

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

Rasım ÇEKİK¹, Sedat TELÇEKEN^{1,2}

EKG SİNYALLERİNİN KABA KÜMELER TEORİSİ KULLANILARAK SINIFLANDIRILMASI

ÖZ

Kaba kümeler teorisi (KKT), bulanık kümeler gibi uzman sistemler için analiz ve bilgi çıkarımında kullanılan kural tabanlı bir yöntemdir. Kaba kümeler eksik, tutarsız ve belirsiz veri kümelerini düzenleyerek değerlendirme için uygun hale getirmektedir. Bu çalışma EKG sinyallerini, kaba kümeler teorisi tabanlı yeni bir modele göre sınıflandırmayı önerir. Eksik, gereksiz ve tutarsız veri kümeleri günümüzde en çok hastalık verilerinde karşımıza çıkmaktadır. Bu veri kümelerinden doğru kural çıkarmak veya bu veri kümelerini doğru sınıflandırmak, hastalık teşhisinde son derece önemlidir. EKG sinyallerinin verilerinin bulunduğu veri kümesine öngörülen modelin uygulanması, sınıflandırmada çalışma zamanında iyileştirmeyi sağlamıştır. Ayrıca bu çalışma ile en az parametre kullanarak, doktorların daha hızlı ve erken teşhisine yardımcı olacak bir model oluşturulmuştur. Bu model ile EKG sinyalleri % 85 yakın bir doğruluk ile sınıflandırılmıştır.

Anahtar Kelimeler: Kaba kümeler teorisi, EKG, Sınıflandırma, Uzman sistemler, Büyük veri analizi, Veri madenciliği.

CLASSIFICATION OF EKG SIGNALS USING ROUGH SETS THEORY

ABSTRACT

Rough sets theory (RST) is a rule-based method used for the analysis and data mining in expert systems such as fuzzy sets. Rough sets organize data sets with missing, inconsistent and ambiguous data and make them suitable for analysis and evaluation. This paper proposes a new rough sets theory -based model for the classification of EKG signals. Missing, unnecessary and inconsistent data sets are encountered mostly in patient data. For correct diagnosis, it is very important to correctly classify and extract rules from these data sets. The application of the proposed method to a data set containing EKG signals improves the running time performance of classification. Additionally, the proposed method requires minimal number of parameters and can be used as an aid for doctors for faster and early diagnosis. EKG signals are classified correctly up to 85% by this model.

Keywords: Rough sets theory, EKG, Classification, Expert systems, Big data analysis, Data mining.

¹. Anadolu Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İki Eylül Kampüsü, Eskişehir.
E-posta: rasimcekik@anadolu.edu.tr, Tel: +90 222 335 05 80 /6569.

². E-posta: stelceken@anadolu.edu.tr, Tel: +90 222 321 35 50 /6563.

1. GİRİŞ

Hayatın her döneminde kalp sağlığı, sağlık alanının en önemli konusu olmuştur. Kalbin fonksiyonel görevinin öneminden dolayı kalp sağlığı ile ilgili problemleri çözmek son derece önem arz etmektedir. Sağlık alanında birçok konuda olduğu gibi kalp hastalıkları tedavisinde de erken teşhis altı çizilmesi gereken bir durumdur. Kalp hastalıklarının erken teşhisi için kullanılan en yaygın ve en etkili yöntem Elektrokardiyografi (EKG)'dir.

Kalp kasının ve sinirsel iletim sisteminin çalışmasını incelemek üzere tasarlanan bu yöntem, kalp genişlemesi, kalp büyümesi, kalbe giden kan miktarındaki azalma, yeni veya eski kalp hasarları, kalp ritim problemleri ve değişik kalp ve kalp zarı hastalıkları hakkında önemli bilgiler verebilmektedir. Bir EKG sinyali; *P*, *Q*, *R*, *S* ve *T* adları verilen dalgalardan, *ST* ve *PR* segmentlerinden ve *RR*, *QT* ve *PR* aralıklardan oluşan bir sinüs sinyalidir. Normal bir EKG sinyalinde *P*, *Q*, *R*, *S*, *T* dalgaları belirli bir sıra ve aralıklarla oluşurlar. Dalgaların bu sıra ve aralıklarla oluşumu dalganın şekli, süresi, *ST segmenti*, *RR aralık* gibi parametrelerini içermektedir. Bu parametrelerdeki anormallikler, tek başına ya da başka tahlillerle birlikte kalp sorunları hakkında bilgi vermektedir. Örneğin, *RR aralığın* anormal olması kalpte ritim bozukluğu olduğunu gösterirken *ST segmentin* çökmesi veya yükselmesi Miyokardiyal Enfarktüsü ya da Miyokardiyal İskemi hastalıklarına işaretir (Efeoğlu, 2013).

