIMI**

Ömer Faruk KARAASLAN tarafından hazırlanan tez çalışması 06 / 07 / 2015 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

**Tez Danışmanı**

Yrd. Doç. Dr. Gökhan BİLGİN

Yıldız Teknik Üniversitesi

**Jüri Üyeleri**

Yrd. Doç. Dr. Gökhan BİLGİN

Yıldız Teknik Üniversitesi

\_\_\_\_\_

Doç. Dr. Banu DİRİ

Yıldız Teknik Üniversitesi

\_\_\_\_\_

Doç. Dr. Mustafa KAMAŞAK

İstanbul Üniversitesi

\_\_\_\_\_

## ÖNSÖZ

---

Biyomedikal işaret ve görüntülerin görgül kip ayrışımıyla çıkarılan özelliklerinin çeşitli sınıflandırma ve biçimbilimsel algoritmalarıyla kullanılmasını ve değerlendirilmesini amaçlayan bu tez çalışması Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü Yüksek Lisans programında yürütülmüştür.

Bu tezin hazırlanması sırasında, bilgi birikimiyle çalışmalarına ışık tutan danışmanım Yrd. Doç. Dr. Gökhan BİLGİN'e teşekkür ederim. İş ortamımı yaşanabilir kılan Bilim ve Teknoloji Uygulama ve Araştırma Merkezi'ndeki çalışma arkadaşlarıma ayrıca teşekkür ederim.

Tezimi; gönlümün iki yarısı anne ve babama ithaf ediyorum.

Haziran, 2015

Ömer Faruk KARAASLAN

## İÇİNDEKİLER

---

	Sayfa
SİMGE LİSTESİ.....	vi
KISALTMA LİSTESİ.....	vii
ŞEKİL LİSTESİ.....	viii
ÇİZELGE LİSTESİ .....	x
ÖZET .....	xi
ABSTRACT.....	xiii
BÖLÜM 1	
GİRİŞ.....	1
1.1    Literatür Özeti .....	2
1.2    Tezin Amacı .....	4
1.3    Hipotez .....	4
BÖLÜM 2	
VERİ TANIMLARI.....	5
2.1    Elektrokardiyografi.....	5
2.1.1    EKG Tarihçesi .....	5
2.1.2    EKG Ölçümü .....	6
2.2    Histopatolojik Görüntüler .....	8
2.2.1    Saptama .....	9
2.2.2    Takip .....	9
2.2.3    Bloktan Kesim .....	9
2.2.4    Boyama .....	9
BÖLÜM 3	
ÖZELLİK ÇIKARIMI .....	10

3.1	Görgül Kip Ayrışımı.....	10
3.2	Bütünleşik Görgül Kip Ayrışımı.....	16
3.3	Görüntü Görgül Kip Ayrışımı .....	18
<b>BÖLÜM 4</b>		
<b>İŞARET ve GÖRÜNTÜLERDE GÜRÜLTÜ GİDERME YÖNTEMLERİ.....</b>		<b>20</b>
4.1	Dalgacık Dönüşümü .....	20
4.2	Biçimbilimsel İşlemler .....	23
4.2.1	Aşınma .....	24
4.2.2	Genişleme .....	24
4.2.3	Açma .....	25
4.2.4	Kapama .....	25
4.2.5	Yapısal Elemanlar.....	26
4.3	Medyan Süzgeci .....	26
4.4	Canny Kenar Belirlemesi .....	27
4.5	Gizli Markov Rastgele Alanları – Beklenti Enbüyüklenmesi.....	29
<b>BÖLÜM 5</b>		
<b>SINIFLAYICILAR.....</b>		<b>32</b>
5.1.	Destek Vektör Makineleri .....	32
5.2.	Rassal Ormanlar .....	36
<b>BÖLÜM 6</b>		
<b>DENEY ve UYGULAMA.....</b>		<b>38</b>
6.1.	EKG Denemeleri .....	38
6.2.	Histopatolojik Görüntü Denemeleri .....	45
<b>BÖLÜM 7</b>		
<b>SONUÇ.....</b>		<b>53</b>
<b>KAYNAKLAR.....</b>		<b>54</b>
<b>ÖZGEÇMİŞ.....</b>		<b>58</b>

## SİMGE LİSTESİ

---

$w$	Ağırlık vektörü
$R$	Etiket bilgisi
$b$	Yanlılık (bias) terimi
$H$	Ayırma düzlemi
$L(.)$	Lagrange fonksiyonu
$\alpha_i$	Lagrange çarpanı
$\xi_i$	Yapay değişken (slack variable)
$I(.)$	Belirteç fonksiyonu
$mg(.)$	Marjin Fonksiyonu
$\varphi_{a,b}(t)$	Temel Dalgacık Fonksiyon Kümesi
$y(t)$	Giriş işareti
$a$	Ölçekleme Miktarı
$b$	Öteleme Miktarı
$\mu(t)$	Ölçekleme Fonksiyonu
$\ominus$	Morfolojik aşındırma işlemi
$\oplus$	Morfolojik Genişletme
$H_{i,j}$	Gauss Süzgeci
$G_x$	Yatayda birinci türev
$G_y$	Dikeyde birinci türev
$\theta$	Yönelim Açısı
$T_H$	Yüksek Eşik
$T_L$	Düşük Eşik
$\Theta$	Parametre Kümesi
$U(y)$	Önsel enerji fonksiyonu olan
$Z$	Normalleştirme Sabiti
$P(y)$	Önsel Olasılık
$\hat{Y}$	Gerçek Etiketleme Değeri

## KISALTMA LİSTESİ

---

ADD	Ayrık Dalgacık Dönüşümü
BDT	Bilgisayar Destekli Teşhis
BGKA	Bütünleşik Görgül Kip Ayrışımı
DD	Dalgacık Dönüşümü
DV	Destek Vektör
DVM	Destek Vektör Makineleri
EBS	En Büyük Sonsal
EKG	Elektrokardiyografi
FD	Fourier Dönüşümü
FIR	Finite Impulse Response
GGKA	Görüntü Görgül Kip Ayrışımı
GKA	Görgül Kip Ayrışımı
GMRA-BE	Gizli Markov Rastgele Alanları – Beklenti Enbüyüklemesi
HHD	Hilbert-Huang Dönüşümü
IEMD	Image Empirical Mode Decomposition
İKF	İçkin Kip Fonksiyonu
KAĞ	Kıvrımlı Ağlar
MRI	Magnetic Resonance Imaging
MAP	Maximum a posteriori
RO	Rassal Ormanlar
SDD	Sürekli Dalgacık Dönüşümü
SPECT	Single Photon Emission Computed Tomography
TBA	Temel Bileşen Analizi
UCSB	University of California at Santa Barbara
XICA	Exclusive Independent Component Analysis
YG	Yer Gerçekliği

## ŞEKİL LİSTESİ

	Sayfa
Şekil 2. 1 Bir EKG işaretinin bölümleri .....	7
Şekil 2. 2 Bir histopatolojik resim örneği.....	8
Şekil 3. 1 Bir EKG işaretinde GKA uygulaması .....	12
Şekil 3. 2 $x(t)$ işaretinin gösterimi.....	13
Şekil 3. 3 $x(t)$ işaretinin Fourier Dönüşümü sonucu .....	14
Şekil 3. 4 $y(t)$ işaretinin gösterimi.....	15
Şekil 3. 5 $y(t)$ işaretinin Fourier Dönüşümü sonucu .....	15
Şekil 3. 6 GGKA uygulaması((a) orijinal resim (b-e) İKF'ler (f) artık işareti) .....	19
Şekil 4. 1 İki seviyeli dalgacık paketi ağaç yapısı .....	23
Şekil 4. 2 (a) aşınmada kullanılan örnek görüntü (b) aşınma sonuç görüntüsü.....	24
Şekil 4. 3 (a) genişlemede kullanılan örnek görüntü (b) genişletme sonuç görüntüsü	24
Şekil 4. 4 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) çıktıya uygulanan açma işlemi sonucu elde edilen görüntü .....	25
Şekil 4. 5 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) çıktıya uygulanan kapama işlemi sonucu elde edilen görüntü .....	26
Şekil 4. 6 (a) 2 nolu histopatolojik resim (b) gürültü eklenmiş hali (c) ortalama süzgeç sonucu (d) medyan süzgeç sonucu .....	27
Şekil 4. 7 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) Canny Kenar Belirlemesi sonucu resmi .....	29
Şekil 6. 1 EKG işaretlerine uygulanacak yöntemlerin genel akış şeması.....	38
Şekil 6. 2 St-Petersburg INCART aritmi veri kümesindeki 5 EKG türü ((a) atriyal erken atım (b) karıncık atım füzyonu (c) normal atım (d) sağ dal blok atımı (e) erken karıncık kasılması) .....	40
Şekil 6. 3 European ST-T veri kümesindeki 4 EKG türü ((a) karıncık atım füzyonu (b) normal atım (c) supraventriküler atım (d) erken karıncık kasılması).....	41
Şekil 6. 4 Histopatolojik resimlere uygulanacak yöntemlerin genel akış şeması.....	45
Şekil 6. 5 (a) iyi huylu hücre örneği (b) kötü huylu hücre örneği .....	46
Şekil 6. 6 Önerilen sistemin 2 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama).....	49
Şekil 6. 7 Önerilen sistemin 27 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama).....	50



Şekil 6. 8	Önerilen sistemin 25 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama).....	51
Şekil 6. 9	Önerilen sistemin 47 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama).....	52

## ÇİZELGE LİSTESİ

---

	Sayfa
Çizelge 6. 1	EKG'lerden çıkarılan İKF'leri sınıflandırmada kullanılan özellik kümeleri 42
Çizelge 6. 2	Orijinal sınıflandırma başarısı %86.00 olan St. Petersburg aritmi veri kümesi için elde edilen yüzdelerlik sınıflandırma başarımları..... 43
Çizelge 6. 3	Orijinal sınıflandırma başarısı %87.17 olan European ST-T veri kümesi için elde edilen yüzdelerlik sınıflandırma başarımları..... 44
Çizelge 6. 4	RGB-EMD ve La*b*-EMD'nin Rassal Orman sınıflandırma sonuçları ..... 47
Çizelge 6. 5	Orijinal RGB ve La*b* değerlerine RO uygulanması sonuçları..... 47
Çizelge 6. 6	RO sonuçlarının işleme sonrası sonuçları..... 48

## BİYOMEDİKAL İŞARET VE GÖRÜNTÜLERDE GÖRGÜL KİP AYRIŞIMI

Ömer Faruk KARAASLAN

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Yrd. Doç. Dr. Gökhan BİLGİN

Biyomedikal veriler, kullanıldıkları otomasyonlar sayesinde hastalık teşhisinde doktorlara büyük kolaylıklar sağlamaktadır. Bu verilerin çeşitli işaret, görüntü işleme algoritmaları kullanılarak incelenmesi ve sınıflandırma algoritmalarıyla yapısal özelliklerinin çıkarılması günümüzde yaygın hale gelmiştir.

Biyomedikal alandaki bu yaygın kullanım, beraberinde bazı sorunları da getirmiştir. Bilgisayar ortamında oluşturulan yapay işaretlerin aksine EKG gibi biyomedikal işaretlerin düzensiz ve gürültülü yapısı, klasik işaret işleme algoritmalarının uyarlanması konusunda yetersizliğe sebep olmuştur. Benzeri durum, iki boyutlu işaretler olan histopatolojik görüntüler için de geçerlidir. Bu yapılardaki gürültü ve belirsizlikler, sınıflandırmada kullanılacak özelliklerin çıkarılmasına engel olmaktadır.

Bu tez çalışmasında, belirli bir formül gerektirmeyen, veriyle uyumlu çalışan Görgül Kip Ayrışımı (GKA, Empirical Mode Decomposition) kullanılmıştır. İki bölümden oluşan bu çalışmanın ilk bölümünde EKG işaretleri incelenmiş, klasik işaret ayrıştırma yöntemleriyle performans kıyaslaması yapılmış ve GKA'nın özellik çıkarımında daha başarılı olduğu gözlenmiştir. Bu başarımların değeri, Destek Vektör Makineleri'yle (DVM, Support Vector Machines) sınıflandırma yapılarak elde edilmiştir. İkinci bölümde de histopatolojik görüntüler incelenmiş, GKA'dan elde edilen özellikler çeşitli morfolojik araçlarla eniyileştirilmiş ve orijinal görüntü ögesi değerleriyle elde edilen özellik kümelerine kıyasla Rassal Ormanlar (RO, Random Forests) yöntemi kullanılarak elde edilen sınıflandırma başarısında ilerleme kaydettiği gözlenmiştir.

Bu çalışmadan elde edilen bilgiler, GKA ile biyomedikal işaretlerden elde edilen özelliklerin, uygun gürültü giderme araçlarının kullanılmasıyla, sınıflandırılma

performansını arttırma konusunda klasik yöntemlerden elde edilen özelliklere göre daha belirleyici olduğunu göstermiştir.

**Anahtar Kelimeler:** Biyomedikal işaretler, görgül kip ayrışımı, sınıflandırma, fourier dönüşümü, dalgacık dönüşümü, destek vektör makineleri, rassal ormanlar, morfolojik işlemler, gizli markov rastgele alanları modeli

**EMPIRICAL MODE DECOMPOSITION ON BIOMEDICAL SIGNALS AND  
IMAGES**

Ömer Faruk KARAASLAN

Department of Computer Engineering

MSc. Thesis

Advisor: Asst. Prof. Dr. Gökhan BİLGİN

Biomedical data provides great convenience to doctors thankfully to automations that are used by them. It has been very common to extract and analyze structural features of these data by implementing various signal, image processing and classification algorithms.

This wide usage in biomedical area accompanies some problems, too. In contrast to artificial signals which are created in computer environment, irregular and noisy structure of biomedical signals such as ECG, EEG reasons lack about adaptation of signal processing algorithms. Same case is valid for histopathology images which are 2-dimensional signals. Uncertainties and noises in these structures prevent to extract features that will be used in classification.

In this thesis study, Empirical Mode Decomposition (EMD), which works adaptively with data and doesn't need for a specific formulation, has been used. In the first chapter of this study which is consisted of two chapters, ECG signals have been analyzed, a performance comparison has been made with classical signal decomposition methods and it has been observed that EMD is more successful about feature extraction. This performance value has been acquired by classifying with Support Vector Machines (SVM). In the second chapter, histopathology images has been analyzed, the features extracted from EMD have been optimized various morphological tools and it has been observed that a progress has been made at classification success obtained by using Random Forest (RF) method comparing with the extracted feature sets of the original pixel values.

The acquired information from this study demonstrates that extracted features from biomedical signals with EMD is more identifier according to extracted features from classical methods about increasing classification performance by using suitable denoise tools.

**Keywords:** Biomedical signals, empirical mode decomposition, classification, fourier transform, wavelet transform, support vector machines, random forests, morphological operations, hidden markov random fields model

## BÖLÜM 1

---

### GİRİŞ

Tıp alanında tedaviyi kolaylaştıran uygulamalar, günümüzde oldukça yaygınlaşmaya başladı. Biyomedikal verilerin gerçek zamanlı değerlendirilmesi, hastalık belirtilerinin kolaylıkla algılanabilmesi bu alanda yapılmış çalışmalardan yalnızca birkaçıdır.

Cihaz kalitesinin artması, biyomedikal araçlarla elde edilen verilerin belirleyici kalitesini arttırmış, bu da Bilgisayar Destekli Teşhis (BDT) uygulamalarının performansına katkı sağlamıştır.

Biyomedikal veri kapsamında kullanılan EKG işaretleri ve histopatolojik görüntüler, bu tez çalışmasında incelenmiştir. EKG, kalbin bulunduğu duruma göre vücut üzerinde oluşturduğu elektriksel işaretlerin gösterimi olarak ifade edilir ve kalp hastalıklarının teşhisinde, ön tanısında yaygın olarak kullanılır. Histopatolojik görüntüler de ilgili dokunun hücresel yapılarını barındıran ve gerektiğinde bozuk hücre şekillerinin incelenmesi aracılığıyla özellikle kanser konusunda ön teşhisi kolaylaştıran görüntülerdir.

Yaygın kullanımlarının dezavantajı olarak, bu işaret ve görüntüler edinildikleri yere göre çeşitlilik arz etmektedir. Bunun için işaret ve görüntünün doğrusal olduğu ve durağan olduğu gibi belli varsayımlara dayalı olarak incelenmesi ve ayrıştırılmaya çalışılması, bulunduğu duruma göre sınıflandırılmasını zorlaştırmaktadır. Çünkü gerçek hayatta elde edilen işaretler, bilgisayar ortamında belli bir fonksiyona bağlı olarak üretilen işaretlere nazaran daha düzensiz yapılar barındırır. Bu yapıların içinde yer alan gürültülerin giderilmesi ve ilgili biyomedikal verinin yeniden yapılandırılması, sınıflandırma işlemini kolaylaştıracaktır.

Bu işlemlerin gerçekleştirilebilmesi için, belirli bir fonksiyon çerçevesinde hareket etmeyen, veriyle uyumlu bir metodun kullanılması gerekmektedir. Görgül Kip Ayrışımı(GKA, Empirical Mode Decomposition) yöntemi, veriyle uyumlu bir şekilde yinelemeli olarak ilgili işaret ya da görüntüye ait frekans ve alt parçalarının edinilmesini sağlayan bir yöntem olarak[1], bu tez çalışması kapsamında incelenmiş ve değerlendirilmiştir.

Bu çalışmanın bölümleri şu şekilde oluşturulmuştur: Bölüm 2’de EKG ve histopatolojik görüntülerin yapılarında bahsedilmiştir. Bölüm 3’te GKA ve yaygın kullanılan sürümleri anlatılmıştır. Bölüm 4’te GKA’dan elde edilen özelliklerdeki gürültülerin giderilmesini sağlayan yöntemlerden bahsedilmiştir. Bölüm 5’te çalışma kapsamında kullanılan sınıflandırma algoritmaları açıklanmıştır. Bölüm 6’da Deney ve Uygulama kısmında yapılan işlemler ve denemeler adım adım anlatılmıştır. Bölüm 7’de Sonuç kısmıyla da çalışmadan elde edilen çıkarımlar aktarılmıştır.

## **1.1 Literatür Özeti**

Bu tez çalışmasında bir boyutlu biyomedikal işaret olan EKG verilerinden hastalık tespiti ve iki boyutlu işaret olan histopatolojik resimlerde hücre bölütleme yapılmıştır. Bu işlemlerle ilgili yapılmış çalışmalar literatürü oldukça zenginleştirmiştir.