Gerçek zamanlı günümüz uygulamalarında, eksik, belirsiz ve tutarsız verilerin olması sık karşılaşılan bir durumdur. Sağlık alanında sezgisel teşhislerin de olmasından dolayı, verilerin rasyonel olarak tanımlanabilmesi sıklıkla güç olmaktadır. Kalbin çalışması, fonksiyonel görevi ve özel yapısı; yukarıda ifade edilen parametreler arasındaki mantıksal bağlantıyı karmaşıklığa sürüklemektedir. Eğer bir veri kümesinde bazı nitelik değerleri bilinmiyorsa, bu veri kümesine eksik veri kümesi denir (Avşar ve Kaya, 2007). Veri kümelerinden çıkarım yapan birçok metot bulunmaktadır. Örneğin, eksik nitelik içeren nesnelere doğrudan silinmesi (Chmielewski ve Grzymala-Busse, 1996), ya da eksik nitelik değerini istatistiksel analiz sonucu elde ettiği mümkün bir değerle ya da eksik veri içeren bu nesne ile aynı giriş verilerine sahip, bir başka nesnenin bu niteliğe ait değeri ile değiştirilmesi (Lingras ve Yao, 1998)

gibi metotlar bulunmaktadır. Bu metotlar eksik veri kümesini tam veri kümesine dönüştürmek için kullanılmaktadırlar. Fakat bu metotlardan yapılan çıkarımlar tümüyle “Kesin kurallar içermektedir” denilemez.

Bu yöntemlerin yanında yüksek başarı gösteren bulanık kümeler (fuzzy sets) ve kaba kümeler (rough sets) gibi yöntemler; özellikle eksik, yetersiz ve belirsiz verileri düzenleyerek, veri analizi için yeterli hale getirmektedir (Telçeken ve Doğan, 2004). Pawlak tarafından ortaya atılan “Kaba Kümeler Teorisi” (KKT) günümüzde çoğu bilim dalında uygulama olarak kendisini kabul ettirmiştir (Pawlak, 1982). KKT eğitim verileri üzerinde yararlı verilerin çıkarımındaki başarısı; veri madenciliği ve veri tabanı üzerine yapılan çalışmalarda da sıklıkla kullanılmasına sebep olmuştur (Avşar ve Kaya, 2007).

Yapılan araştırmalarda ve literatürde EKG'nin tek başına bir kalp hastalığının kesin teşhisi için yeterli olmadığı, bunun yanında mutlaka başka tahkiklerin yapılması gerektiği görülmüştür. Bununla birlikte EKG sinyalleri, teşhis için çok değerli bilgiler taşımaktadır. Bu nedenle EKG'nin doğru yorumlanması çok önemlidir.

EKG ile tanı konulabilecek en genel durumlar; Ritimler, Aktivasyon Serisindeki Bozukluklar, Atria ve Ventricles Duvarında Kalınlaşma veya Boyutunda Büyüme, Miyokardiyal Enfarktüs ve İskemi vb. gibi ana başlıklar altında toplanabilir. Doktorlar bu tanıları EKG kâğıdındaki bilgiler ışığında belirlemektedir. Ancak bir tanıyı koymak için birden çok EKG belirtisini değerlendirmek gerekebilir. Bu işlemler zaman ve dikkat isteyen işlemlerdir. Doktorların dikkatsizliğinden kaynaklanan hatalar yanlış tanı konulmasına ve önemli detayların gözden kaçmasına neden olabilir. Bu çalışma ile bu problemlere çözüm bulma ve doktorların zamandan kazanım sağlaması amaçlanmıştır.

Yapılan çalışmada Ventricles Duvarında Kalınlaşma veya Boyutunda Büyüme ve Miyokardiyal Enfarktüs ve İskemi ana başlığının alt başlık tanıları olan Sol Ventricular Büyümesi, Atrial Büyümesi, Miyokardiyal İskemi ve Miyokardiyal Enfarktüs tanıları üzerinde durulmuştur.

Yürütülen çalışma ile en az parametre kullanılarak, doktorların daha hızlı ve erken teşhisine yardımcı olacak bir model oluşturulmuştur. KKT ve sınıflandırma yöntemlerinden oluşan model, gereksiz, tekrarlı ve eksik verinin sık olduğu EKG alanındaki verileri tutarlı hale getirmektedir. Bu tutarlı veri kümesi en etkili sınıflandırma yöntemi ile sınıflandırmaktadır. Bu sınıflandırma doktorlara daha hızlı ve daha az hata ile tanı koyma imkânı sunar.

2. ELEKTROKARDİYOGRAFI

EKG, kalbin elektriksel aktivitesinin özel kâğıtlara yazdırılma işlemine denilmektedir. EKG esas olarak elektriksel olayları gösterir ve mekanik olaylar hakkında bilgi vermez. EKG, dönme hızı genellikle saniyede 25 mm'ye ayarlanmış bir kâğıt üzerine kaydedilir. Bu EKG kâğıdının üzerinde 1x1 mm'lik küçük ve 5x5 mm'lik büyük kareler mevcuttur.

Bir EKG sinyali; dalga, segment ve aralıklardan oluşan bir sinüs sinyalidir. Dalgalar *P*, *Q*, *R*, *S*, *T* dalgalarıdır. Bu dalgaların tanımlamalarını ilk defa Willem Einthoven (1985), yapmıştır ("Doktor EKG", t.y.).

EKG ile tanı konulabilecek durumları en genel hali ile aşağıdaki gibi sınıflandırabiliriz.

- A. Ritm ve iletim bozuklukları
- B. Kalbin (Miyokardın) kanlanması
 - a. İskemi (yeterli kanlanamama)
 - b. Lezyon (zararlanma)
 - c. Nekroz
- C. Kalp adalesinin durumu (Hipertrofi)
- D. Kan iyonları dengesizlikleri
 - a. Ca, K, Mg

2.1. EKG Üzerindeki Dalga, Segment ve Aralıklar

2.1.1. Dalgalar

Bir EKG sinyalinde *P*, *Q*, *R*, *S*, *T* dalgaları belirli bir sıra ve aralıklarla meydana gelirler.

- P*: Atriyum depolarizasyonu.
Q: *P* dalgasından sonraki ilk negatif dalga.
R: *P* dalgasından sonraki ilk pozitif dalga.
S: *P* dalgasından sonraki ikinci negatif dalga.
T: Ventrikül repolarizasyon dalgası.

2.1.2. Segmentler

PR-Segment: *PR* segmenti *P* dalgası bitişi ve *QRS* kompleksi başlangıcı arasındaki yassı ve genellikle izoelektrik segmenttir (Efeoğlu, 2013).

ST-Segment: *ST* segmenti, *S* dalgası sonu (*J* noktası) ile *T* dalgası arasındaki yassı, izoelektrik EKG kısmıdır. Ventriküler depolarizasyon ve repolarizasyonu gösterir. *ST segment* anormalliklerinin (Elestasyon/ Depresyon) en yaygın sebebi miyokardiyal iskemisi/ enfarktüs (Efeoğlu, 2013).

2.1.3. Aralıklar

QT-Aralık: *QT* aralığı *Q* dalga başlangıcından *T* dalga sonuna kadar geçen süredir. Ventriküler depolarizasyon ve repolarizasyon için geçen süreyi belirtir (Efeoğlu, 2013).

PR-Aralık: *PR* aralığı *P* dalgası ile *QRS* kompleksi arasında geçen süredir. AV düğümünden geçen iletiyi gösterir (Efeoğlu, 2013).

RR-Aralık: İki kompleks *QRS* arasında geçen süre olarak değerlendirilebilir. Anlık kalp ritmi iki kompleks *QRS* arasında geçen süre ile hesaplanabilir. Dolayısıyla kalp ritmi *RR aralık* ile hesaplanabilir. Bu yöntemin kötü tarafı ise sağlıklı bir insanın kalp ritmi anormal çıkabilir veya tam tersi bir durum yaşanabilir (Efeoğlu, 2013).

3. HASTALIKLAR

Miyokardiyal Enfarktüsü (MI); kalp kasının iskemik nekrozudur. Yani kalbin koroner kan dolaşımının belli bir bölgede yetersiz kalması sonucu, o bölgedeki kalp kası dokusunun ölmesidir. Miyokard enfarktüsü sıklıkla koroner damarların ateroskleroz sonucu daralma ya da tıkanmalarına bağlıdır. Miyokard enfarktüsü % 80 vakada kendisini göğüs ortasındaki şiddetli, sıkıştırıcı bir ağrı ile belli eder. % 20 vakada ise ağrı ortaya çıkmaz. Yaşlı ya da şeker hastalığı olan kimselerde, ağrısız miyokard enfarktüslerine daha sık rastlanmaktadır ("Sağlık.im", t.y.).