Bir boyutlu işaretler olan EKG verilerinin tespiti hakkında yapılan çalışmalar, işaretin incelenen kısımlarına göre değişkenlik göstermektedir. İşaretin QRS bölgesinin tespitiyle ilgili bilinen en yaygın çalışma 1985’te Tompkins[2] tarafından ortaya konmuştur. Yapmış olduğu çalışmada Tompkins, öncelikle doğrusal süzgeçlerle gürültü etkilerinin giderilmesini sağlamış, ardından fark alma(differentiator) ve karesini alma(squaring) ile EKG sinyalindeki bölgeler arası farkın belirginleşmesine sebep olmuştur. Son olarak da hareketli pencere yöntemiyle de QRS bölgesinin tespit etmiştir. Uslu ve Bilgin[3] de aritmik EKG sınıflandırması konusunda yerel kesirli Fourier Dönüşümü’nü kullanmıştır. Fetüs durumundaki bebeğin iki kalp atımı arasını gösteren RR aralığındaki değişikliğin tespiti de GKA kullanılarak Ortiz vd. [4] tarafından yapılmıştır. Bu çalışmada Ortiz, yüksek frekansları GKA ile elde etmiş, düşük frekansların da soluk alıp vermek etkisi olarak öngörmüştür. Bu sayede EKG işaretini gürültülerinden arındırmış ve yüksek frekans bilgilerini kullanarak yeniden oluşturmuş. Pan vd. [5], GKA kullanarak EKG’lerde



cihaz ya da kullanıcı tarafından meydana gelen anahat sapmalarını gidermek amacıyla bir çalışma yapmıştır. Dalgacık dönüşümünün, GKA ile kullanıldığı uygulamalar mevcuttur. Yang ve Tang'ın [6] EKG tespiti ve Khiari vd. [7] nin R dalga tespiti konusunda yaptığı çalışmalar bunlardan birkaçıdır. Pinheiro vd. [8], da EKG işaretlerinden GKA ile çıkardıkları özellikleri Temel Bileşen Analizi (TBA) ile incelemiştir. De Ridder vd. [9] GKA'nın, dalgacık ve FIR süzgeçleri ile kıyaslamasını yapmış ve GKA'nın daha iyi olduğunu gözlemişlerdir. Karagiannis ve Constantinou [10] da yaptıkları çalışmada, GKA'nın gürültü giderme konusunda başarılı olduğunu göstermişlerdir.

İki boyutlu işaretler olan görüntülerde bölütleme konusunda da çalışmalar mevcuttur [11], [12]. Anoraganingrum[13] tıbbi resimlerde medyan süzgeciyle ön işleme yapmış, eşikleme ve matematiksel biçimbilim yöntemlerini kullanarak kan hücreleri ile tümör hücrelerinin ayırt edilebilmesini sağlamıştır. Bamford ve Lovell[14] etkin çevrimleri (active contours) ve eğitimsiz sınıflandırma algoritmalarını kullanarak hücre çekirdeği ve hücre bölütlemesi yapmışlardır. Zhang vd.[15] MR görüntülerinin bölütlenmesi konusunda Gizli Markov Rastgele Alan (GMRA, Hidden Markov Random Field) yöntemini kullanmışlardır. Huang ve Lee[16] histopatolojik görüntülerinde mitozlu ve mitozsuz hücrelerin sınıflandırılmasını ayrıcalıklı bağımsız bileşen analizi (XICA; exclusive independent component analysis) yöntemiyle gerçekleştirmiştir. Bu çalışmada, eşikleme ve bölütleme işlemleri yardımıyla hücreler tespit edilmiş ve dokusal tanımlayıcılar yardımıyla öz nitelikleri çıkarılıp sınıflandırılmıştır. Hošťálková ve Procházka[17] biyomedikal görüntülerde gürültü giderimi için Dalgacık Dönüşümü (DD, wavelet transform) kullanmış ve başarılı sonuç almışlardır. Ben Saïd vd.[18] karaciğer tümör hücrelerinin belirlenmesi için hücrelerin uzamsal bilgilerini ve üç boyutlu komşuluk yapılarını (voxels) GMRA-BE yöntemiyle sınıflandırmış, ardından biçimbilimsel işlemleri ve etkin çevrimleri kullanmışlardır. Pang vd.[19] hücre çekirdeği bölütlemesinde Kıvrımlı Ağlar(KAĞ, Convolutional Networks) yöntemini kullanmışlardır. Krinidis vd.[20] renk histogramını GKA ile ayrıştırarak görüntü bölütleme kapsamında yeni bir yöntem geliştirmişlerdir. Sertel vd.[21], doğrusal olmayan renk miktarlama ile histopatolojik resimlerin bölütlenmesini sağlamışlardır. Gallix vd.[22] GKA'yı kullanarak beyin SPECT (Single-Photon Emission Computed Tomography) görüntülerini

bölütlemeyi başarmışlardır. Chen vd.[23] iris görüntülerinden özellik çıkarımı konusunda GKA'yı kullanmışlardır. Mahmood vd.[24] olasılık haritaları kullanarak beyin MRI(Magnetic Resonance Imaging) görüntülerinde bölütleme yapmıştır. Albayrak ve Bilgin[25] kanser hücrelerinin tespiti ve sınıflandırılması konusunda dokusal özellikleri kullanmıştır.

## **1.2 Tezin Amacı**

Bir ve iki boyutlu biyomedikal işaretler, hastanın durumunu belirten niteliklere sahiptir. Bu niteliklerin ilgili tıp mensubuna en iyi şekilde sunulması, hastalık teşhisinde büyük kolaylıklar sağlayacaktır. Doğrusal ve durağan olmayan işaretlerin işlenmesinde sıkıntı oluşturan Fourier gibi klasik işaret ayrıştırıcı yöntemlerin aksine, süreci kolaylaştırmak adına; daha önceden işaretlenmiş verilerin kullanılması, ilgili verinin kendisiyle uyumlu bir şekilde özelliklerinin çıkarılması ve çıkarılan bu özelliklerin uygun sınıflandırma parametreleriyle niteliğinin belirlenmesi gerekmektedir.

Bu tezin amacı bir ve iki boyutlu biyomedikal işaretlerin eğitici sınıflandırma yöntemlerinde kullanılması için veriyi belirgin kılacak niteliklerin edinilmesidir. Elde edilen niteliklerin uygun gürültü giderme algoritmalarıyla en iyileştirilmesi ve sınıflandırma performanslarının bu sayede geliştirilmesi amaçlanmaktadır.

## **1.3 Hipotez**

Biyomedikal işaret ve görüntülerin son zamanlarda detaylı bir şekilde elde edilmesi, biyolojik yapıların incelenmesi ve seçilmesi konusunda büyük kolaylıklar sağlamaktadır. Verilerdeki detayın artması; gürültü miktarını da etkilemekte, FD gibi formül esaslı işaret ayrıştırma yöntemleri tek başına yeterli olmamakta, verinin kendisiyle uyumlu olan GKA yönteminin kullanımını özellik çıkarımı bakımından zorunlu kılmaktadır.

Bu çalışmanın birinci kısmında bir boyutlu biyomedikal işaret olan EKG'lerin özelliklerinin GKA ile çıkarılacak ve gürültülerinin DD ile giderilmesiyle sınıflandırma performansı artacaktır. Çalışmanın ikinci kısmında da histopatolojik görüntülerin öge bazında özellikleri GKA ile çıkarılacak ve çeşitli biçimbilimsel (morfolojik) işlemlerle sınıflandırma başarısı artacaktır.

### VERİ TANIMLARI

Bu bölümde, çalışma esnasında kullanılan Elektrokardiyografi (EKG) işaretleri ve histopatolojik görüntülerin tanımı, yapısı ve genel kullanım alanlarından bahsedilmektedir.

#### 2.1 Elektrokardiyografi

Günümüzde en yaygın ölüm sebepleri arasında olması nedeniyle kalp hastalıklarının ve ritim bozukluklarının bilgisayar destekli teşhis (BDT) sistemleri ile mümkün olduğunca erken tanımlanması ve bunun paralelinde BDT sistemlerinin geliştirilmesi gittikçe daha önemli hale gelmektedir. Özellikle kalp-damar alanında doktor ve uzmanlara teşhis ve tedavi aşamasında büyük kolaylıklar sağlayan en önemli BDT sistemlerinden biri de Elektrokardiyografi(EKG)'dir.

Yunanca elektriksel aktivite anlamına gelen “elektro”, kalp anlamına gelen “kardio” ve yazmak anlamına gelen “graph” kelimelerinin birleşiminden oluşan EKG, vücuda yayılan elektriksel biyomedikal işaretleri kullanarak kalbin durumunu gösteren, kalp ritim bozukluklarının ve hastalıklarının erken tespitinde kullanılan bir görüntüleme biçimidir. Temel çalışma prensibi, vücut yüzeyindeki elektriksel işaretlerin elektrotlar ile alınması ve sayısallaştırılarak gösterimine dayanır.

##### 2.1.1 EKG Tarihçesi

EKG'nin günümüzdeki yapıya benzer ilk kullanımı 1872'de St Bartholomew Hastanesi'nde bileğe bağlanan kablolar ile yapılmıştır. Kalp atımlarının bir cihaza

aktarılarak gerçek zamanlı kaydedilmesini sağlayan ilk makine Augustus Waller tarafından oluşturulmuş ve St. Mary Hastanesi'nde kullanılmıştır.

Bu alanda çığır açan ilk gelişmeyi, 1901'de Willem Einthoven sicimli akımölçeri icat ederek yapmıştır. Akımölçerin uçlarını iki adet içleri tuzlu çözelti barındıran kaplara bağlamış ve kalp bilgisini öğrenecek şahsın bir ayağı ve bir kolunu, bu kapların içine daldırmıştır. Bu sayede suyun içindeki iyon hareketini gözlemlemiş ve bunu kaydetmiştir. 1897'de Clément Ader tarafından icat edilen cihazdan ve Waller'in makinesinden daha hassas bir kılcal damar ölçümü gerçekleştirmiştir. Daha sonra elde ettiği kayıtların genel yapısına bakarak, EKG çıktısının belirli bölümlerine P, Q, R, S ve T harflerini atamış ve bu bölümlere bakarak kalp ritimindeki düzensizlikleri sınıflandırmıştır. 1924'te de yaptığı çalışmalar sebebiyle tıp alanında Nobel ödülünü almıştır.

Günümüzde daha gelişmiş teknolojilere sahip olan EKG cihazlarının temel çalışma ilkeleri Einthoven'ın cihazının yapısıyla benzerlikler taşır. Bu alanda yapılan bilimsel araştırmalar, ölçüm tutarlılık ve sistemlerinin iyileştirilmesi, anlık kalp ritim bozukluğu tespitinin hızlandırılması üzerinedir.

### **2.1.2 EKG Ölçümü**

Klasik 12 başlıklı EKG cihazlarında elektrotlar hastanın belirli kaburga ve göğüs yüzeyine özel yapıdaki bir jel kullanılarak yerleştirilir. Jelin etkisini arttırmak amacıyla elektrotların yerleştirileceği bölgedeki tüy ve kıllar temizlenir. 12 başlıktan elde edilen elektrik potansiyel büyüklüğü periyodik kaydedilir. Bu sayede kalbin elektriksel kutuplanma büyüklüğü ve yönü belirlenmiş olur. Elde edilen gerilim-zaman grafiğine de EKG adı verilir.

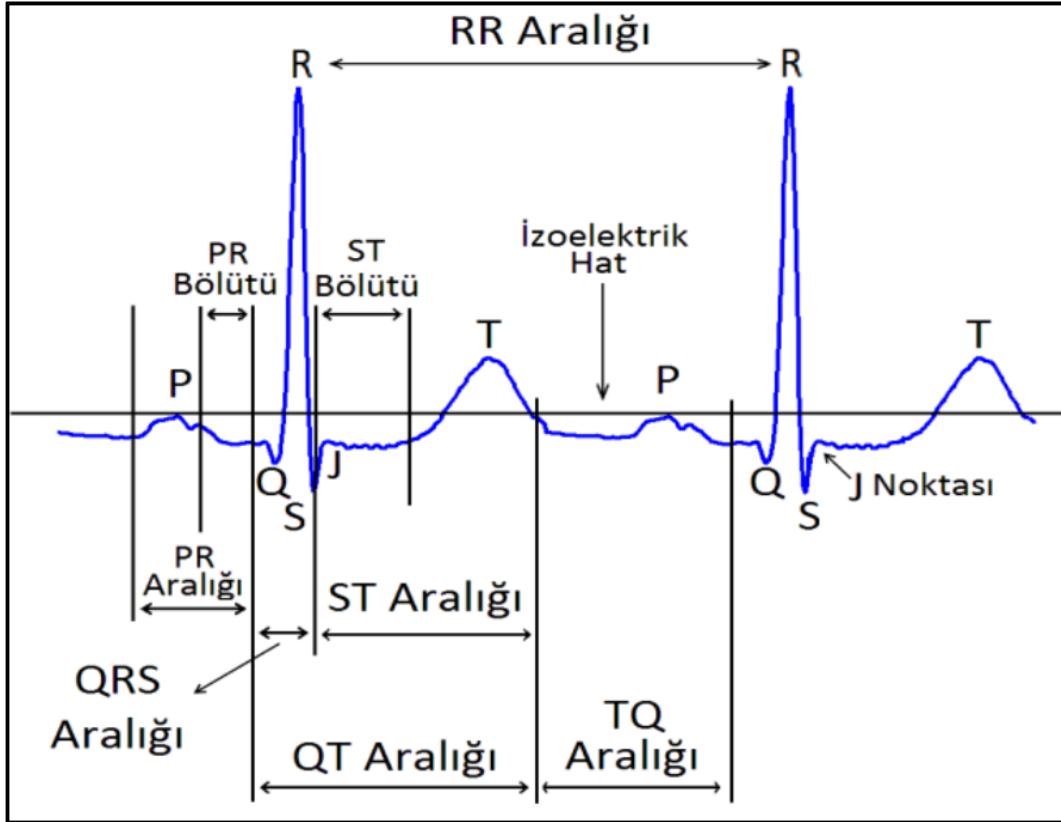
Her kalp atımı esnasında, sağlıklı kalp düzenli bir kutuplanma süreci içindedir. Bu süreç sinoatriyal düğümdeki kalp atışı düzenleyici hücreler tarafından başlatılır, kulakçık boyunca kutuplanma devam eder. Son olarak karıncık boyunca aşağı ve sol yönde Purkinje lifleri aracılığıyla süreç tamamlanır.

Bir EKG çıktısı,

- Kalp atımlarının hız ve ritmi

- Kalp karıncık ve kulakçıklarının boyut ve yeri
- Kalp kaslarındaki ve dolaşım sistemindeki yıpranmaların büyüklüğü
- Kalp ilaçlarının etkileri

başlıkları altında uzman kullanıcılara geniş bilgiler sağlar. Sağlıklı bir kalp vurusuna ait bölütlerin ve aralıkların gösterimi Şekil 2.1'de verilmektedir.



Şekil 2. 1 Bir EKG işaretinin bölümleri

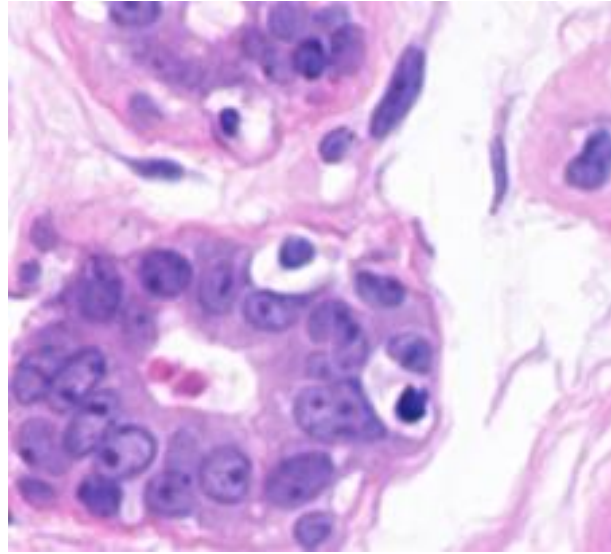
Sağlıklı kalp vurularının belirli bir ritmi vardır. Kalp-damar sistemindeki herhangi bir sorun ya da gecikme, bu ritmi etkiler ve ortaya Aritmik EKG çıkar. Aritmik EKG'lerin yapısı, normal EKG'lerden farklıdır. Hastalık türüne göre çeşitli aritmik EKG türleri mevcuttur(erken, karıncıkla ilgili, iletimsel vs...). EKG ölçümleri sırasında, insan hareketi, şehir şebekesi, nefes alış veriş kaynaklı gürültüler de EKG bilgisinin tespitini zorlaştırmaktadır. Literatür Özeti'nde de belirtildiği üzere, yapılan çalışmaların büyük çoğunluğu bu tarz gürültülerin mümkün olduğunca azaltılıp EKG'den hastalık tespitini kolaylaştıracak özelliklerin edinilmesi üzerinedir.

## 2.2 Histopatolojik Görüntüler

Doku anlamına gelen “Histos”, hastalık anlamına gelen “pathos”, ve bilim anlamına gelen “logos” kelimelerinin birleşiminden oluşan histopatoloji, doku düzeyindeki yapıların yıpranmalarını gözlemleyerek hastalıkların önceden tedavisi konusunda tıp mensuplarına kolaylık sağlayan bir bilim dalıdır.

Patoloji biliminin tarihi, optik mikroskopun tarihiyle benzerlik gösterir. 17. Yüzyılda Robert Hooke'un Micrographia isimli çalışmasıyla insan hücre yapıları görsel olarak sunulmuş ve büyük ilgi toplamıştır. 19. yüzyılda ayrı bir bilim dalı olarak kabul edilen histopatoloji, 20. yüzyılda elektron mikroskoplarının icadıyla büyük ilerlemeler kaydetmiştir.

Histopatolojide temel olarak uygulanan işlem, hasta dokudan alınan bir kesitin yapısal veya işlevsel olarak bozulmaya uğrayıp uğramadığı kontrol edilerek gerekli teşhisin konmasıdır. Bu teşhisin konabilmesi için hücre ile hücre dışı yapıların birbirinden ayrılması ve böylelikle kullanıcı uzmana kolaylık sağlanması gerekmektedir.



Şekil 2. 2 Bir histopatolojik resim örneği

Hücrelerin ve yapıların tespitini kolaylaştırmak amacıyla ilgili dokuya uygulanan işlemler sırasıyla şu şekildedir:

### **2.2.1 Saptama**

Vücuttan alınan dokular canlılıklarını bir süre daha devam ettirir. Yapıların bu canlılık süresinde saptanabilmesi ve çevresel etmenlerden etkilenmemesi için ilgili doku özel bir sıvıda bekletilir. Sıvı olarak genelde, doku yapısını uzun süre muhafaza edebilen %10'luk formalin çözeltisi kullanılır.

### **2.2.2 Takip**

Saptamadan belirli bir süre sonra, dokunun belirgin ve sert olması amacıyla formalin ya da kullanılan sıvı uzaklaştırılır. Dokunun sertliğini azalttığından, alkol kullanılmışsa ksilol yardımıyla ortamdan uzaklaştırılması gerekir. Daha sonra da, alkol ve ksilolle steril hale getirilen sıvı parafin dokuya aktarılır. Oda sıcaklığında katılaştıran parafin sayesinde ilgili dokular daha belirgin ve işlenecek hale gelir. Ardından numaralandırma ve takip işlemi gerçekleştirilir. Vakumlama, mikrodalga gibi yöntemlerle dokuların işlenmesi oldukça kolaylaşır.

### **2.2.3 Bloktan Kesim**

İşlenen dokular, dikdörtgen prizma biçimindeki kalıplara konular ve ince film halinde kesilebilmesi amacıyla üzerlerine ısıtılmış parafinin dökülüp soğutulur.

Parafin bloklar; Mikrotom aygıtıyla belirli kalınlıkta kesilir, daha sonra kesitler ılık su banyosuna, oradan da lamlara aktarılır. Kesitlerin ksilol ile parafinden arındırılmasının ardından boyama işlemi başlatılır.