Miyokardiyal İskemi (MIS); doku perfüzyonunun azalması sonucunda gelişen oksijen yoksunluğu olarak tanımlanabilir (Bilal ve Sarıoğlu, 1992). Halsizlik, terleme, kusma ve nefes alma güçlüğü, gibi belirtileri taşıyan MIS, diyabet hastaları ve yaşlılarda daha sık görülmektedir.

Sol Ventricular Büyümesi (Hipertrofisi) (SVH); hipertansiyonu olan hastalarda kalp sürekli yüksek basınçta kan pompalar. Bu artmış iş yükünün sonucu olarak, akciğerlerden gelen temiz (oksijenlenmiş) kanı vücuda pompalayan kalbin sol karıncığının duvarı (kas tabakası) kalınlaşır. Buna sol ventrikül hipertrofisi denilir. Zamanla bu karıncığın iç boşluğu genişler, fonksiyonları bozulur, temiz kanı vücuda yeterince pompalayamamaya başlar. Sol kalbin pompalama fonksiyonunu yeterince gerçekleştiremediği bu durum sol ventrikül disfonksiyonu olarak adlandırılır (“Abbott Türkiye”, t.y.).

Atrial Büyümesi (Hipertrofisi) (AH); kalbin odacıklarındaki büyüme olarak bilinir. Atriyal genişleme (büyüme) atriyumda basınç veya volüm artmasına bağlı bir problemdir.

4. EKG SİNYALLERİNİN SINIFLANDIRILMASI

4.1. Kural Çıkarımı

4.1.1. Sınıf Belirleme

Bu çalışmada EKG ile tanı konulacak hastalıklar;

- Miyokardiyal Enfarktüs
- Miyokardiyal İskemi
- Sol Ventricular Büyümesi (Hipertrofisi)
- Atrial Büyümesi (Hipertrofisi)

olarak sınıflandırılmıştır. Bunlara ek olarak sağlıklı bireylerin EKG’lerinden oluşan “Normal Birey” sınıfı da eklenmiştir. Bu sınıflar tasarlanmış olan modelin de sınıflarıdır. Hastaların EKG’leri ile hastaların hastalık sınıfları belirlenmiştir. Bunun için en etkili sınıflandırma yöntemleri kullanılmıştır.

4.1.2. Sınıfların EKG Belirtileri

Sınıflar için EKG belirtileri Tablo 1’de verilmiştir.

Tablo 1. Sınıflar İçin EKG Belirtileri

Sınıf	EKG Belirtileri
Miyokardiyal İskemi	ST Segment Çökmesi
	ST Segment Yükselmesi
	T Dalgası Negatif Olması
	T Dalgasının Sivri Olması (Pozitif Olması)
	Arritmeler
	QT Aralığı Uzaması
	Negatif U Dalgası
	Dal Bloğu Olması
	QRS Atternansı
Miyokardiyal Enfarktüs	ST Segment Çökmesi
	St Segment Yükselmesi
	R Dalga Progresyon Kaybı
	T Dalgasının Sivri Olması (Pozitif Olması)
	Patolojik Q Dalgası Olması
Sol Ventricular Büyümesi	QRS Yükseklik Artışı
	ST Segment Yükselmesi
	QT Aralık Uzaması
	QRS Genişlik(Süre) Uzaması
	T Dalgası Negatif Olması
	R Dalga Progresyon Kaybı
Atrial Büyümesi	P Dalgası Uzaması
	P Dalgası Kısa olması
	Negatif P Dalgası Olması
	P Dalgası Çentikli(Çift tepe) olması
	P Dalgasının Düşük olması
	P Dalgası Yüksek Olması

Tablo 1’deki Arritmeler, *QT Aralığı Uzaması*, *Negatif U Dalgası*, *Dal Bloğu Olması*, *QRS Atternansı* belirtilerinin sınıfı belirlemede, yapılan araştırma ve literatürde, zayıf belirtiler olduğu görülmüştür (Malmivuo ve Plonsey, 1995). Bu sebeple bu belirtiler göz ardı edilmiştir.

4.1.3. Nitelik Belirleme

Bir EKG sinyalindeki dalga, aralık ve segmentler EKG parametreleri olarak isimlendirilebilir. Bu parametreler sayesinde EKG yorumlanır.

EKG parametreleri uzman hekimlerle yapılan görüşme ve literatür çalışması ile önem teşkil eden ve önemsiz görülen olarak sınıflandırılmıştır. Çalışmada belirleyici olacak parametreler, nitelik olarak alınmıştır. Buna göre seçilen nitelikler aşağıda verilmiştir.

Tablo 2. Seçilen Nitelikler

Nitelik	Açıklama
Rate	Kalp atış hızı.
ST-Segment	S dalgası sonu (J noktası) ile T dalgası arasındaki yassı, izoelektrik kısımdır. Ventriküler depolarizasyon ve repolarizasyonu gösterir.
T-Dalgası	T dalgasının her QRS kompleksinden sonraki pozitif sapmasıdır. Ventrikül repolarizasyonunu temsil eder.
Patolojik Q	Aşağıdakilerin varlığında Q dalgaları patolojik olarak kabul edilir: <ul style="list-style-type: none"> • > 40 ms (1 mm) genişlik • > 2 mm derinlik • QRS kompleksinin %25'inden daha fazla derinlik • V1-3'te görülmesi
R-Dalga Progresyon Kaybı	V3'te ≤ 3 mm R dalga yüksekliği.
PR-Aralık	PR aralığı P dalgası ile QRS kompleksi arasında geçen süredir. AV düğümünden geçen ileticiyi gösterir.
P-Dalgası Geniş	P dalgasının yüksekliği normal sınırlardadır. Ancak süresi 120 ms'den uzundur.
P-Dalgası Yükseklik	Genişliği değişmeyen (< 120 ms) ancak normalden daha yüksek P dalgaları meydana gelir (2,5 mm) .
QRS-Genişlik	QRS > 100 ms olma durumudur.
QRS-Yükseklik	Voltajın normalinden yüksek olma durumudur.
QTc	Düzeltilmiş QT aralığı (QTc), 60/dk kalp hızındaki QT aralığını hesaplar.
T-Reciprocity	T dalgasındaki karşıtlık.
QT-Aralık	QT aralığı Q dalga başlangıcından T dalga sonuna kadar geçen süredir. Ventriküler depolarizasyon ve repolarizasyon için geçen süreyi belirtir.
Cinsiyet	Hastanın Cinsiyeti

4.1.4. Niteliklerin Değer Kümesinin Belirlenmesi

Yapılan çalışmada her nitelik için değer kümesi belirlenmiştir.

Rate = {B, N, T}, ST Segment = {C, N, Y},
T Dalgası = {P, N, Ne}, Patolojik Q = {V, Y},
R Progresyon = {V, Y}, PR Aralık = {N, A},
P Dalgası Yükseklik = {N, A}, QTc = {N, A},
P Dalgası Genişlik = {N, A}, QRS Genişlik = {K, N, U},
QRS Yükseklik {D, N, Y}, T Reciprocity {V, Y},
QT Aralık = {K, N, U}, Cinsiyet = {C, YK, YE}

Değer Kümesindeki her bir değer farklı anlamları mevcuttur. Bunlar Rate; B: Bradycardia, N: Normal (Sağlıklı) T: Tachycardia, ST segmenti; C: Çökme, N: Normal, Y: Yükselme, T dalgası; P: Pozitif (Sivri); N: Normal, Ne: Negatif, Patolojik Q; V: Var, Y: Yok, R progresyon; V: Var, Y: Yok, PR aralık; N: Normal, A: Anormal, P dalgası yükseklik; N: Normal, A: Anormal, QTc; N: Normal, A: Anormal, P dalgası genişlik; N: Normal, A: Anormal, QRS genişlik; K: Kısa, N: Normal, U: Uzun, QRS yükseklik; D: Düşük, N: Normal, Y: Yüksek, T reciprocity; V: Var, Y: Yok, QT aralık; K: Kısa, N: Normal, U: Uzun, Cinsiyet; C: Çocuk, YK: Yetişkin Kadın, YE: Yetişkin Erkek olarak sınıflandırılmıştır.

4.2. Kaba Kümeler Teorisi

Pawlak tarafından 1982 yılında önerilen kaba küme teorisi, uzman sistemler için değerlendirme ve kural çıkarımında kullanılan matematiksel bir metottur (Grzymala-Busse, 1988; Orłowska, 1994). KKT ve bulanık mantık gibi kuramlar eksik ve belirsizlik bilgi tabanından faydalı bilgi çıkarımını sağlayan teoremlerdir (Avşar ve Kaya, 2007). Şüpheli, eksik, tutarsız gizli bilgilerin sık olduğu sağlık alanında KKT kendine geniş bir çalışma alanı bulmuştur.

4.2.1. Bilgi Sistemi

KKT'nin ilk aşamasını oluşturan bilgi sistemi; problem ile ilgili en geniş ve kapsamlı bilgi içeren sistemdir.