### **2.2.4 Boyama**

Histopatolojik çalışmaların son aşaması olan boyamada Hematoksilen-Eozin boyama yöntemi (H&E) kullanılır. Yapıları belirleyebilmek amacıyla; hücre çekirdekleri Hematoksilen (mavi) ile, hücre sıvısı da Eosin (kırmızı) ile boyanır. Yapılan son çalışmalarda, belirli yapıların tespiti için antikorlar da kullanılmaya başlanmıştır.

### ÖZELLİK ÇIKARIMI

Bu bölümde, biyomedikal işaret ve görüntülerin sınıflandırılmasında kullanılacak özelliklerin çıkarımında kullanılan yöntemlerden bahsedilecektir.

#### 3.1 Görgül Kip Ayrışımı

Hilbert–Huang Dönüşümü (HHD, Hilbert-Huang Transform)’nün birinci aşaması olan Görgül Kip Ayrışımı 1998’de Huang vd. [1] tarafından doğrusal olmayan ve durağan olmayan zaman serilerinin incelenmesi için ortaya atılmıştır. GKA, en karmaşık yapıdaki verileri bile İçkin Kip Fonksiyonu (İKF, Intrinsic Mode Function) adı verilen sınırlı sayıda parçaya ayrıştırabilmesi sebebiyle HHD’nin en önemli kısmı olarak görülür.

İKF’ler tipik olarak işaretin anlık frekanslarını zaman fonksiyonu olarak sunarlar. İlgili  $x(t)$  işareti yinelemeli bir şekilde adım adım ayrıştırılır. Ardından işaretin özelliğine göre belirli sayıda İKF ( $İKF_i, i = 1, 2, \dots, n$ ) ve artık işareti ( $art$ ) elde edilir:

$$x(t) = \sum_{i=1}^n İKF_i + art \quad (3.1)$$

GKA işlemi şu şekilde uygulanır:

**Adım 1.** Öncelikli olarak işaretin tüm yerel minimum ve maksimum noktaları çıkarılır. İşarete ait maksimum noktaları ile üst zarf ( $env_{max}$ ); minimum noktaları ile de alt zarf ( $env_{min}$ ) elde edilir.

**Adım 2.** Elde edilen zarfların ortalaması, giriş işaretinden çıkarılır:



$$x(t) = x(t) - \frac{(env_{min} + env_{max})}{2} \quad (3.2)$$

**Adım 3.** Ortalama zarf işareti uç noktalar barındırdıkça bu işleme birinci adımdan itibaren devam edilir. Uç noktaların tamamlanması, ilgili adımda İKF'nin bulunduğunu gösterir.

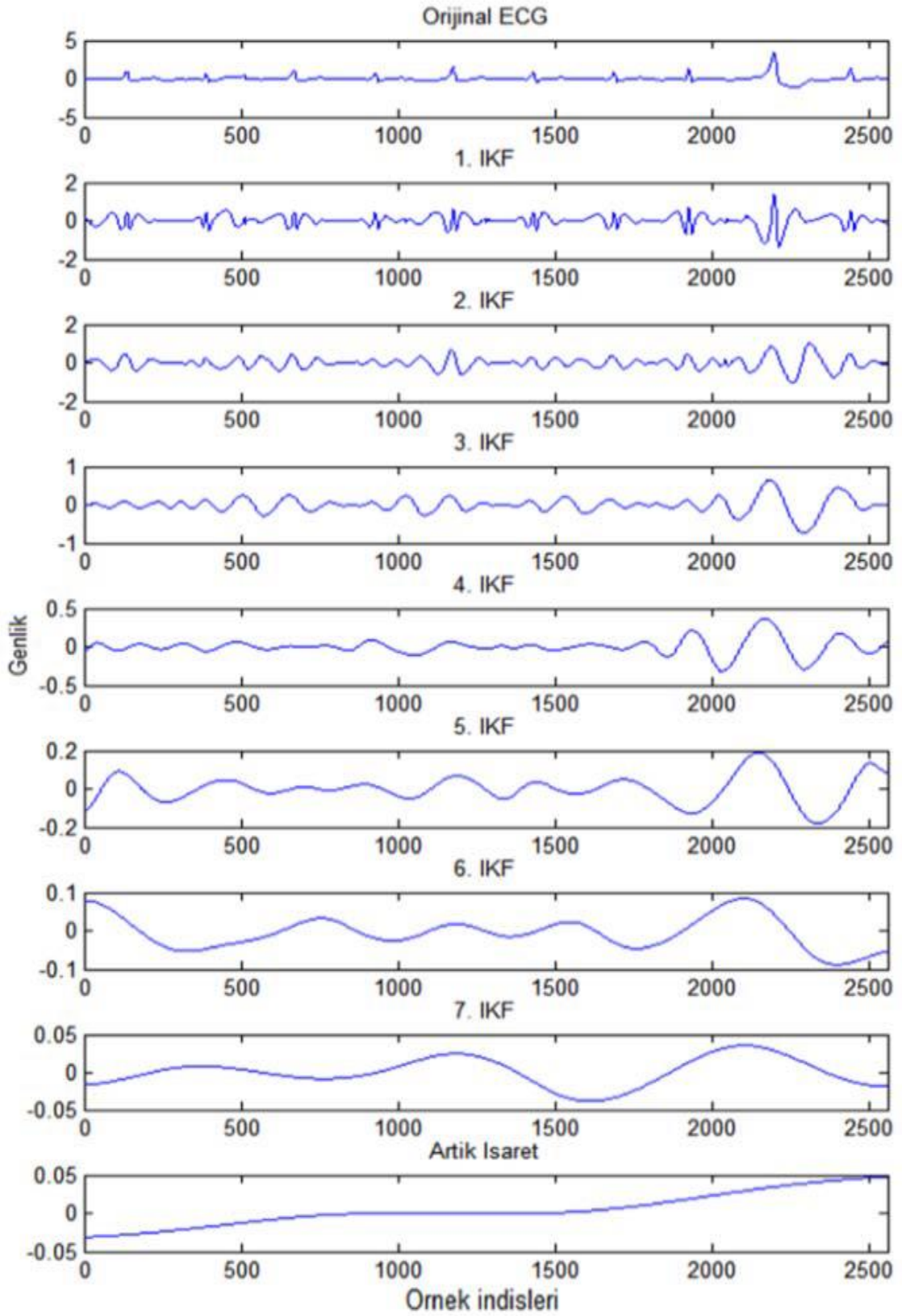
**Adım 4.** Üçüncü adımda elde edilen İKF, giriş işaretinden çıkarılır ve kalan işaret giriş işareti gibi değerlendirilerek işlem birinci adımdan itibaren yinelemeli olarak devam ettirilir. Giriş işareti uç nokta barındırmadığı zaman, artık işareti (*art*) elde edilir ve işlem tamamlanır.

Örnek bir EKG işareti üzerine uygulanan GKA'nın sonuçları Şekil 3.1'de verilmiştir:

Elde edilen İKF'ler iki koşulu sağlamalıdır:

- 1) Sıfırdan geçiş ve uç noktaların sayısı ya eşit olmalı ya da en fazla 1 farklı olmalıdır.
- 2) Yerel maksimum noktalarıyla belirlenen üst zarf ile yerel minimum noktalarıyla belirlenen alt zarfın ortalaması sıfır olmalıdır.

GKA, veri ile uyumlu olduğu ve ilgili işaret frekansının zaman bilgisini kapsayacak şekilde yerel özellikleri incelediği için, doğrusal olmayan ve durağan olmayan işaretler ile oldukça verimli çalışmaktadır.



Şekil 3. 1 Bir EKG işaretinde GKA uygulaması

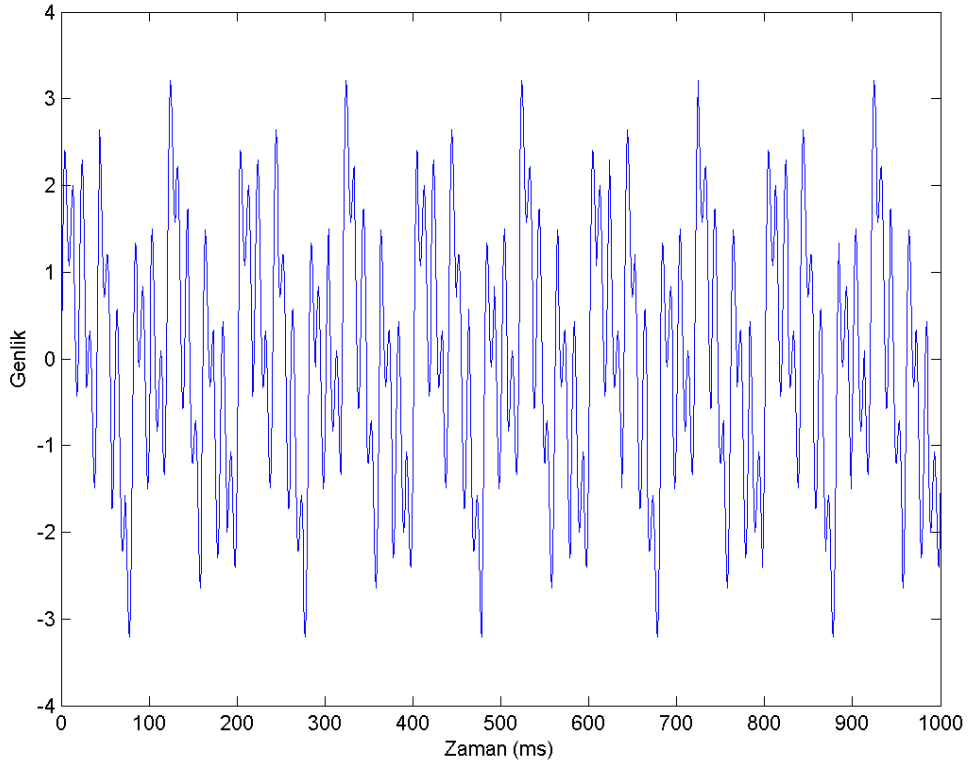
Yaygın olarak kullanılan Fourier Dönüşümü ve Dalgacık Dönüşümü'nün aksine GKA, giriş verisi için işaretin durağan olması gibi herhangi bir varsayıma gerek duymaz. Fourier Dönüşümü, işaretin durağan ve doğrusal olduğunu varsayar, ancak bu varsayım işareti ayrıştırıcı bir özelliğe sahip değildir.

Aşağıdaki formüle sahip bir frekansı ele alalım:

$$x(t) = \cos(2 * \pi * 10 * t) + \cos(2 * \pi * 25 * t) + \cos(2 * \pi * 50 * t) + \cos(2 * \pi * 100 * t) \quad (3.3)$$

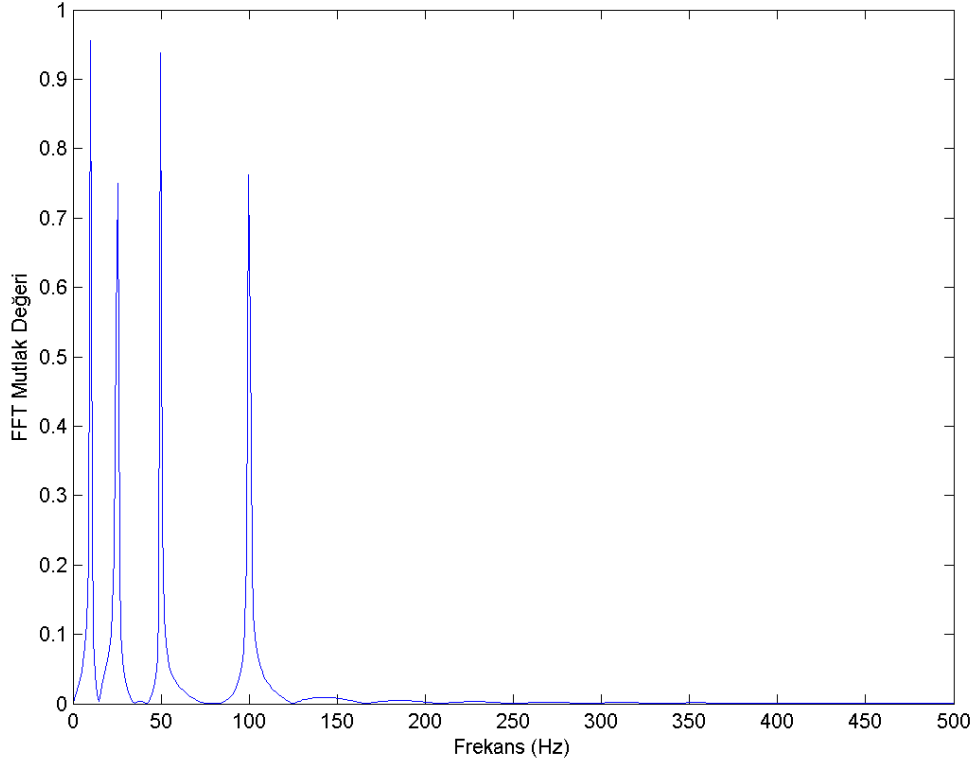
Bu, durağan bir işarettir; çünkü her  $t$  anında 10, 25, 50 ve 100 Hz'lik işaretler barındırır.

Gösterimi de şu şekildedir:



Şekil 3. 2  $x(t)$  işaretinin gösterimi

Bu işaretin Fourier Dönüşümü sonucu da şu şekildedir:



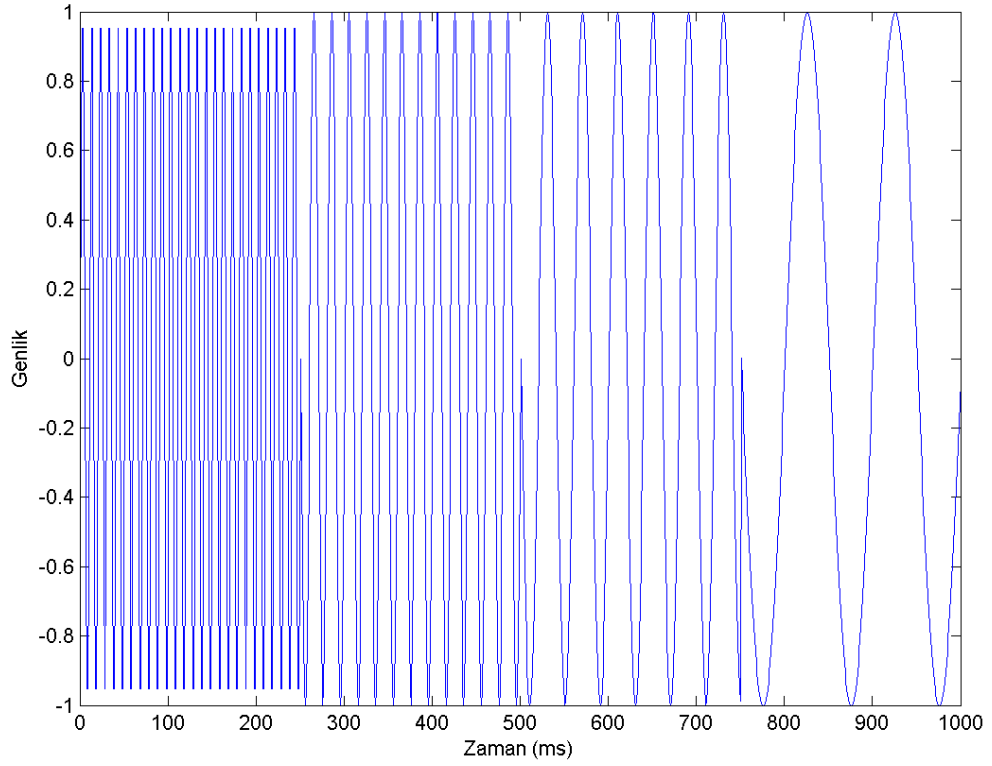
Şekil 3. 3  $x(t)$  işaretinin Fourier Dönüşümü sonucu

Görüldüğü üzere Fourier Dönüşümü ile, bu işarettaki 10,25, 50 ve 100 Hz'lik işaretler elde edilmiştir.  $x(t)$ , durağan bir işaret olduğu için bu frekanslar her  $t$  anında gözlemlenebilir.

Aynı frekansların farklı zaman aralıklarında yer aldığı başka bir  $y(t)$  işareti oluşturulduğunda Fourier Dönüşümü, benzeri bir sonuç verecektir. İşaretimizin yapısı aşağıdaki gibi belirlenmiş olsun:

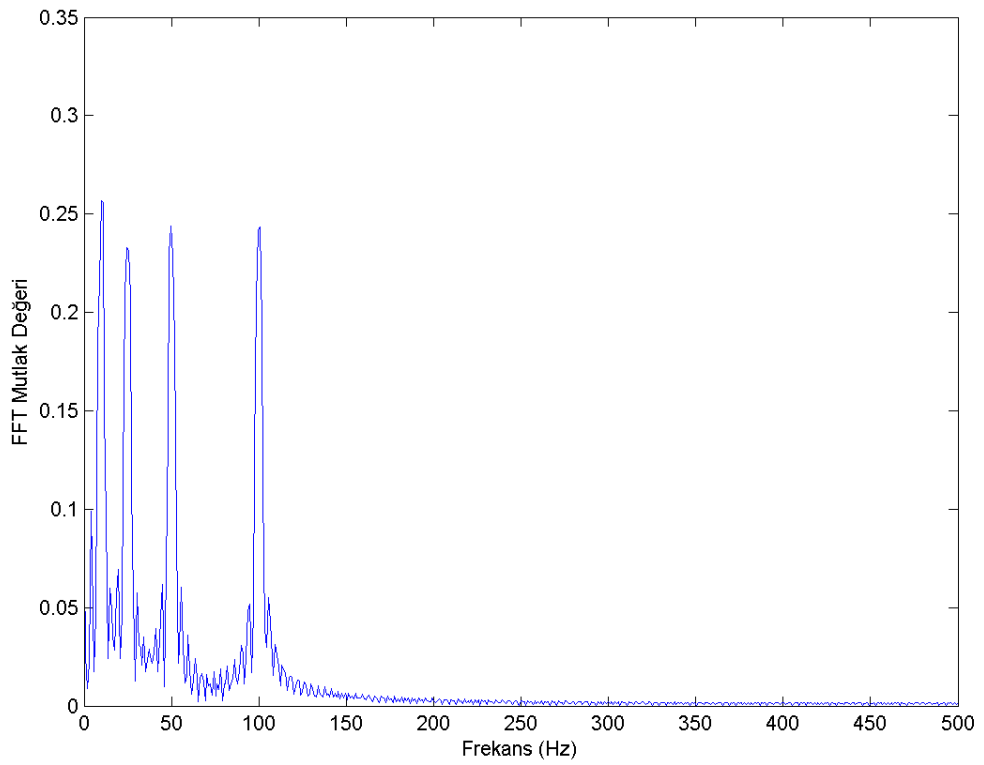
- 0 - 250 ms = 100 Hz
- 250 - 500 ms = 50 Hz
- 500 - 750 ms = 25 Hz
- 750 - 1000 ms = 10 Hz

İşaretin gösterimi şu şekilde olacaktır:



Şekil 3. 4  $y(t)$  işaretinin gösterimi

Bu işaretin Fourier Dönüşümü sonucu da şu şekildedir:



Şekil 3. 5  $y(t)$  işaretinin Fourier Dönüşümü sonucu

Şekil 3.3 ile Şekil 3.5 karşılaştırıldığında, yüksekten alçağa geçişteki değişiklikler dışında, belirgin bir benzerlik görülecektir. Çünkü Fourier Dönüşümü, işaretin zaman bilgisini değil, frekans miktarını hesaplar. Dolayısıyla her iki işarete de 10, 25, 50 ve 100 Hz'lik sinusoid işaretler kullanıldığı için Fourier Dönüşümü sonucuna göre ikisi de aynıdır. Ancak şekil 3 ve şekil 4 kıyaslanacak olursa görülecektir ki, bu iki işaret birbirinden tamamen farklıdır. Fourier Dönüşümü, bu açıdan durağan olmayan yapılar barındıran EKG gibi biyomedikal işaretlerin işlenmesi ve incelenmesinde yetersiz kalmaktadır.