Bilgi Sistemi

$$IS = (U, A) \quad (1)$$

U: Boş olmayan sonlu nesne kümesi ($\{x_1, x_2, \dots, x_n\}$) A: Boş olmayan sonlu nitelik kümesi ($\{a_1, a_2, \dots, a_n\}$) olarak tanımlanır. Her $a \in A$ olmak üzere, V_a gibi bir değer kümesine ve $f_a: U \rightarrow V_a$ olmak üzere de bir bilgi fonksiyonu sahiptir.

4.2.2. Ayırt Edilemezlik

Bir karar çizelgesi (tablosu), oluşturulmuş olan model hakkında var olan tüm bilgileri içermektedir. Buna bağlı olarak bu çizelge gereksiz yere çok büyümüş olabilir. Aynı nesne ya da ayırt edilemeyen nesnelere birden fazla tekrarlanmış olabilir. Ayrıca bazı nitelikler de gereksiz veya fazladan olabilir (Aydoğan ve Gencer, t.y.).

$S = (U, A)$ bir bilgi sistemi ise, herhangi bir $B \subseteq A$ için, B özelliklerinin her alt kümesi ayırt edilemezlik bağıntısı adı verilen U evreninin bir denklik bağıntısını tanımlar. (x, y) , U'dan nesne çiftleri olmak üzere, aşağıda tanımlanan bir $IND_S(B)$ denklik bağıntısına B-ayirt edilemezlik bağıntısı denir (Aybar, 2011).

$$IND_S(B) = \{(x, y) \in U^2 \mid \forall a \in B, a(x) = a(y)\} \quad (2)$$

Ayirt edilemezlik bağıntısındaki S alt indisi, hangi bilgi sisteminin kastedildiği açıksa genellikle ihmal edilir. Eğer $(x, y) \in IND_S(B)$ ise, x ve y nesnelere B'ye göre ayırt edilemezdirler.

4.2.3. Karar Sınıfları

d karar özelliği U nesnelere evreninin bir bölüntüsünü tüme varır. Genellemeyi bozmadan V_d 'nin $\{1, \dots, r(d)\}$ tamsayılar kümesi olduğu kabul edilebilir. Burada r(d), d'nin "rank"ıdır. Dolayısıyla tüme varılan bölüntüler; $\{X_1, \dots, X_r(d)\}$ "karar sınıfları" denilen denklik sınıflarının bütünüdür. Burada iki nesne karar özelliği için aynı değere sahipse, aynı karar sınıfına ait oldukları söylenir.

$$X_i = \{x \in U \mid d(x) = i\} \quad (5)$$

Bir S karar sistemi ve $B \subseteq A$ ise, S'nin B-genelleştirilmiş kararı adı verilen $\delta_B: U \rightarrow P(V_d)$ fonksiyonu şöyle tanımlanır:

$$\delta_B(x) = \{v \in V_d : \exists X' \in U(X' \cap IND_S(B)x \text{ ve } d(x) = v)\} \quad (6)$$

δ_S S-genelleştirilmiş kararına S'nin genelleştirilmiş kararı denir.

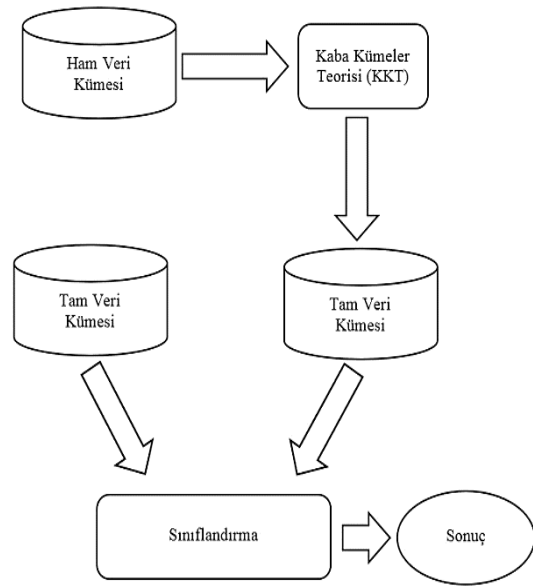
Bir S karar sistemi, ancak herhangi bir $x \in U$ için $|\delta_S(x)| = 1$ ise tutarlıdır, aksi takdirde tutarlı değildir. Buradan bir A karar tablosu ancak ve ancak $POSA(d) = U$ ise tutarlı olduğu kolaylıkla görülebilir (Skowron ve ark., 2002).