Dalgacık Dönüşümü temelli ayrıştırma yöntemlerine bakıldığında, GKA ile arasında iki temel ayrım görülmektedir. Bunlardan birincisi, farklı dalgacık tiplerinin kullanımı mümkündür ve bu da dönüşüm performansını dalgacık tipine bağımlı kılar. GKA'nın, Dalgacık Dönüşümü'ndeki dalgacık tipine benzeyen temel bir fonksiyonu olmadığı için, ilgili işareti içkin (intrinsic) karakterine göre ayırır. İkincisi, frekans ölçeği Dalgacık Dönüşümü'nde örnekleme frekansı ve ayrıştırma düzeyine göre her zaman sabittir. Buna karşın GKA'da İKF'lerin sayısı ve frekans içeriği işaretin yerel özelliklerine bağlı olarak değişkenlik gösterebilir.

### 3.2 Bütünleşik Görgül Kip Ayrışımı

Bilgisayarda yapay olarak oluşturulan işaretlerin aksine, gerçek hayattaki işaretlerin frekansları doğrusal olmayan bir yapıya sahiptir. GKA yönteminin en büyük dezavantajlarından biri, "kip karıştırma" etkisi olarak bilinen kavramdır. Kip karıştırma, farklı zaman ölçeklerindeki salınımların bir İKF'de yer alabilmesi ya da aynı zaman ölçeğindeki salınımların farklı İKF'lerde yer alabilmesi durumudur.

Bu sorunu ortadan kaldırmak adına, Bütünleşik Görgül Kip Ayrışımı (BGKA, Ensemble Empirical Mode Decomposition) ortaya konulmuştur [26]. BGKA'nın temel ilkesi, birkaç deneme ile işarete beyaz gürültü eklemektir.

**Tanım 1.1** Beyaz gürültü, ortalaması sıfır, sınırlı çeşitlilikte birbiriyle bağlantısı olmayan örnekler barındıran işarettir. Örnekleri normal dağılım özelliğine sahipse, buna Gauss beyaz gürültüsü adı verilir[27].

GKA üzerinde beyaz gürültüyle yapılan çalışmalar, sınırlı genlikteki beyaz gürültünün istatistiksel özelliklerde belirleyici olduğunu ortaya koymuştur[28], [29].

BGKA işlem adımları şu şekilde gerçekleştirilir:

**Adım 1.** Beyaz gürültü adeti ( $M$ ) ve beyaz gürültü genliği ( $A$ ) belirlenir. Adım sayısı olan  $m$ 'nin ilk değeri 1 olacak şekilde ayarlanır.

**Adım 2.** İncelenecek  $x(t)$  işaretine beyaz gürültü serilerinden biri ( $n_m(t)$ ) eklenir.

$$x_m(t) = x(t) + n_m(t) \quad (3.4)$$

**Adım 3.** Bölüm 3.1'de anlatıldığı şekilde  $x_m(t)$  işareti İKF'lerine ve artık işaretine ayrıştırılır. İKF $_{mi}$ ,  $m$ . yinelemede elde edilen  $l$  adet İKF'den  $i$ . İKF'yi gösterir.

$$x_m(t) = \sum_{i=1}^n \text{İKF}_{mi} + \text{art}_m \quad (3.5)$$

**Adım 4.**  $m < M$  olduğu sürece işlem Adım 2'den itibaren farklı beyaz gürültü serileriyle devam ettirilir.

Adım 5. Her İKF için ayrı ayrı ortalama ( $c_i$ ) hesaplanır ve elde edilen ortalama incelenen işaretin  $i$  numaralı İKF'si olarak kabul edilir.

$$c_i = \frac{1}{M} * \sum_{m=1}^M \text{İKF}_{mi}, i = 1, 2, \dots, l \quad (3.6)$$

BGKA'nın etkili olmasını sağlayan 2 parametre mevcuttur: Yineleme sayısı ( $N$ ) ve beyaz gürültü genliği ( $a$ ). Bu iki parametre arasında

$$e = \frac{a}{\sqrt{M}} \quad (3.7)$$

şeklinde bir eşitlik söz konusudur. Burada  $e$ , standart sapma hatasını göstermektedir. BGKA gürültü kaynaklı bir işaret işleme türü olduğundan, beyaz gürültüdeki genlik değeri ile yineleme sayısının orantılı olması, gürültünün işaretin ayrıştırılmış parçalarındaki özellikleri etkilememesi bakımından gereklidir.

### 3.3 Görüntü Görgül Kip Ayrışımı

Görüntü Görgül Kip Ayrışımı (GGKA, Image Empirical Mode Decomposition), GKA'nın iki boyutlu işaretler olan görüntüler için kullanılan sürümüdür[30]. Adımları bir boyutlu işaretlere uygulanan GKA ile benzer olup, şu şekildedir:

**Adım 1.** Öncelikli olarak giriş resminin  $(x(w, h), w: \text{genişlik}, h: \text{yükseklik})$  tüm yerel minimum ve maksimum noktaları 2-d spline enterpolasyonu yöntemiyle çıkarılır. Resme ait maksimum noktaları ile üst zarf ( $env_{max}$ ); minimum noktaları ile de alt zarf ( $env_{min}$ ) elde edilir.

**Adım 2.** Elde edilen zarfların ortalaması, giriş resminden çıkarılır:

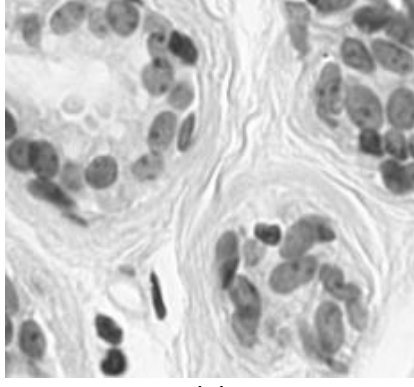
$$x(w, h) = x(w, h) - \frac{(env_{min} + env_{max})}{2} \quad (3.8)$$

**Adım 3.** Ortalama zarf işareti uç noktalar barındırdıkça bu işleme birinci adımdan itibaren devam edilir. Uç noktaların tamamlanması ve  $x(w, h)$  değerinin belirli bir eşik değerinin altında olması, ilgili adımda İKF'nin bulunduğunu gösterir.

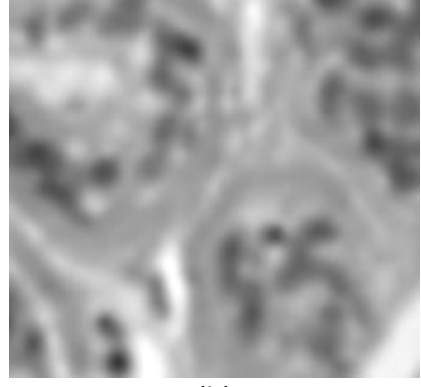
**Adım 4.** Üçüncü adımda elde edilen İKF, giriş resminden çıkarılır ve kalan resim, giriş resmi gibi değerlendirilerek işlem birinci adımdan itibaren yinelemeli olarak devam ettirilir. Giriş resmi uç nokta barındırmadığı zaman, artık işareti ( $art$ ) elde edilir ve işlem tamamlanır.

Bir histopatolojik resim üzerinde GGKA uygulaması Şekil 3.6 da gösterilmektedir:





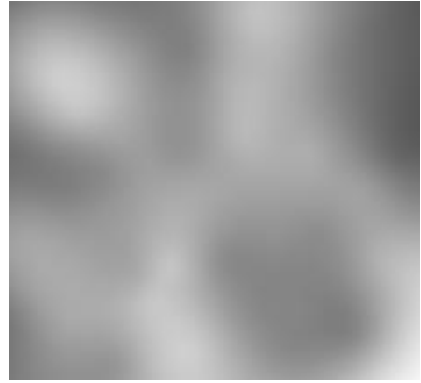
(a)



(b)



(c)



(d)



(e)



(f)

Şekil 3. 6 GGKA uygulaması((a) orijinal resim (b-e) İKF'ler (f) artık işareti)

### İŞARET ve GÖRÜNTÜLERDE GÜRÜLTÜ GİDERME YÖNTEMLERİ

Bu bölümde Görgül Kip Ayrışımı ile elde edilen özelliklerdeki gürültülerin giderilmesi konusunda EKG işaretlerine ve histopatolojik resimlere uygulanan yöntemlerden bahsedilmektedir.

#### 4.1 Dalgacık Dönüşümü

Klasik Fourier Dönüşümü'nde frekans bilgisi başarılı bir şekilde edinilmesine rağmen, işaretin karakterini belirleyecek zaman bilgisinin edinilememesi özellik çıkarımı konusunda sıkıntı oluşturmaktadır. Kısa Zamanlı Fourier Dönüşümü'nde de işaretin belirli kısımlarının doğrusal ve durağan olduğunun varsayılması, biyomedikal işaretlerin işlenmesinde sorun oluşturmaktadır. Kısa Zamanlı Fourier Dönüşümü'nde işaretin hangi "zaman aralığında" olduğu bilgisi edinilmekte, zaman bilgisine tam olarak erişilememektedir.

Dalgacık Dönüşümü, Fourier Dönüşümü'nün eksik kaldığı ilgili frekansın zaman karakteristiğini edinmeyi sağlayan bir işaret işleme yöntemidir. Böylelikle bilgisayar ortamında oluşturulan durağan ve doğrusal yapay işaretlerin aksine, biyomedikal alanda sıkça rastlanan doğrusal olmayan ve durağan olmayan EKG gibi işaretlerin karakteristik özellikleri kolaylıkla ortaya çıkacaktır[31].

Dalgacık dönüşümü,  $f(t)$  işaretin temel dalgacık fonksiyon kümesi ile integrali olarak tanımlanabilir. Temel dalgacık fonksiyon kümesi  $\varphi_{a,b}(t)$ , temel dalgacığın ölçeklenmesi ve ötelenmesiyle elde edilir.

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}} \varphi\left(\frac{t-b}{a}\right) \quad a > 0; a, b \in R \quad (4.1)$$

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

$$W_t(a, b) = \int_{-\infty}^{+\infty} f(t) * \varphi_{a,b}(t) dt \quad (4.2)$$

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[32]. Böylece işaret belirli ölçeklerde incelenerek hesapsal verimlilik artırılmış olur. Bu yöntemle çoklu çözünürlük analizi adı verilir. Bu durumda ayrık dalgacık kümesi

$$\varphi_{j,k}(x) = 2^{-j/2} * \varphi(2^{-j} * x - k) \quad j, k \in Z \quad (4.3)$$

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

$$c_{j,k} = 2^{j/2} \int_{-\infty}^{+\infty} f(x) * \varphi(2^j x - k) dx \quad (4.4)$$

ile bulunur. Elde edilen dalgacık katsayıları yardımıyla geri dönüşüm formülü ise

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} c_{j,k} * \varphi_{j,k}(x) \quad (4.5)$$

şeklinde elde edilir.

Dalgacık dönüşümünde işaret, ayrıştırılırken ölçekleme ve dalgacık fonksiyonları cinsinden ifade edilebilir. Ölçekleme fonksiyonu  $\mu(t)$  ve dalgacık fonksiyonu  $\varphi(t)$

aşağıdaki şekilde tanımlanabilir. Burada  $g(k) = (-1)^k$  ve  $h(1-k)$  olmak üzere süzgeç katsayılarıdır.

$$\mu(x) = \sqrt{2} * \sum_k h(k) * \mu(2x - k) \quad (4.6)$$

$$\varphi(x) = \sqrt{2} * \sum_k g(k) * \mu(2x - k) \quad (4.7)$$

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

$$f(x) = \sum_k c_{j_0,k} \mu_{j_0,k}(k) + \sum_{j=j_0}^J \sum_k d_{j,k} \varphi_{j,k}(x) \quad (4.8)$$

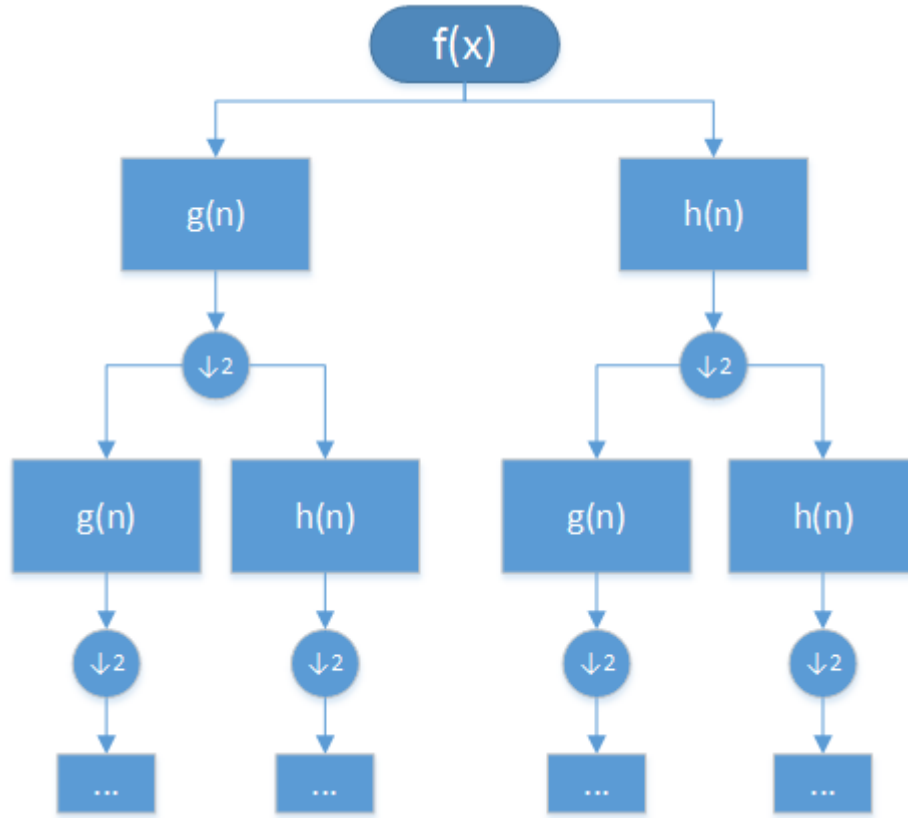
Burada  $J \geq j \geq j_0$  olmak üzere  $j_0$  keyfi bir başlangıç ölçeğidir.  $j$  seviyesindeki  $c_{j,k}$  ve  $d_{j,k}$  katsayılarını elde etmek için,  $j-1$  seviyesindeki  $c_{j-1,k}$  yaklaşıklık katsayılarıyla alçak ve yüksek geçiren süzgeçler evriştirilmiştir (convoluted). Örnek sayısı ise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

$$c_{j,n} = \sum_k c_{j-1,k} * h(k - 2n) \quad (4.9)$$

$$d_{j,n} = \sum_k c_{j-1,k} * g(k - 2n) \quad (4.10)$$

EKG işaretlerinde ayrıştırma işleminde her bir seviyede alçak geçiren bantların kullanılması ile doğrusal dalgacık özellik çıkarımı elde edilir. Doğrusal dalgacık dönüşümünde elde edilen düşük boyutlu yaklaşıklık katsayıları işaretin sonlu elemanlı yaklaşıklılığını verir. Dalgacık paketlerinde ise sadece alçak geçiren bantlar değil aynı zamanda yüksek geçiren bantlar da hesaba katılır. Böylece elde edilen tüm alt bantlarda yinelemeli olarak alçak ve yüksek geçiren süzgeçler yardımıyla paket katsayıları elde edilir. Yaklaşıklık katsayıları haricinde ayrıntı katsayılarının da hesaba katılması ile işaretin daha ayrıntılı bir biçimde analizine olanak sağlanır.

2-seviyeli bir dalgacık paket ağaç yapısı Şekil 4.1’de gösterilmiştir. Oluşturulan dalgacık paketi ağaç yapısının uç düğümlerinde işaret alt uzayının değişik zaman ve frekans yerleşim karakteristikleri bulunur. Bu karakteristikler de kendi içerisinde birleştirilip sınıflayıcıya düşük boyutlu özellikler sağlanabilir.



Şekil 4. 1 İki seviyeli dalgacık paketi ağaç yapısı

#### 4.2 Biçimbilimsel İşlemler

Biçimbilimsel (morfolojik) işlemler, küme teorisi temel alınarak ortaya konmuş matematiksel işlemler türüdür[33]. Görüntü bölütleme, görüntüdeki örüntünün sınırları ve iskelet gibi yapıların tanımlanması, gürültü giderimi, görüntü iyileştirme, kenar tespiti, gibi uygulamalarda sıkça kullanılmaktadır.

Ön bilginin elde edilmesi amacıyla kümelenen histopatolojik görüntüde hücrel yapıların dışında gürültü olarak sayılabilecek belirli hücrel olmayan yapılar da oluşmaktadır. Bundan sonra yapılacak işlem, normal hücrel büyüklüğe sahip olmayan bu yapıların temizlenmesi olacaktır. Bunun için belirleyici bilgiler 1, arkaplan bilgileri de 0 değerine sahip olur.

#### 4.2.1 Aşınma

Aşınma işlemi ikili bir görüntü üzerinde resmin içindeki nesnelerin küçülmesini, daralmasını sağlayan bir işlemdir. Bu işlem sonrasında resim üzerindeki bazı gürültüler yok olur, açıklıklar büyür, zayıf bağlantılar kopar, şekiller daralır. Aşınma işlemi genel olarak birbirine ince bir çizgiyle bağlanmış birden fazla nesneyi birbirinden ayırmak için kullanılmaktadır.  $\ominus$  simgesiyle ifade edilir.



(a)



(b)

Şekil 4. 2 (a) aşınmada kullanılan örnek görüntü (b) aşınma sonuç görüntüsü

#### 4.2.2 Genişleme

Genişleme işlemi, ikili görüntüdeki nesnelere büyüyen ve kalınlaştıran morfolojik işlemdir. İkili görüntüdeki bir nesnenin birbirinden ayrı iki nesne gibi görünmesine engel olmak için kullanılmaktadır.  $\oplus$  simgesiyle ifade edilir.