5. ÖNERİLEN MODEL

Sınıflandırma yöntemleri daha çok tam veri kümeleri üzerinde çalışırlar. Başka bir deyişle eksik veri kümesi üzerinde sınıflandırma metodlarının çalışma performansları düşük olurlar. Bunun nedeni, veri kümesinde eksik ve belirsiz verilerin olmasıdır. Daha önce de bahsedildiği gibi kaba kümeler ve bulanık

kümeler gibi araçlar, eksik, belirsiz ve tutarsız veri kümelerini tam veri kümelerine dönüştürürler. Bu çalışmada öngörülen ve önerilen model olarak, eksik veri kümesinin sınıflandırma yöntemine tabi tutulmadan önce bu araçlarla tam veri kümesine dönüştürülmesidir. Yani veri kümesi kural çıkarım yapan kaba kümeler ve bulanık kümeler araçlarına tabi tutulduktan sonra sınıflandırma yöntemlerine tabi tutulması önerilir.

Aşağıda önerilen modelin şeması verilmiştir.



Şekil 1. Önerilen Model

Bir nesnenin hangi sınıfa ait olduğunu belirlemek, niteliğin ayırt ediciliğine bağlıdır. Sınıf belirleme işleminde bir niteliğin değer kümesindeki bir değer, birden fazla sınıf için ortak olabilir. Örneğin, ST segment değeri, MI ve MIS sınıfları için ortak bir değerdir. Bu ortak değerler karmaşıklık yaratmakta ve ayırt ediciliği azaltmaktadır. Dolayısıyla bahsi geçen problemler sınıf belirlemeyi zorlaştırmaktadır. Bu problemlere KKT ile çözüm sunulmuştur. Gereksiz, tekrarlı ve eksik bilgi içeren ham veri kümesi KKT için bilgi sistemidir. KKT ayırt edilebilirlik matrisi ve ayırt edilebilirlik fonksiyonlar yardımı ile bu bilgi sistemini, özgünlüğünü bozmadan gereksiz ve tekrarlı bilgiden arındırarak tam veri kümesine dönüştürür. Aynı zamanda nesne ile nitelik arasındaki bağıntı karmaşıklığını sadeleştirir.

Ayırt Edilebilirlik Matrisi (Discernibility Matrix); $S = (U, A)$ U nesne, A nitelik kümesi olmak üzere ve n adet nesnesi olan bir bilgi sistemi olsun. S'nın ayırt edilebilirlik (discernibility) matrisi, M_S aşağıda verilen d_{ij} elemanları olan simetrik bir nxn matristir. Bu matrisin her d_{ij} elemanı x_i ve x_j nesnelerini farklı kılan özellikler kümesinden oluşmaktadır (Skowron ve ark., 2002).

$$d_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \quad i, j = 1, 2, \dots, n \quad (3)$$

Ayırt Edilebilirlik Fonksiyonları, mantıksal bir fonksiyondur. Ayırt edilebilirlik matrisindeki elemanlar kendi aralarında “VEYA (+)” işlemi, birbirleri arasında ise “VE (*)”işlenme tabidir. Ayırt edilebilirlik fonksiyonu indirgenerek indirgenmiş nitelik kümeleri bulunur. Bir bilgi sistemi için ayırt edilebilirlik fonksiyonu f_S, a_1, \dots, a_m özelliklerine karşılık gelen m adet a_1^*, \dots, a_m^* Boolean değişkeninin bir Boolean fonksiyonudur (Cios ve ark., 2000).

$$f_S(a_1^*, \dots, a_m^*) = \Delta\{v d_{ij}^* \mid 1 \leq j \leq i \leq n, d_{ij}^* \neq \emptyset\}$$

Burada $d_{ij}^* = \{a: a^* \in d_{ij}\} \quad (4)$

Ayırt edilebilirlik matrisi ve ayırt edilebilirlik fonksiyonları işlemleri için ROSE (Rough Set Data Explorer) programından faydalanılmıştır. ROSE programı KKT'nin temel işlevlerini yerine getiren ve kural çıkarımı yapan bir modül yazılımıdır (Predki ve ark., 1998).

Sınıflandırma kısmında WEKA adlı program kullanılmıştır. WEKA bugün birçok insan tarafından kullanılan Veri Madenciliği uygulaması geliştirme programıdır. Bu program Java platformu üzerinde geliştirilmiş açık kaynak kodlu bir programdır (Dener ve ark., 2009). WEKA'da çeşitli sınıflandırma yöntemleri mevcuttur. Önerilen modelde yüksek doğruluk oranına sahip sınıflandırma yöntemleri seçilmiştir.

Bundan sonra makalede önerilen model yerine model terimi kullanılacaktır.