(a)



(b)

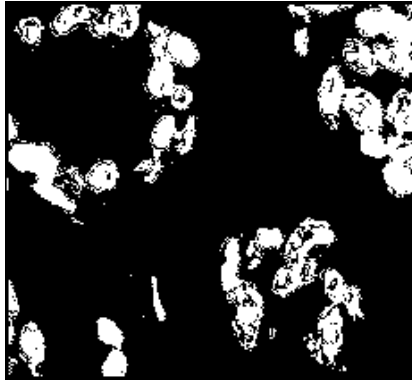
Şekil 4. 3 (a) genişlemede kullanılan örnek görüntü (b) genişletme sonuç görüntüsü

### 4.2.3 Açma

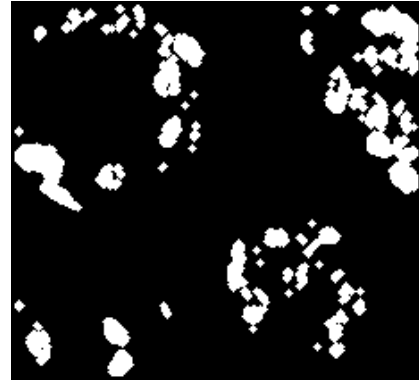
Görüntü işlemede aşınma işleminin hemen ardından gerçekleştirilen genişleme işlemi sonucunda açma işlemi elde edilir.

$$A \circ B = (A \ominus B) \oplus B \quad (4.11)$$

Açma işleminde temel olarak nesnelerin fazla bir değişikliğe uğramadan birbirinden ayrılması amaçlanmaktadır.



(a)



(b)

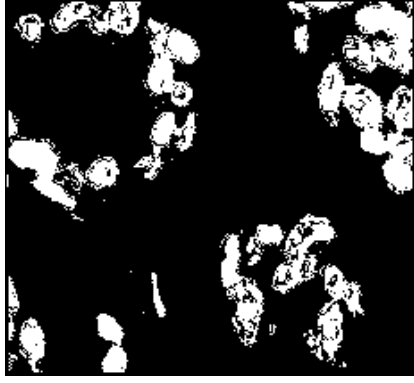
Şekil 4. 4 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) çıktıya uygulanan açma işlemi sonucu elde edilen görüntü

### 4.2.4 Kapama

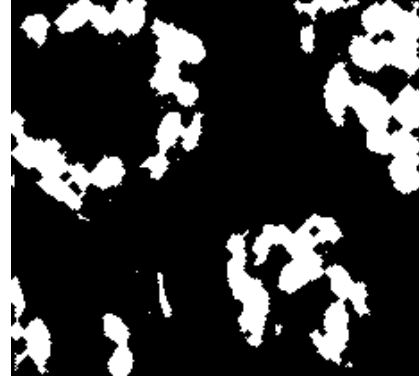
Görüntü üzerinde genişleme işleminden hemen sonra aşınma işlemi uygulanarak kapama işlemi elde edilir.

$$A \bullet B = (A \oplus B) \ominus B \quad (4.12)$$

Kapama işleminde amaçlanan şey birbirine yakın iki nesnenin görüntüde fazla değişikliğe gidilmeden bağlanmış olmasıdır.



(a)



(b)

Şekil 4. 5 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) çıktıya uygulanan kapama işlemi sonucu elde edilen görüntü

#### 4.2.5 Yapısal Elemanlar

Aşınma, genişletme, açma, kapama gibi morfolojik işlemlerde ilgili morfolojik işlemin etkisini belirleyecek unsurlardan biri, görüntü ögesinin çevresine büyütme ya da küçültme sağlayacak olan yapısal elemanlardır. Yapısal elemanlar, ikili formatta, çeşitli geometrik şekle sahip olan görüntülerden oluşmaktadır. İşlenmek istenen görüntünün ilgili bölgesi ile yapısal elemanın kesişmesinden oluşan nesneyi yapısal eleman belirlemektedir. Bu nesne yapısal elemanın şekline göre değişiklik göstermektedir. Örneğin yuvarlak hücre yapılarını belirgin kılabilme amacıyla disk ya da daire şeklindeki yapısal elemanlar kullanılır.

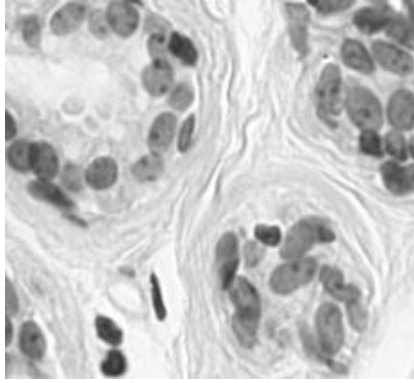
#### 4.3 Medyan Süzgeci

Dijital verilerde yer alan, ilgili işaret elemanının çevresiyle uyumlu bir değer taşımadığı duruma gürültü denir. Medyan süzgeci, bir ve iki boyutlu işaret işlemede kullanılan ön işleme ve gürültü giderme tekniklerinden biridir.

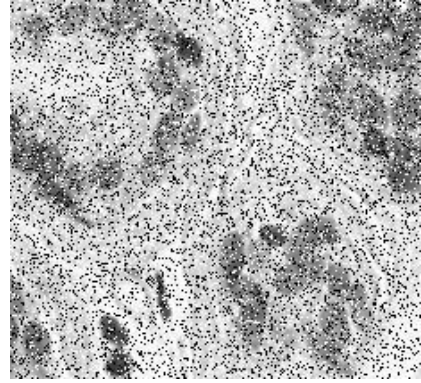
Bu süzgeç yönteminde, ilgili işaret parçasını ya da görüntü ögesinin komşu değerleri tek sayılı bir boyuta sahip pencere ile sıralama işlemine tabi tutulur. Sıralamada ortadaki eleman medyan değeri olarak belirlenir ve bu değer ilgili işaret parçasının ya da görüntü ögesinin yeni değeri olarak atanır[34].



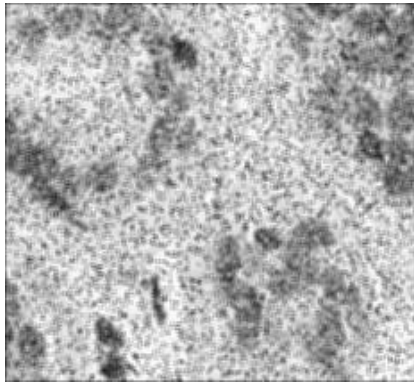
Yöntem olarak basit bir yapısının olmasının yanı sıra, ortalama süzgeci gibi doğrusal süzgeçlere göre daha performanslı sonuçlar üretmektedir. Şeklin kenar yapıları da ortalama süzgecine göre daha iyi korunur. Medyan süzgecinin ortalama süzgeciyle kıyaslanması Şekil 4.6'da görülebilir.



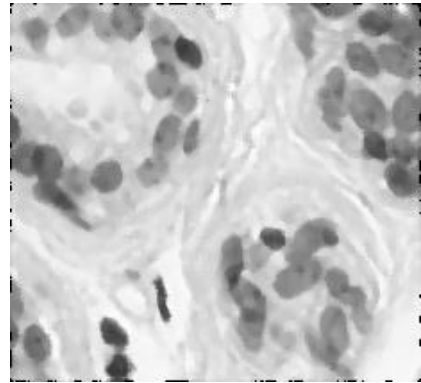
(a)



(b)



(c)



(d)

Şekil 4. 6 (a) 2 nolu histopatolojik resim (b) gürültü eklenmiş hali (c) ortalama süzgeç sonucu (d) medyan süzgeç sonucu

#### 4.4 Canny Kenar Belirlemesi

Bir resmin içindeki nesnelerin yapısal olarak belirlenmesini sağlayan özelliklerinden biri de kenar özellikleridir. Canny Kenar Belirlemesi (Canny Edge Detection) de bir resimdeki kenar özelliklerinin ortaya çıkarılmasında kullanılan yöntemlerden biridir [35]. Adım adım şu şekilde uygulanır [36]:

**Adım 1.** Kenarların belirlenmesini engelleyen gürültülerin giderilmesi için resme Gauss Süzgeci uygulanır.  $(2k + 1)$  kenarlı bir Gauss Süzgeç penceresi şu denklem ile oluşturulmaktadır:

$$H_{i,j} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-k-1)^2 + (j-k-1)^2}{2\sigma^2}\right) \quad (4.13)$$

Oluşturulan bu süzgeç, daha sonra her bir görüntü ögesine evrişim şeklinde uygulanır ve böylece gürültülerin büyük bir kısmı giderilmiş olur:

$$B = A * H \quad (4.14)$$

**Adım 2.** Kenarların yatay, dikey ve çapraz yönelimlerini tespit etmek için 4 süzgeç kullanılır. Resmin yatay yönde birinci türevi ( $G_x$ ) ve dikey yönde birinci türevi ( $G_y$ ) kullanılarak kenarın eğilimi (gradient) ve yönü bulunur:

$$G = \sqrt{G_x^2 + G_y^2} \quad (4.15)$$

$$\theta = \text{atan2}(G_y, G_x) \quad (4.16)$$

Daha sonra kenar yönelim açısı, 4 yönü temsil eden açılardan birine ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  ve  $135^\circ$ ) yuvarlanır. Mesela 0 ile 22.5 arası 0 derece olarak kabul edilebilir.

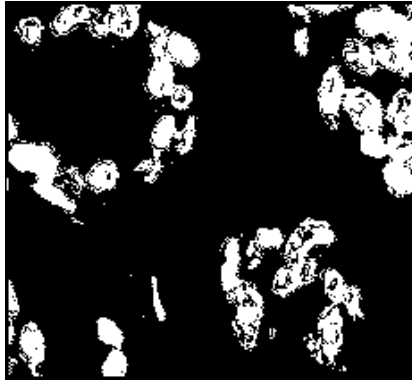
**Adım 3.** Eğilim hesabından sonra bulanık kenarları inceltmek için Azami Olmayanı Bastırma (Non-Maximum Suppression) yöntemi uygulanır. Bunun için ilgili görüntü ögesinin eksi ve artı eğilim yönlerindeki kenar gücü kıyaslanır. Eğer ilgili görüntü ögesinin kenar gücü aynı çerçevedeki ve yöndeki diğer görüntü ögelerine göre daha fazlaysa değeri korunur, küçükse sıfırlanır.

**Adım 4.** Azami olmayanı bastırma uygulamasından sonra, eğilim değeri düşük kenarlı görüntü ögelerini süzmek ve yüksek eğilimli kenar görüntü ögelerini muhafaza etmek için deneysel olarak belirlenen iki eşik değeri kullanılır: Yüksek eşik ( $T_H$ ) düşük eşik ( $T_L$ ). Eğer kenar görüntü ögesinin eğilim değeri:

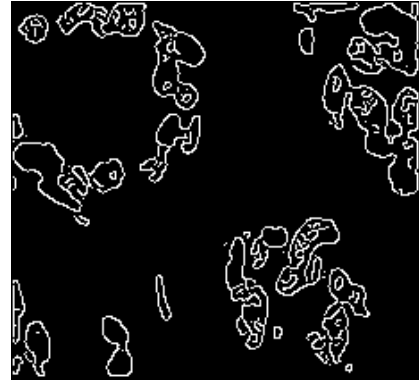
- $T_H$  den büyükse, o değer güçlü kenar görüntü ögesi olarak işaretlenir.

- $T_H$  den küçük; ama  $T_L$  den büyükse, zayıf kenar görüntü ögesi olarak işaretlenir.
- $T_L$  den de küçükse sıfırlanır.

**Adım 5.** İkili eşikleme uygulamasından sonra güçlü ve zayıf kenarlar belirlenmiş olur. Şeklin kenarlarının belirgin olanları ortaya çıkarılır. Zayıf kenarlardaki görüntü öğelerinin çevresindeki 8 öğeden en az birinin güçlü bir kenara ait olup olmadığı denetlenir. Eğer bir tane bile güçlü kenar görüntü ögesi bağlantısı yoksa o kenar sıfırlanır, varsa güçlü kenar olarak kabul edilir.



(a)



(b)

Şekil 4. 7 (a) 2 nolu histopatolojik resmin Rassal Orman çıktısı (b) Canny Kenar Belirlemesi sonuç resmi

#### 4.5 Gizli Markov Rastgele Alanları – Beklenti Enbüyüklenmesi

Markov Rastgele Alanlar teorisi, olasılık kuramının uzamsal (spatial) veya bağlamsal (contextual) bağımlılıkları incelediği bir çalışma konusudur. Temel olarak doku ve nesne özellikleri gibi içerik bağımlı örüntülerin önsel (a priori) olasılıklarının modellenmesi ile ilgilidir. MRA sahip olduğu nesnel fonksiyonlarının uygun değer çözümünü sağlayabilmek amacıyla istatistiksel karar verme ve kestirim yöntemlerini kullanılır. Bu bağlamda kullanılan en popüler istatistiksel kıstas 'En Büyük Sonsal' (EBS, Maximum a posteriori) olasılığıdır. Çerçeve olarak MRA-EBS, özellikle görüntü ve video işleme alanlarında bölütleme, iyileştirilme, doku analizi, kenar belirleme ve görsel birleştirme gibi çeşitli temel işlemlerde yararlanılmakta olan bir yöntemdir.

Gizli Markov Rastgele Alanları – Beklenti Enbüyüklenmesi (GMRA-BE, Hidden Markov Random Fields with Expectation Maximization) yöntemi ise ilk olarak beyin MR görüntülerini bölütlemek amacıyla Zhang v.d. [15] tarafından kullanılmıştır. GMRA ile görüntülerde birbiri ile uzamsal komşulukların birbirine etkisi modellenenmektedir. Bu modelleme gerçekleştirilirken gerekli parametrelerin en iyileştirilmesi ise ‘beklenti enbüyüklenmesi’ (BE, Expectation Maximization) algoritması kullanılarak gerçekleştirilmektedir.

Verilen bir histopatolojik görüntüde ( $x = x_1, x_2, \dots, x_n$ ) her bir  $x_i$  yoğunluk değerini gösterebilir. Bu görüntüde her bir öğeye denk düşen etiket değerleri ( $y = y_1, y_2, \dots, y_n$ ) belli bir değer almaktadır ( $y_i \in L$  ve ikili bir bölütleme probleminde  $L = \{l_1 = 0, l_2 = 1\}$  olarak tanımlanır.). En doğru bölütlemeyi sağlamak amacıyla gerçek etiketleme ( $\hat{y}$ ), EBS kistasına göre aranır:

$$\hat{y} = \operatorname{argmax}_y \{P(X|Y)P(y)\} \quad (4.17)$$

Burada önsel olasılık  $P(y)$  Gibbs dağılımındadır:

$$P(y) = \frac{1}{Z} \exp(-U(y)) = \frac{1}{Z} \exp\left(-\sum_{c \in C} V_c(y)\right) \quad (4.18)$$

Z normalleştirme sabitini göstermekte olup bölme fonksiyonu olarak adlandırılır. Önsel enerji fonksiyonu olan  $U(y)$  ise olası tüm komite (clique) potansiyellerinin toplamından oluşmaktadır. Ayrıca görüntüde her sınıfa ait ( $y_i = l$ ) öğe yoğunluklarının Gauss dağılım parametrelerine sahip olduğu varsayılır ( $\theta_i = \{\mu_l, \sigma_l\}$ ).

$$P(x_i|y_i) = g(x_i; \theta_i) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{(y_i - \mu_l)^2}{2\sigma_l^2}\right) \quad (4.19)$$

Böylece koşullu dağılım olasılığı şu şekilde hesaplanabilir:

$$P(x|y) = \prod_i p(x_i | y_i) = \frac{1}{Z'} \exp(-U(x|y)) \quad (4.20)$$

$Z' = (2\pi)^{N/2}$  olmak üzere benzerlik enerjisi ( $U(x|y)$ ):

$$U(x|y) = \sum_i U(x_i|y_i) = \sum_i \left[ -\frac{(x_i - \mu_i)^2}{2\sigma_i^2} + \log(\sigma_{x_i}) \right] \quad (4.21)$$

şeklinde tanımlanır. Koşullu olasılık görüldüğü üzere sınıflara ait Gauss dağılımının parametrelerine bağlıdır:

$$P(x|y, \theta) = \prod_i p(x_i | y_i, \theta_{y_i}) \quad (4.22)$$

En iyileştirilmesi istenen bu parametreler ( $\theta_{y_i} = (\mu_{y_i}, \sigma_{y_i}), \theta = \{\theta_l | l \in L\}$ ) BE algoritması ile elde edilirler. BE algoritmasına temel adımlarıyla kısaca değinecek olursak:

**Adım 1.** Başlangıç parametre kümesi ( $\theta^0$ ) belirlenir.

**Adım 2.**  $t$ . yinelemede koşullu beklenti ( $\theta^t$ ) hesabı şu şekilde gerçekleştirilir:

$$Q(\theta|\theta^t) = E[\ln P(y|x, \theta)|x, \theta^t] = \sum_y P(y|x, \theta^t) \ln P(y|x, \theta) \quad (4.23)$$

**Adım 3.** Bir sonraki kestirimi elde edebilmek için  $Q(\theta|\theta^t)$  ifadesinin en büyüklenmesi gerekir:

$$\theta^{t+1} = \operatorname{argmax}_{\theta} Q\{\theta|\theta^t\} \quad (4.24)$$

Ardından  $\theta^{t+1} \rightarrow \theta^t$  yapılarak B-adımından itibaren işlem adımları tekrarlanır.

BE algoritmasında  $t$  anı için geçerli parametreler için etiket değerleri EBS kestirimi ile belirlenirken sonsal enerji fonksiyonundan yararlanır.

$$y^t = \operatorname{argmax}_x \{P(x|y, \theta^t)P(y)\} = \operatorname{argmin}_x \{U(x|y, \theta^t) + U(x)\} \quad (4.25)$$

Etiket değerleri en iyilenirken kullanılacak sonsal enerji ise şu şekilde tanımlanır:

$$U(y|x) = U(x|y) + U(y) + \text{const.} \quad (4.26)$$

Bu durumda EBS kestirimi sonsal enerji fonksiyonunun en küçüklenmesi ile bulunur.

$$\hat{y} = \operatorname{argmin}_y \{P(X|Y)P(y)\} \quad (4.27)$$

#### 5.1. Destek Vektör Makineleri

Destek vektör makineleri (DVM, Support Vector Machines), Vapnik ve Cortes tarafından ortaya atılan, iki sınıfı birbirinden ayırmak için kullanılan bir sınıflandırma algoritmasıdır [37]. Daha sonraki yıllarda, Watkins ve Weston tarafından, çok sınıflı verilerin sınıflandırılması amacıyla çok sınıflı DVM (multi-class SVM) algoritması ortaya atılmıştır[38].

Bu algorithmada veri kümesindeki etiketli eğitim verisi kümesi  $\mathbf{X} = \{\mathbf{x}_i, r_i\}$ ,  $i \in 1, 2, \dots, m$  şeklinde bir model oluşturulur. Etiket bilgileri ise  $r_i \in \{-1, +1\}$  olarak kullanılabilir. DVM, oluşturulduğu modelle veri kümesindeki elemanları birbirinden ayıracak (5.1)'deki gibi bir düzlem oluşturur.

$$f(x) = \mathbf{w}^T \mathbf{x} + b \quad (5.1)$$

Bu ifadede yer alan  $w$  ağırlık vektörü ayırma düzlemine hiperdüzlem (hyperplane) olan normali,  $b$  ise yanlılık (bias) terimini göstermektedir.

Sınıflar doğrusal olarak ayrılabilirse bu sınıfları ayıran sonsuz düzlem oluşturulabilmesi beklenmektedir. Destek vektör makinelerinde ise ayırma düzlemi, her iki sınıfa eşit olan düzlem temel alınarak gerçekleştirilir. Test aşamasında hiperdüzleme en yakın noktalar (destek vektörler) kullanılarak yeni gelen örneğe sınıf ataması yapılır. Destek vektörlerin hiperdüzleme uzaklığı  $1/\|\mathbf{w}\|$  olup. En uygun ayırma düzlemi destek vektörlerden en uzak olandır. Bu yüzden  $2/\|\mathbf{w}\|$  ifadesini maksimize

etmek ya da aynı anlama gelen  $\frac{1}{\|\mathbf{w}\|^2} = \mathbf{w}^T \mathbf{w}$  ifadesini (5.2)'deki kısıt da gözetilerek minimize etmek gerekmektedir.

$$\text{Minimize et: } \mathbf{w}^T \mathbf{w}, \quad \text{kısıtlar: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad \forall i = 1, \dots, N \quad (5.2)$$

Ayırma düzlemi  $H'$ 'ye paralel birinci sınıfa ait  $H_1$  ve diğer sınıfa ait  $H_2$  olmak üzere iki adet ayırma düzlemi(5.3) vardır.

$$H_1 = f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b = 1 \quad (5.3)$$

$$H_2 = f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b = -1$$

$H_1$  ve  $H_2$  arasındaki mesafe eğitim aşamasında maksimize edilmeye çalışılır. Yanlılık( $b$ ) ve hiperdüzlem normalinin( $w$ ) en uygun değerlerini bulmak amacıyla (5.4)'de belirtilen Lagrange denklemi kurulur.

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{\mathbf{w}^T \mathbf{w}} - \sum_i \alpha_i [y_i \langle \mathbf{w}^T \mathbf{x}_i + b \rangle - 1] \quad (5.4)$$

Bu ifadede yer alan  $\alpha_i$ , Lagrange çarpanıdır ve sifıra eşit veya sıfırdan büyük değerler alır. Eğer  $w$  ve  $b$  en uygun değerleri alırsa (5.5)'de gösterildiği gibi kısmi türevleri sıfır olur.

$$\frac{\partial L}{\partial \mathbf{w}} = 0, \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \quad (5.5)$$

$$\frac{\partial L}{\partial b} = 0, \sum_i \alpha_i y_i = 0$$

(5.5)'deki eşitlikler (5.4)'deki ifadede yerlerine koyulur ve düzenlenirse (5.6)'daki eşitlik elde edilir

$$L(\mathbf{w}, b, \alpha_i) = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i \quad (5.6)$$

$$\sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \forall i = 1, \dots, N \quad (5.7)$$

Elde edilen (5.6)'daki denklemin (5.7)'deki kısıtlar gözetilerek minimize edilmesi karesel (quadratic) programlama problemidir.

$H_1$  veya  $H_2$  ayırma düzlemi üzerinde bulunan  $\mathbf{x}'$ 'in  $\alpha$  değerleri sıfırdan farklıdır ve bu vektörlere destek vektör (DV, Support Vector) denilmektedir.

Veri kümelerinin geneli pratikte doğrusal ayrılamayan verilerden oluşur. Böyle durumlarda yumuşak marjin algoritması kullanılır. Bu algoritma en az yanlış sınıflandırma olacak şekilde çalışmaktadır. Doğrusal ayrılan DVM' den farklı olarak  $C$  düzenleme parametresi ve yapay değişken  $\xi$  (slack variable) tanımlanır.

$$\text{minimize} : \mathbf{w}^T \mathbf{w} + C \sum_i \xi_i \quad (5.8)$$

$$\text{kısıtlar: } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall i = 1, \dots, N \quad (5.9)$$

Bu ifadelerde yer alan yapay değişken  $\xi_i$  yanlış sınıflandırmaya karşı toleransı göstermektedir. Bunu anlamı da belli bir hata oranına kadar yanlış sınıftan olan örneklerin diğer sınıfta bulunabilmesine izin verilmesidir.

(5.8)'in ikinci terimi yanlış sınıflandırmaya karşı kullanılan ceza terimidir ve  $C$  ifadesi ikinci terimin eşitliğe katkısını göstermektedir.

Yumuşak marjin yöntemi doğrusal çatı altında kullanılan destek vektör makinesinin bir uzantısıdır. Doğrusal olmayan durumlarda ise çekirdek yöntemler (kernel methods) kullanılmaktadır. Çekirdek yöntemlerin ana düşüncesi vektör uzayını daha yüksek boyutlu uzaya dönüştürmektir.

Yüksek boyutlu Hilbert uzayına geçiş için kullanılan çekirdek fonksiyonu (5.10)'da belirtilmektedir.

$$\mathbf{K}(\mathbf{x}, \mathbf{x}') = \Phi(\mathbf{x})^T \cdot \Phi(\mathbf{x}') \quad (5.10)$$



Bu denklem çekirdek fonksiyonunun  $\Phi$  fonksiyonu ile yüksek boyutlu uzaya izdüştürülen  $\mathbf{x}$  ve  $\mathbf{x}'$  arasındaki mesafenin eşit olduğunu göstermektedir. Çekirdek fonksiyonu kullanılarak minimizasyon işlemi gerçekleştirilirse, doğrusal olmayan sınır elde edilir. İz düşüm uzayındaki ayırma düzlemi şu şekilde ifade edilebilir:

$$\mathbf{w}^T \Phi(\mathbf{x}) + b = 0 \quad (5.11)$$

Çekirdek fonksiyonları ile (5.6)'daki ifade yeniden yazılacak olursa (5.13)'teki kısıtlarla birlikte (5.12)'deki eşitlik elde edilir.

$$L(\mathbf{w}, b, \alpha_i) = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) + \sum_i \alpha_i \quad (5.12)$$

$$\sum_i \alpha_i y_i = 0, \quad \alpha_i \geq 0, \quad \forall i = 1, \dots, N \quad (5.13)$$

Kullanılmakta olan birçok çekirdek fonksiyonu vardır bunlardan bazıları şu şekildedir:

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (\text{Doğrusal Çekirdek fonksiyonu})$$

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}^T \mathbf{x}' + 1)^P \quad (\text{Polinom Çekirdek Fonksiyonu}) \quad (5.14)$$

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (\text{Gauss Çekirdek Fonksiyonu})$$

Destek vektör makineleri ikili sınıflayıcılar oldukları için çoklu sınıflama için ikili sınıfları birleştiren yöntemler sunulmuştur. Çoklu sınıflandırma için en yaygın kullanılan yöntem bire-bir ve bire-hepsi destek vektör makineleridir. Bire-hepsi DVM'lerde sınıf sayısı kadar DVM modeli üretilir. Bire-bir DVM'lerde,  $n$  sınıf sayısı olmak üzere  $n/(n-1)/2$  adet DVM modeli üretilerek sınıflama işlemi gerçekleştirilir.

## 5.2. Rassal Ormanlar

Rassal Ormanlar (RO, Random Forests) algoritması, Breimann[39] tarafından ortaya atılan, birden fazla ağaç öngörücüsünün kendi arasında bağımsız bir şekilde oluşturduğu bir sınıflama algoritmasıdır.

Rassal ormanı oluşturan ağaç topluluklarının büyümesini ve her bir topluluğun en popüler sınıfa oy vermesini sağlamak, sınıflama başarımında önemli artışlar sağlamaktadır. Ağaç topluluklarının büyümesi için torbalama (bagging) mekanizması ile her bir ağaç için rastgele vektör seçilir.

Ormanın büyümesini sağlamak için  $k$ 'inci ağaç için önceki vektörlerden  $(\Theta_1, \dots, \Theta_{k-1})$  bağımsız fakat aynı dağılıma sahip rastgele  $\Theta_k$  vektörü torbalama mekanizması ile seçilir. Bu ağaç, seçilen rastgele vektör  $\Theta_k$  ve ağaç sınıflayıcı  $h(\mathbf{x}, \Theta_k)$  ile büyütülür.

Torbalama (bagging) aynı zamanda önyükleme (bootstrap) olarak da bilinen istatistiksel sınıflama modeli alanında kullanılan bir makine öğrenmesi topluluğu algoritmasıdır. Genellikle ağaç modellerinde çeşitliliği azaltmak ve aşırı öğrenmeyi (over fitting) önlemek için kullanılır. Torbalama yöntemi, verilen  $n$  boyutundaki bir  $D$  veri setinden  $m$  adet yeni  $D_i$  veri setleri oluşturur. Torbalama, rastgele giriş seçimi ile kullanıldığında rastgele ormanın başarımını arttırmaktadır.

Bir sınıflayıcı topluluğu  $h_1(x), h_2(x), \dots, h_k(x)$  ile torbalama yöntemi kullanılarak marjin fonksiyonu (5.16)'daki gibi oluşturulur.

$$mg(x, y) = av_k I(h_k(x) = y) - \max_{j \neq y} av_k I(h_k(x) = j) \quad (5.16)$$

Bu ifadede yer alan  $I(\cdot)$  belirteç fonksiyonudur. Marjinin büyük olması sınıflandırmanın daha iyi olmasını sağlar. Marjin için genelleştirilmiş hata fonksiyonu (5.17)'deki gibidir.

$$PE^* = P_{x,y}(mg(x, y) < 0) \quad (5.17)$$

Bu ifadede yer alan  $x, y$  indisleri  $x, y$  uzayı üzerindeki olasılıkları göstermektedir. Ağaç sayısı arttıkça genelleme hatası belli bir limite ulaşır. Rastgele ormanın ağaç sayısı

arttıkça aşırı öğrenme durumu meydana gelmez fakat genelleme hatasının limit değerine ulaşmasına sebep olur. Rastgele ormanda meydana gelen hata oranı orman içerisinde yer alan iki ağaç arasındaki ilintiye ve ağaçların bireysel olarak dayanıklılığına bağlıdır.

Torbalandırmaz hata tahmini (out of bag error estimation) genelleştirilmiş hata tahmini yapmak için kullanılan bir yöntemdir [40]. Ağaçların büyütülmesi için kullanılan mevcut önyüklemeli örneklerin yaklaşık yüzde otuzu kullanılmamaktadır. Bu durum mevcut ağaç için torbalandırmaz durum olarak adlandırılır. Torbalandırmaz hata tahmininin temelini, ağaçların eklendiği ana birleşim devam ederken ağaç sınıflayıcıların da yüzde otuzunu birleştirmek oluşturur. Hata oranı, birleşen ağaç sayısı ile birlikte azalır ve test veri setinin yakınsadığı noktaya erişir.

$N$  eğitim veri sayısı ve  $M$  test veri sayısı olmak üzere rastgele orman işlem adımları şu şekilde sıralanabilir:

**Adım 1.**  $N$  boyutlu torbalama örnekleri eğitim verisinden seçilir.

**Adım 2.** Budama yapılmaksızın ağaç maksimum boyuta çıkarılır.

**Adım 3.** Ağacın her yaprağında en iyi bölünmeyi sağlayacak değişkenler seçilir.

**Adım 4.** Sınıflandırma yapılırken torbalandırmaz durum indirilir.

**Adım 5.** Ağaçlar ormana eklenmeden önce torbalandırmaz durum hata tahmini hesaplanır.

### DENEY ve UYGULAMA

#### 6.1. EKG Denemeleri

Bu kısımda GKA yönteminin EKG verileri üzerindeki uygulaması ve deney sonuçları anlatılmaktadır. Yapılan işlem adımları genel hatlarıyla şu şekildedir[41]:

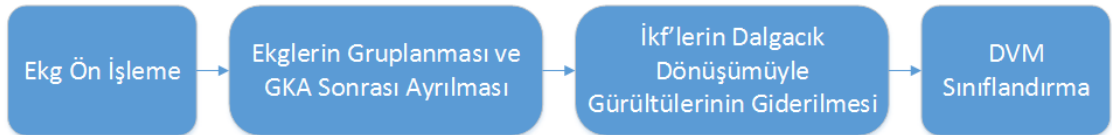
**Adım 1.** EKG verilerinin sınıflama öncesi başarımı arttırabilmek amacıyla insan, şebeke ve benzeri kaynaklı gürültüler çeşitli süzgeçlerle giderilir.

**Adım 2.** GKA ile elde edilecek olan İKF'lerin sayısını arttırmak amacıyla veri kümelerindeki EKG'ler sınıflarına göre gruplandırılarak İKF'lerine ve artık işaretine ayrıştırılır.

**Adım 3.** Gruplara ayrılmış olan EKG işaretleri daha sonra QRS bileşenlerindeki R tepesi merkezde olmak üzere birim vurulara ayrıştırılır.

**Adım 4.** Ayrıştırılan bu İKF'ler çeşitli özellik kümelerine ayrılır ve dalgacık dönüşümüyle kümelerin gürültüleri giderilir.

**Adım 5.** Elde edilen özellikler DVM ile sınıflandırılır ve performansları değerlendirilir.



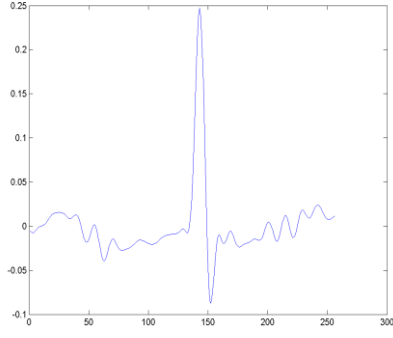
Şekil 6. 1 EKG işaretlerine uygulanacak yöntemlerin genel akış şeması

Bu aşamada PhysioNet üzerinden araştırmacıların hizmetine sunulan çeşitlik EKG verileri içerisinde St-Petersburg INCART [42]ve European ST-T [43] veri kümeleri kullanılmıştır.

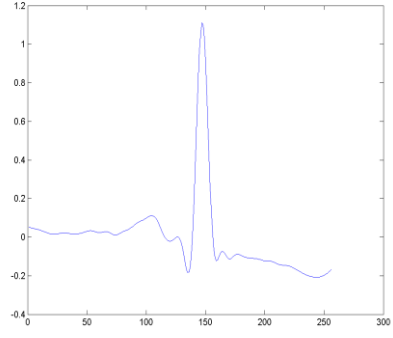
*St-Petersburg INCART aritmi veri kümesi*, 32 holter cihazından elde edilmiş 75 adet etiketlenmiş kayıttan oluşmaktadır. Her bir kayıt 12-başlıktan elde edilen, 275 Hz örnekleme frekansı ile örneklenmiş 30 dakika uzunluğundaki EKG işaretlerinden oluşmaktadır. Veri kümesi toplam 5 sınıftan oluşmaktadır. Bu sınıflar; atriyal erken atım, karıncık atım füzyonu, normal atım, sağ dal blok atımı ve erken karıncık kasılması olarak belirlenmiştir.

Her bir sınıftan 200 eğitim ve 200 test örneği alınarak toplam 1000 örnekten oluşan eğitim veri kümesi ve 1000 örnekten oluşan test veri kümeleri oluşturulmuştur.

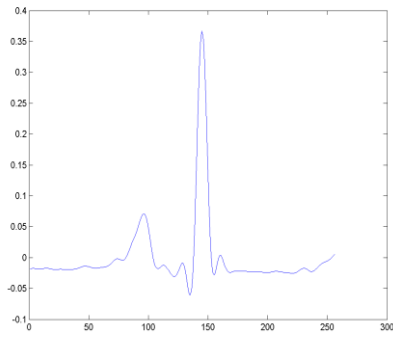
*European ST-T veri kümesi* ise EKG işaretlerindeki ST ve T-dalgalarının değerlendirilmesi için oluşturulmuştur. Veri kümesi 79 hastadan seyyar EKG cihazıyla alınmış 90 adet etiketlenmiş kayıttan oluşmaktadır. Her bir kayıt 2 saat uzunlukta kaydedilmiş olup, iki işaret taşımaktadır. Her bir işaret 250 Hz ile örneklenmiş olup 12 bitlik çözünürlüğe sahiptir. Veri kümesi toplam 4 sınıftan oluşmaktadır. Bu sınıflar; karıncık atım füzyonu, normal atım, supraventriküler atım ve erken karıncık kasılması olarak belirlenmiştir.



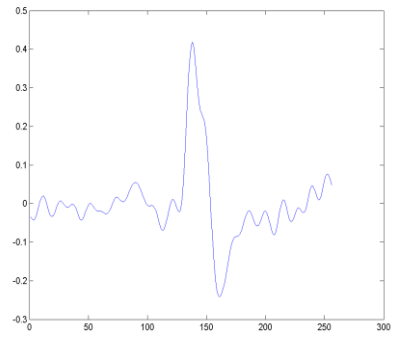
(a)



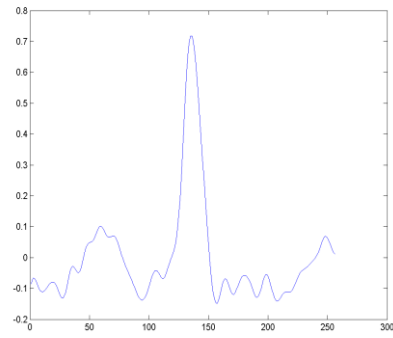
(b)



(c)

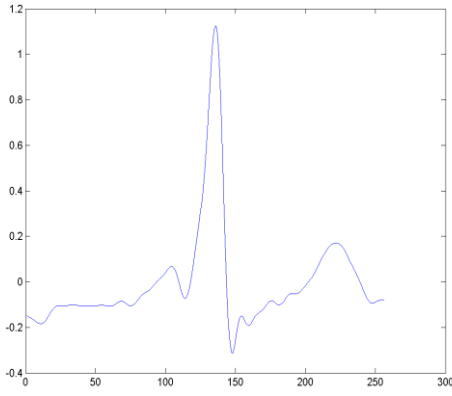


(d)

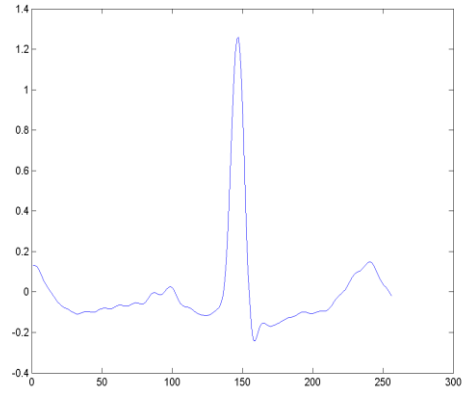


(e)

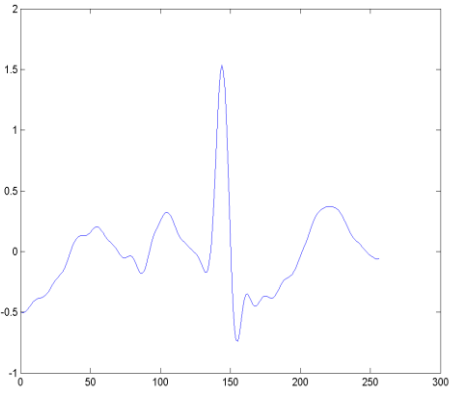
Şekil 6. 2 St-Petersburg INCART aritmi veri kümesindeki 5 EKG türü ((a) atriyal erken atım (b) karıncık atım füzyonu (c) normal atım (d) sağ dal blok atımı (e) erken karıncık kasılması)



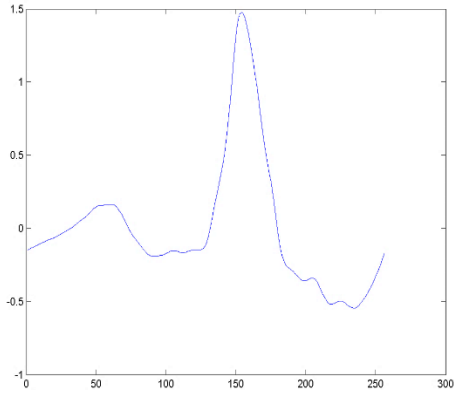
(a)



(b)



(c)



(d)

Şekil 6. 3 European ST-T veri kümesindeki 4 EKG türü ((a) karıncık atım füzyonu (b) normal atım (c) supraventriküler atım (d) erken karıncık kasılması)

Her bir sınıftan 400 eğitim ve 400 test örneği alınarak toplam 1600 örnekten oluşan eğitim veri kümesi ve 1600 örnekten oluşan test veri kümeleri oluşturulmuştur.

EKG verilerinin sınıflama öncesi başarımı arttırabilmek amacıyla çeşitli kaynaklardan gelen gürültülerin temizlenmesi gerekmektedir. EKG cihazlarını etkileyen temel gürültü kaynakları: ülkenin şebeke frekansı gürültüsü (50 veya 60 Hz), kas ve elektrotlardan kaynaklanan yüksek frekanslı gürültüler ve kayıt cihazının kendisinden kaynaklanan temel seviye sürüklenmesidir. Şebeke frekans gürültüsünü gidermek amacıyla çentik süzgeç uygulanmıştır. Temel seviye sürüklenme gürültüsünü gidermek için 0,67 Hz üstündeki frekansları geçiren süzgeç seçilmiştir. Kas ve elektrot kaynaklı gürültünün giderilmesi için 40 Hz kesim frekansındaki alçak geçiren FIR süzgeç kullanılmıştır.

Çizelge 6. 1 EKG'lerden çıkarılan İKF'leri sınıflandırmada kullanılan özellik kümeleri

Özellik Kümesi	Açıklama
ÖK-1	Artık işareti çıkarıldıktan sonra İKF'lere dalgacık dönüşümü uygulanmıştır.
ÖK-2	Artık ve öncesindeki 1. İKF çıkarılıp, kalan İKF'lere dalgacık dönüşümü uygulanmıştır.
ÖK-3	Artık ve öncesindeki 1. ve 2. İKF'ler çıkarılıp, kalan İKF'lere dalgacık dönüşümü uygulanmıştır.

Öncelikli olarak karşılaştırma amacıyla orijinal işaretlerin sınıflandırma performansı DVM'leri ile belirlenmiştir. DVM sınıflandırması Matlab® ortamında[44], GKA ise R platformunda[45] gerçekleştirilmiştir.

ST-Petersburg veri kümesi için orijinal işaretlerle elde edilen DVM sınıflandırma başarımı **%86.00** olarak bulunmuştur. European ST-T veri kümesindeki DVM sınıflandırma yüzdesi ise **%87.17** olarak belirlenmiştir.

Veri kümelerindeki EKG işaretleri ise 10'lu gruplar halinde birleştirilmiş ve ardından GKA gerçekleştirilmiştir. Ardından elde edilen İKF'ler üzerinde Çizelge 6.2 ve Çizelge 6.3'te gösterilmekte olan dalgacık tipleri kullanılarak 3 seviyeli evrensel eşikleme yöntemi ile gürültü giderimi gerçekleştirilmiştir.



Çizelge 6. 2 Orijinal sınıflandırma başarısı %86.00 olan St. Petersburg aritmi veri kümesi için elde edilen yüzdelerle sınıflandırma başarımları

Dalgacık Tipi	ÖK-1	ÖK-2	ÖK-3
DB1 (Haar)	88.3	87.8	87.3
DB2	87.6	87.4	87.7
DB4	87.1	87.4	87.4
DB6	87.4	87.2	87.3
COIF1	88.1	87.7	88
COIF3	87.4	87.4	87.3
COIF5	87.9	87.3	87.7
SYM2	87.6	87.4	87.7
SYM4	87.7	87.7	88.1
SYM8	86.9	87.1	87.6
DMEY	86.5	87.1	86.6
BIOR1.3	87.8	87.8	88
BIOR2.4	87.4	87.3	87.7
BIOR3.1	87.7	87.7	87.3

Çizelge 6. 3 Orijinal sınıflandırma başarısı %87.17 olan European ST-T veri kümesi için elde edilen yüzdelik sınıflandırma başarımları

Dalgacık Tipi	ÖK-1	ÖK-2	ÖK-3
<b>DB1 (Haar)</b>	86.83	87.17	86.83
<b>DB2</b>	87	86.75	87.08
<b>DB4</b>	87.25	86.83	86.5
<b>DB6</b>	87.08	87.25	86.25
<b>COIF1</b>	87.08	86.42	86.67
<b>COIF3</b>	87	86.58	85.83
<b>COIF5</b>	87	86.75	85.58
<b>SYM2</b>	87	86.75	86.5
<b>SYM4</b>	87.25	86.67	86.17
<b>SYM8</b>	87	86.42	86.25
<b>DMEY</b>	86.42	86.17	84.92
<b>BIOR1.3</b>	<b>87.67</b>	86.58	86.5
<b>BIOR2.4</b>	86.75	86.25	86.08
<b>BIOR3.1</b>	86.83	86.33	86.25

DVM'lerinde radyal tabanlı çekirdek fonksiyonlu seçilerek, ceza parametresi (c) 1-50 aralığında; ( $\gamma$ ) parametresi ise 0.01-10 aralığında belirli adımlarla arttırılarak en yüksek sınıflandırma başarımları bulunmuştur. Elde edilen sonuçlar St. Petersburg aritmi veri kümesi için Çizelge 6.2 ve European ST-T veri kümesi için de Çizelge 6.3'te sunulmuştur.

Tablolardan elde edilen sonuçlara göre oluşturulan bütün özellik kümeleri için orijinal işaretlerle elde edilen sınıflandırma başarımından daha yüksek sonuçlara ulaşılmıştır. Her iki veri kümesi için de elde edilen en yüksek başarımların ÖK-1 kullanılarak elde edilmiştir. St. Petersburg aritmi veri kümesi için en yüksek başarımların sonucu **Haar (DB1)** dalgacık tipi ile **%88.30**; European ST-T veri kümesi için ise en yüksek başarımların sonucu **Biorthogonal 1.3** dalgacık tipi ile **%87.67** olarak bulunmuştur.

## 6.2. Histopatolojik Görüntü Denemeleri

Bu kısımda GKA yönteminin histopatolojik resimler üzerindeki uygulaması ve deney sonuçları anlatılmaktadır. Yapılan işlem adımları genel hatlarıyla şu şekildedir:

**Adım 1.** İlgili resmin RGB renk uzayındaki R, G, B katmanlarının ve  $La^*b^*$  renk uzayındaki L,  $a^*$  ve  $b^*$  katmanlarının özellikleri GGKA ile çıkarılır.

**Adım 2.** GGKA işlemi sonrasında ilgili resimler belirli eğitim ve test sınıflarına bölünür ve tüm katmanlardaki İKF'lerin gürültüleri 5X5 lik medyan süzgeç ile giderilir.

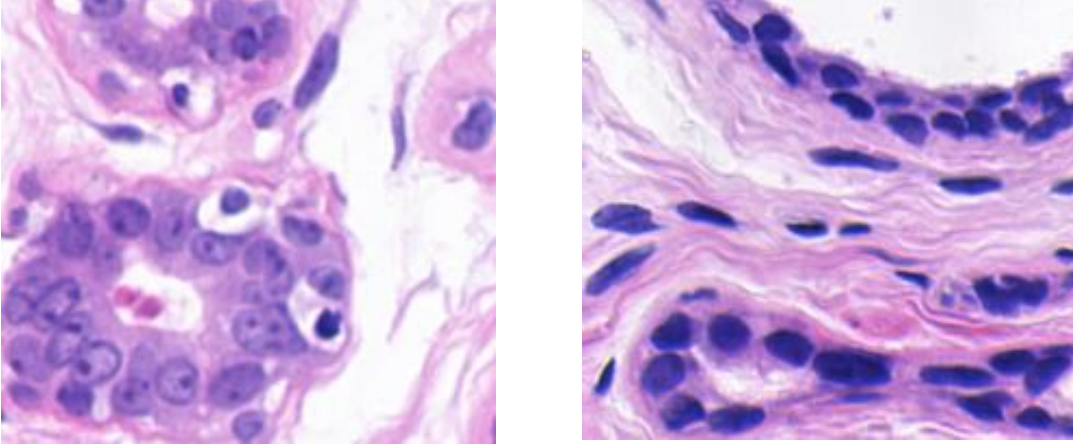
**Adım 3.** Gürültüsü giderilmiş İKF'ler birleştirilir ve elde edilen veriler ile eğitim ve test kümeleri oluşturulur. Ardından bu kümeler, Rassal Ormanlar algoritması ile sınıflandırılır.

**Adım 4.** GMRA-BE aşamasında taban resim olarak, ilgili resimlerin R katmanlarındaki İKF'leri temel alınır. Kenar belirleme aşamasında Canny Kenar Belirleyicisi kullanılır ve son olarak morfolojik işlemlerle süreç tamamlanır. Elde edilen sonuçlar, veri seti içindeki Yer Gerçeklikleri(YG, Ground Truth) ile birebir karşılaştırılarak performans ölçümü yapılır.



Şekil 6. 4 Histopatolojik resimlere uygulanacak yöntemlerin genel akış şeması

Bu çalışmanın histopatolojik görüntü işleme kısmında, Yale Üniversitesi'ndeki David Rimm Laboratuvarı tarafından sağlanan histopatolojik resimler kullanılmıştır. Bu veri kümesinde 32 iyi huylu ve 26 kötü huylu tümör barındıran resimler mevcuttur. Her resmin kendine ait bir yer gerçekliği vardır. Bu veri kümesi özellikle histopatoloji alanında çalışan araştırmacıların değerlendirmelerinde kullanılmak üzere University of California at Santa Barbara (UCSB) Bio-image Informatics Center'de araştırmacıların hizmetine sunulmuştur [46].



(a)

(b)

Şekil 6. 5 (a) iyi huylu hücre örneği (b) kötü huylu hücre örneği

Deney aşamasının bu bölümündeki tüm işlemler Matlab™ R2014a platform üzerinde gerçekleştirilmiştir. Bunun yanında kullanılan diğer yazılım araçları ve parametreleri aşağıda belirtilmektedir:

**IEMD:** Bu araç, Anna Linderhed[47] tarafından hazırlanan bir GGKA aracıdır. Resimlerin RGB ve La\*b\* katmanları bu araçla İKF'lerine ayrılmıştır. Sınırlandırma bakımından, oluşturulması istenen en yüksek İKF sayısı "6" olarak, İKF belirlenmesini sağlayan eşik değeri (epsilon) "0.5" olarak ve üst-alt zarfların edinilmesinde kullanılacak interpolasyon metodu da "thin-plate" olarak belirlenmiştir.

**RandForest:** Bu araç, Abhishek Jaiantilal [48] tarafından oluşturulan bir Rassal Orman sınıflandırma aracıdır. Bu sınıflandırmada, paketteki varsayılan parametreler kullanılmıştır. Veri kümesindeki 58 resmin rastgele seçilen 20'si eğitim kümesi olarak, 38 tanesi de test kümesi olarak belirlenmiştir. Sonuç olarak her görüntü ögesi için ilgili renk uzayına göre 3'er adet özelliği oluşturulmuştur (RGB-EMD ve La\*b-EMD). Hücre görüntü ögesi sayıları ve hücre özelliği barındırmayan görüntü ögesi sayıları 500-2500 arasında seçilmiştir. Çizelgedeki ilk sütunu eğitim kümesindeki görüntü ögesi sayısını, ikinci sütunu test kümesindeki görüntü ögesi sayısını belirtmektedir. 38 resme ait ortalama sınıflandırma sonuçları Çizelge 6.4'te verilmektedir.

Çizelge 6. 4 RGB-EMD ve La\*b\*-EMD'nin Rassal Orman sınıflandırma sonuçları

Eğitim Kümesindeki Görüntü Ögesi Sayısı	Test Kümesindeki Görüntü Ögesi Sayısı	RGB-EMD	La*b*-EMD
1000	1500	85.76	89.30
	2000	85.41	89.91
	2500	85.94	<b>90.12</b>

Bu sınıflandırma süreci, orijinal RGB ve La\*b\* verilerine GGKA uygulanmadan da yapılmıştır (Çizelge 6.5) ve GGKA'nın sınıflandırma performansını artırıcı bir etkiye sahip olduğu gözlenmiştir.

Çizelge 6. 5 Orijinal RGB ve La\*b\* değerlerine RO uygulanması sonuçları

Eğitim Kümesindeki Görüntü Ögesi Sayısı	Test Kümesindeki Görüntü Ögesi Sayısı	RGB	La*b*
1000	1500	89.27	89.21
	2000	89.23	88.94
	2500	89.24	89.22
	3000	<b>89.28</b>	<b>89.25</b>

Görüldüğü üzere, La\*b\*-EMD'deki sınıflandırma başarısı, RGB-EMD'deki sınıflandırma başarısından ve orijinal verilerle elde edilen sınıflandırma başarısından daha iyidir. Bu yüzden bir sonraki adımda La\*b\*-EMD'nin Rassal Orman ile sınıflandırılmış verileri kullanılmıştır.

**HMR-EM-Image:** Bu araç Quan-Wang [49] tarafından hazırlanan GMRA-BE algoritması aracıdır. R katmanındaki resmin kenarlarının belirlenebilmesi için Canny Kenar Belirlemesi[35] yöntemi, eşik (ESK) parametresi 0.50 - 0.75 arasında olacak şekilde kullanılmıştır. Kenarlar belirlendikten sonra R katmanı 3 \* 3 boyutundaki bir Gauss süzgeciyle bulanıklaştırılmış ve ardından GMRA-BE algoritması bulanıklaştırılmış bu resme Rassal Orman sonuçları da temel alınarak yineleme sayısı (EM\_ITER) 1-10 arasında olacak şekilde, EBS yineleme sayısı (MAP\_ITER) 1-80 arasında olacak şekilde uygulanmıştır.

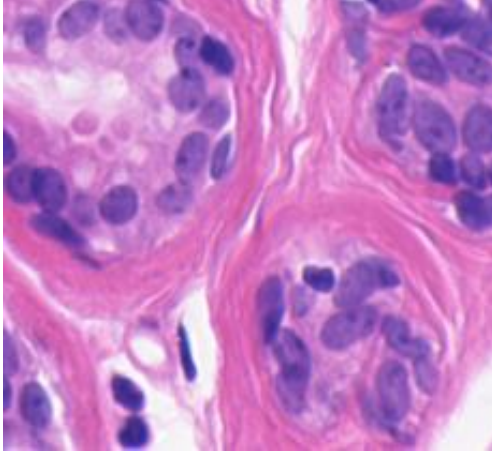
Son olarak sınıflandırma performansını daha da arttırabilmek amacıyla iki Matlab™ fonksiyonu kullanılmıştır. Bunlardan birincisi, sınıflandırmayı etkileyen küçük görüntü ögesi gruplarının silinmesini sağlayan *bwareaopen* fonksiyonudur. Bu çalışmada 100 öğeden daha az yer kaplayan alanları temizlemek için kullanılmıştır. İkincisi de bir grup öğe tarafından çevrilen boşlukların doldurulmasını sağlayan *imfill* fonksiyonudur. Bu fonksiyon da 'holes' parametresiyle kullanılmıştır.

Yapılan bu işlemler sonucunda elde edilen sınıflandırma başarısı Çizelge 6.6'da görülmektedir.

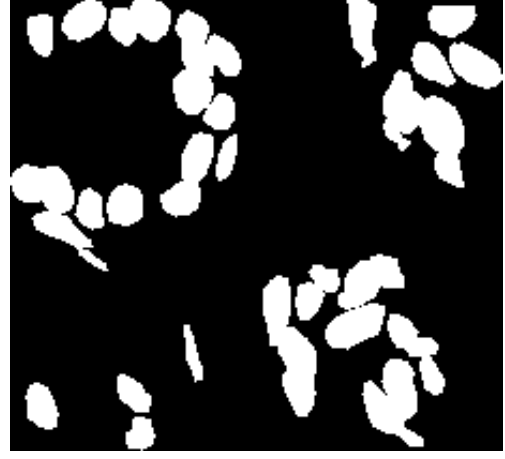
Çizelge 6. 6 RO sonuçlarının işleme sonrası sonuçları

MAP_ITER	EM_ITER	THRS	İşleme Sonrası Sonuçları
1	1	0.75	91.7042
1	2	0.75	91.5270
5	1	0.75	91.8031
5	1	0.50	91.8338
10	1	0.50	91.8674
20	1	0.50	91.8794
40	1	0.50	91.8809
80	1	0.50	<b>91.8829</b>

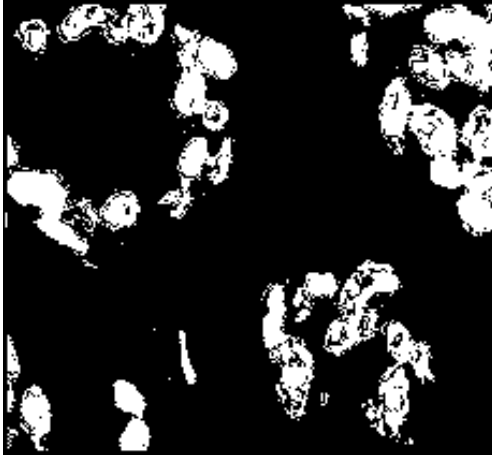
Çizelge 6.6'ten de görüleceği üzere, işleme sonrası süreç, sınıflandırma performansını **90.12**'den **91.88**'e çıkartmıştır. Sadece GGKA bile özellik çıkarma yöntemi olarak öngörülürse, bu durumda sınıflandırma performansının **89.25**'ten **90.12**'ye çıktığı görülür. Örnek bir sınıflandırma süreci Şekil 6.3, Şekil 6.4, Şekil 6.5 ve Şekil 6.6'da görülmektedir. Önerilen yöntem, ilgili resmin yer gerçekliğinde işaretlenmemiş hücreleri bile kolaylıkla tespit etmektedir.



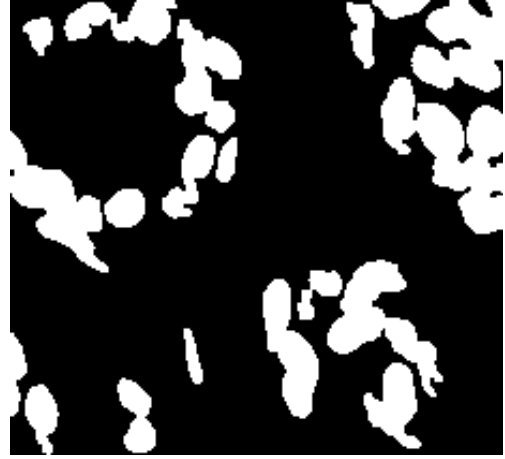
(a)



(b)

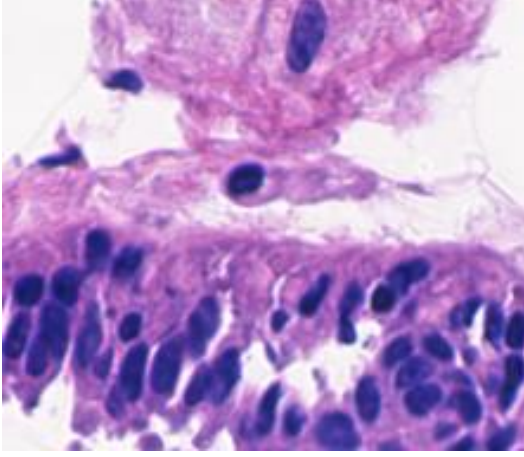


(c)



(d)

Şekil 6. 6 Önerilen sistemin 2 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama)



(a)



(b)



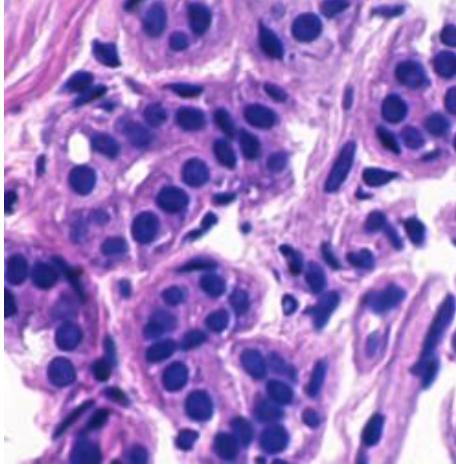
(c)



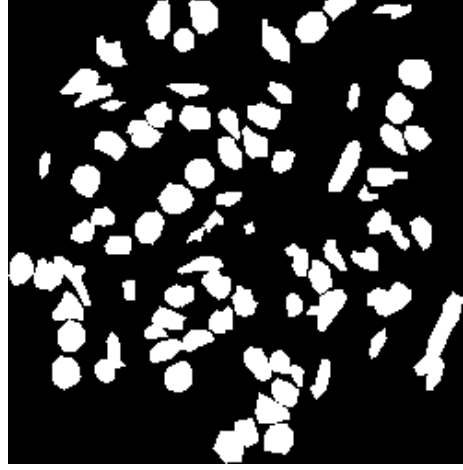
(d)

Şekil 6. 7 Önerilen sistemin 27 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama)

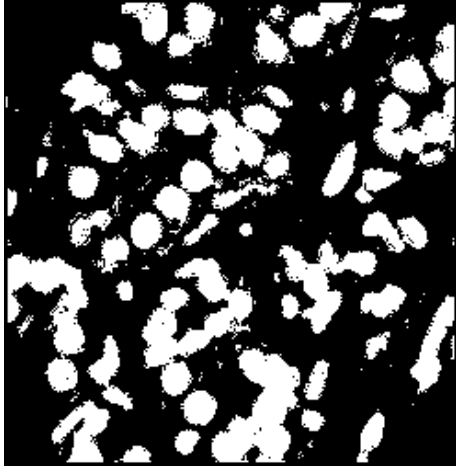




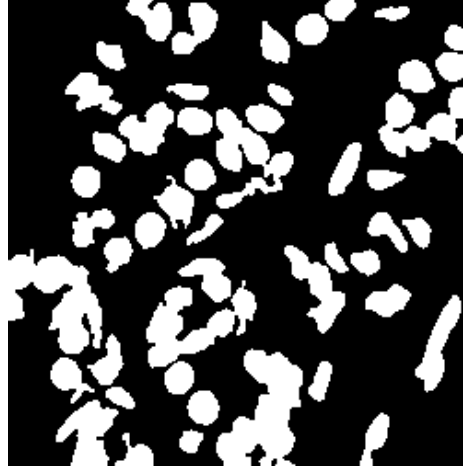
(a)



(b)

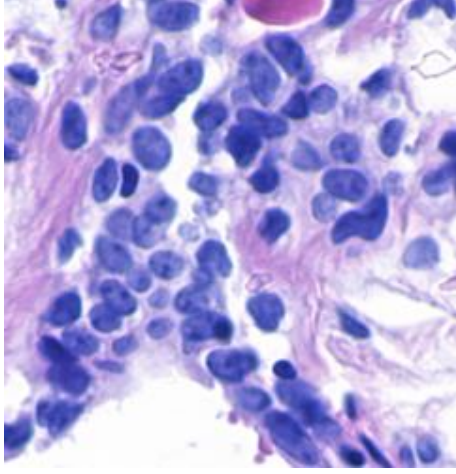


(c)

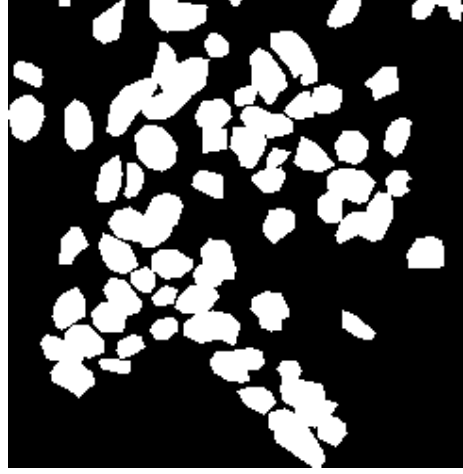


(d)

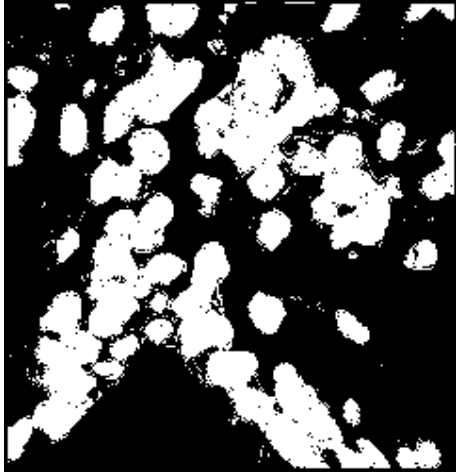
Şekil 6. 8 Önerilen sistemin 25 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama)



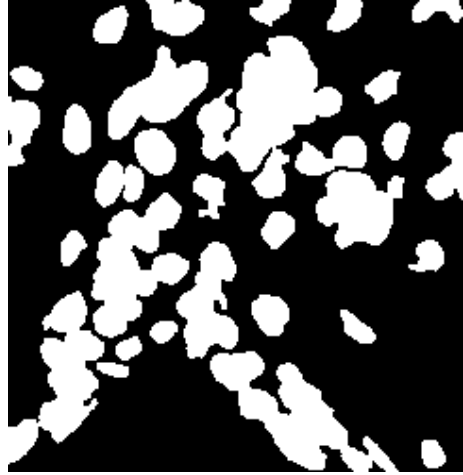
(a)



(b)



(c)



(d)

Şekil 6. 9 Önerilen sistemin 47 nolu histopatolojik resim üzerinde adım adım uygulanması ((a) orijinal histopatolojik resim (b) yer gerçekliği (c) RO sınıflandırması sonrası (d) son aşama)

### SONUÇ

Bu tez çalışmasında, insan sağlığını belirleyici ve erken teşhis sayesinde tedaviyi kolaylaştırıcı etkiye sahip EKG işaretleri ve histopatolojik görüntüler, veriyle uyumlu bir yapıda ayrıştırma özelliğine sahip GKA ile incelenmiş ve sonuçları irdelenmiştir. Veri seti olarak birinci aşamada St-Petersburg INCART ve European ST-T, ikinci aşamada da Yale Üniversitesi'ndeki David Rimm Laboratuvarı tarafından sağlanan histopatolojik resimler kullanılmıştır. Bu inceleme kapsamında eğitici sınıflandırma yöntemi olarak Destek Vektör Makineleri ve Rassal Ormanlar algoritmaları kullanılmıştır.

Biyomedikal verilerin gürültülü yapıları da hesaba katılarak sınıflandırma öncesi ve sonrasında çeşitli süzgeç ve araçlar GKA ile elde edilen özelliklere uygulanmıştır. EKG işaretlerinden elde edilen GKA özelliklerine DD uygulanmış ve rastgele seçilen EKG'lerden oluşturulan çeşitli özellik kümeleriyle DVM algoritması uygulanmış ve sınıflandırma performansının arttığı gözlenmiştir.

Benzer şekilde histopatolojik görüntülerde de RGB ve La\*b\* renk uzayları kapsamında GKA öncesi ve sonrasında gürültü giderme ve morfolojik işlemler uygulanmış ve hücrelerin belirlenebilmesi hususunda belirgin bir artışın olduğu gözlenmiştir.

Sonuç olarak, GKA'nın EKG ve histopatolojik görüntülerden elde ettiği özelliklerin sınıflandırma performansına olumlu etki edecek kadar belirleyici olduğu ve yardımcı araçlarla kullanıldığı takdirde bu sınıflandırmanın daha ileri düzeye taşındığı gözlenmiştir.

## KAYNAKLAR

---

- [1] Huang, N.E. Shen, Z. Long, S.R. Wu, M.C. Shih, H.H. Zheng, Q. Yen, N.-C. Tung, C.C. ve Liu, H.H., (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454:903-995.
- [2] Pan, J. ve Tompkins, W.J., (1985). "A real-time QRS detection algorithm", *Biomedical Engineering, IEEE Transactions*, 230-236.
- [3] Uslu, E. ve Bilgin, G., (2013). "Arrhythmia classification by Local Fractional Fourier Transform", *Signal Processing and Communications Applications Conference (SIU), IEEE*, 1-4.
- [4] Ortiz, M. Bojorges, E. Aguilar, S. Echeverria, J. Gonzalez-Camarena, R. Carrasco, S. Gaitan, M. ve Martinez, A., (2005). "Analysis of high frequency fetal heart rate variability using empirical mode decomposition", *Computers in Cardiology, IEEE*, 675-678.
- [5] Pan, N. Mang, V. ve Un, M.P., (2007). "Accurate removal of baseline wander in ECG using empirical mode decomposition", *Noninvasive Functional Source Imaging of the Brain and Heart and the International Conference on Functional Biomedical Imaging, Joint Meeting of the 6th International Symposium on. IEEE*, 177-180.
- [6] Yang, X.-L. ve Tang, J.-t., (2008). " ", *WiCOM'08. 4th International Conference on. IEEE*, 1-4.
- [7] Khiari, B. Ben Braiek, E. ve Jemni, M., (2013). "R-wave detection using EMD and bionic wavelet transform", *International Conference on Electrical Engineering and Software Applications (ICEESA), IEEE*, 1-5.
- [8] Pinheiro, E. Postolache, O. ve Girão, P., (2012). "Empirical mode decomposition and principal component analysis implementation in processing non-invasive cardiovascular signals", *Measurement*, 45: 175-181.
- [9] De Ridder, S. Neyt, X. Pattyn, N. ve Migeotte, P.-F., (2011). "Comparison between EEMD, wavelet and FIR denoising: Influence on event detection in impedance cardiography", *Engineering in Medicine and Biology Society, IEEE*, 806-809.

- [10] Karagiannis, A. ve Constantinou, P., (2008). "Comparative study of empirical mode decomposition applied in experimental biosignals", IEEE International Conference on BioInformatics and BioEngineering, 1-6.
- [11] Gurcan, M.N. Boucheron, L.E. Can, A. Madabhushi, A. Rajpoot, N.M. ve Yener, B., (2009). "Histopathological image analysis: A review", IEEE Reviews in Biomedical Engineering, 2: 147-171.
- [12] Meijering, E., (2012). "Cell segmentation: 50 years down the road [life sciences]", Signal Processing Magazine, IEEE, 29: 140-145.
- [13] Anoraganingrum, D., (1999). "Cell segmentation with median filter and mathematical morphology operation", Proceedings of International Conference on Image Analysis and Processing, 1043–1046.
- [14] Bamford, P. ve Lovell, B., (1998). "Unsupervised cell nucleus segmentation with active contours", Signal Processing, 71: 203-213.
- [15] Zhang, Y. Brady, M. ve Smith, S., (2001). "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm", IEEE Transactions on Medical Imaging, 20: 45-57.
- [16] Huang, C.-H. ve Lee, H.-K., (2012). "Automated mitosis detection based on exclusive independent component analysis", International Conference on Pattern Recognition (ICPR), 1856-1859.
- [17] Hošťálková, E. ve Procházka, A., (2007). "Complex wavelet transform in biomedical image denoising".
- [18] Ben Saïd, T. Azaiz, O. Chaieb, F. M'hiri, S. ve Ghorbel, F., (2010). "Segmentation of liver tumor using HMRF-EM algorithm with Bootstrap resampling", International Symposium on I/V Communications and Mobile Network (ISVC), , 1-4.
- [19] Pang, B. Zhang, Y. Chen, Q. Gao, Z. Peng, Q. ve You, X., (2010). "Cell nucleus segmentation in color histopathological imagery using convolutional networks", Chinese Conference on Pattern Recognition (CCPR), 1-5.
- [20] Krinidis, M. Krinidis, S. ve Chatzis, V., (2011). "A novel workspace for image clustering", International Conference on Digital Signal Processing (DSP), IEEE, 1-6.
- [21] Sertel, O. Kong, J. Lozanski, G. Shana'ah, A. Catalyurek, U. Saltz, J. ve Gurcan, M., (2008). "Texture classification using nonlinear color quantization: Application to histopathological image analysis", International Conference on Acoustics, Speech and Signal Processing, IEEE, 597-600.
- [22] Gallix, A. Górriz, J. Ramírez, J. Illan, I. ve Lang, E.W., (2012). "On the empirical mode decomposition applied to the analysis of brain SPECT images", Expert Systems with Applications, 39: 13451-13461.
- [23] Chen, W.-K. Lee, J.-C. Han, W.-Y. Shih, C.-K. ve Chang, K.-C., (2013). "Iris recognition based on bidimensional empirical mode decomposition and fractal dimension", Information Sciences, 221: 439-451.

- [24] Mahmood, Q. Chodorowski, A. ve Persson, M., (2015). "Automated MRI brain tissue segmentation based on mean shift and fuzzy c-means using a priori tissue probability maps", *IRBM*, 36:185-196.
- [25] Albayrak, A. ve Bilgin, G., (2013). "Breast cancer mitosis detection in histopathological images with spatial feature extraction", *Sixth International Conference on Machine Vision (ICMV 13)*, 90670L-90670L-6.
- [26] Wu, Z. ve Huang, N.E., (2009). "Ensemble empirical mode decomposition: a noise-assisted data analysis method", *Advances in Adaptive Data Analysis*, 1: 1-41.
- [27] Diebold, F.X., (1998). *Elements of forecasting*: South-Western College Pub, Ohio.
- [28] Flandrin, P. Rilling, G. ve Goncalves, P., (2004). "Empirical mode decomposition as a filter bank", *Signal Processing Letters, IEEE*, 11: 112-114.
- [29] Wu, Z. ve Huang, N.E., (2004). "A study of the characteristics of white noise using the empirical mode decomposition method", *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460:1597-1611.
- [30] Linderhed, A., (2002). "2D empirical mode decompositions in the spirit of image compression", *SPIE Wavelet and Independent Component Analysis Applications*, 4738:1-8.
- [31] Torrence, C. ve Compo, G.P., (1998). "A practical guide to wavelet analysis", *Bulletin of the American Meteorological society*, 79: 61-78.
- [32] Mahmoodabadi, S. Ahmadian, A. Abolhasani, M. Eslami, M. ve Bidgoli, J., (2006). "ECG feature extraction based on multiresolution wavelet transform", *IEEE-EMBS 2005*, 3902-3905.
- [33] Matheron, G. ve Serra, J., (2002). "The birth of mathematical morphology", *Mathematical morphology*.
- [34] Gonzalez, R.C. Woods, R.E. ve Eddins, S.L., (2004). *Digital image processing using MATLAB*, Pearson Education, India.
- [35] Canny, J., (1986). "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 679-698.
- [36] Kalra, P.K., (2009). "Canny Edge Detection Lecture Note", <http://www.cse.iitd.ernet.in/~pkalra/csl783/canny.pdf>, 05 Haziran 2015.
- [37] Cortes, C. ve Vapnik, V., (1995). "Support-vector networks", *Machine learning*, 20: 273-297.
- [38] Weston, J. ve Watkins, C., (1999). "Support vector machines for multi-class pattern recognition", *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, 99:219-224.
- [39] Breiman, L., (2001). "Random forests", *Machine learning*, 45: 5-32.

- [40] Wolpert, D.H. ve Macready, W.G., (1999). "An efficient method to estimate bagging's generalization error", *Machine learning*, 35: 41-55.
- [41] Karaaslan, O.F. ve Bilgin, G., (2014). "ECG classification with empirical mode decomposition denoised by wavelet transform", *Signal Processing and Communications Applications Conference (SIU), 2014 22nd. IEEE*, 694-697.
- [42] 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., (2000). "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals", *Circulation*, 101: e215-e220.
- [43] Taddei, A. Distanti, G. Emdin, M. Pisani, P. Moody, G. Zeelenberg, C. ve Marchesi, C., (1992). "The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography", *European heart journal*, 13: 1164-1172.
- [44] Chang, C.-C. ve Lin, C.-J., (2011). "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3): 27.
- [45] Kim, D. ve Oh, H.-S., (2009). "EMD: A package for empirical mode decomposition and hilbert spectrum", *The R Journal*, 1: 40-46.
- [46] Gelasca, E.D. Obara, B. Fedorov, D. Kvilekval, K. ve Manjunath, B., (2009). "A biosegmentation benchmark for evaluation of bioimage analysis methods", *BMC bioinformatics*, 10: 368.
- [47] Linderherd, A., (2009). "Image empirical mode decomposition: A new tool for image processing", *Advances in Adaptive Data Analysis*, 1: 265-294.
- [48] Jaientilal, A., "Random Forest implementation for MATLAB ", <https://code.google.com/p/randomforest-matlab/> , 5 Haziran 2015.
- [49] Wang, Q., (2012). "HMRF-EM-image: implementation of the Hidden Markov random field model and its expectation-maximization algorithm", *arXiv preprint arXiv:1207.3510*.

## ÖZGEÇMİŞ

---

### KİŞİSEL BİLGİLER

**Adı Soyadı** : Ömer Faruk KARAASLAN  
**Doğum Tarihi ve Yeri** : 20.05.1989, Eyüp  
**Yabancı Dili** : İngilizce, Fransızca  
**E-posta** : [omerfk@yildiz.edu.tr](mailto:omerfk@yildiz.edu.tr)

### ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Lisans	Bilgisayar Müh.	Yıldız Teknik Üniversitesi	2012
Lise	Sayısal	Kabataş Erkek Lisesi	2008

### İŞ TECRÜBESİ

Yıl	Firma/Kurum	Görevi
2014 – Devam	YTÜ Merkez Laboratuvarı	Araştırma Görevlisi
2012 – 2012	Açı Eğitim Öğretim Hizmetleri A.Ş.	iOS Yazılımcısı
2012 – 2012	Eyüp Belediyesi	Donanım Stajyeri
2011 – 2011	Łódz Teknik Üniversitesi	Java Yazılımcısı



## **YAYINLARI**

### **Ulusal Bildiri**

- 1.(2014) **Karaaslan Ö. F.**, Bilgin G., Dalgacık Dönüşümüyle Gürültüsü Giderilmiş Görgül Kip Ayrışımı ile EKG Sınıflandırma, SİU, 2014