6. UYGULAMA SONUÇLARI

Bu çalışmada Eskişehir Osmangazi Üniversitesi (ESOGÜ) Kardiyoloji Anabilim dalı doktorlarının katkılarıyla elde edilen hastaların verileri kullanılmıştır. Toplamda 689 hastanın gerçek EKG sinyal kayıtları alınmıştır. Bu 689

hastadan 145'i Sol Ventricular Büyümesi, 95'i Atrial Büyümesi, 211'i Myocardial Ischemia, 118'i Myocardial Infarction ve 120'i de Normal Sağlıklı bireylerinden oluşmaktadır. Alınan veri kümesi, 1 karar niteliği ve 14 adet koşul niteliğinden oluşmaktadır. Karar niteliği değer kümesi 5 farklı değer almaktadır. Bu değer aşağıdaki gibi sınıflandırılmıştır.

- Miyokardiyal Enfarktüs (MI) = 1
- Miyokardiyal İskemi (MIS) = 2
- Atrial Büyümesi (AB) = 3
- Sol Ventricular Büyümesi (SVB) = 4
- Sağlıklı Bireyler (N) = 5

Koşul nitelikleri ise hastalıklar üzerinde etkili olan EKG belirtilerini ifade ederler. Ayrıca bu niteliklerin her biri farklı değer kümesine sahiptirler. Örneğin, kalp atış ritmini gösteren *Rate* koşul niteliği için,

Koşul	Aldığı Değer Kümesi
< 60 kalp atış / dk	Bradycardia
60-100 kalp atış / dk	Normal
100 <	Tachycardia

Yukarıdaki koşullara bağlı olarak *Rate* niteliğinin değer kümesi; Bradycardia, Normal ve Tachycardia olmak üzere 3 farklı değerden oluşmuştur.

Aşağıda Tablo 3'de veri kümesinin tam (ham) hali ile sınıflandırma sonuçları gösterilmiştir. Başka bir deyişle model olmadan sınıflandırma sonuçları verilmiştir.

Tablo 3. Model Olmadan Sınıflandırma Sonuçları

Algoritma	Nitelik Sayısı	Eğitim Kümesi (%)	Doğruluk (Accuracy) (%)	Zaman (Running-Time) (sn)	Mutlak Ortalama Hata
NaiveBayes	12	60	82.2464	0.012	0.0957
BayesNET	12	60	84.0580	0.020	0.0900
IB1	12	60	76.0870	0.010	0.0957
IBK	12	60	83.3333	0.001	0.0807
KStar	12	60	83.3333	0.001	0.1146
LWL	12	60	60.1449	0.001	0.1989
JRip	12	60	83.3333	0.110	0.1111
NBTree	12	60	84.0580	1.500	0.0925
VFI	12	60	67.3913	0.010	0.3154

Tablo 3 ve Tablo 4 incelendiğinde sınıflandırma algoritmalarından JRip hariç, diğerlerinin tümünde sınıflandırma doğruluklarında (accuracy) değişme olmaz iken, JRip algoritması da dâhil tüm algoritmaların performanslarında (çalıştırma zamanı bakımında) iyileşmeler olduğu görülmektedir. JRip, Ridor, J48 vb. algoritmalar kural tabanlı çalışan algoritmalarlardır. Genel çalışma mantıkları;

1. Boş kural kümesi ile başla.
2. Learn-One-Rule fonksiyonunu kullanarak kural üret.
3. Kurallar yardımıyla belirsiz eğitim kayıtlarını sil.
4. Kriterler yeterli olana kadar 2. ve 3. adımları tekrarla.

Jrip algoritması kural tabanlı çalıştığında nesne ile nitelik arasındaki bağıntı karmaşıklığı doğruluğunu etkilemektedir. Model ile bu karmaşıklık azaltıldığında JRip algoritmasının doğruluğunda iyileşme olmuştur.

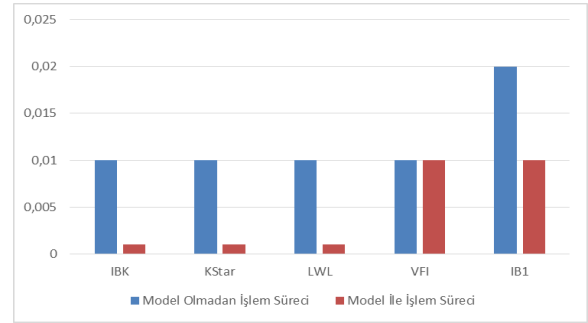
Aşağıda Tablo 5'te model ile model olmadan, sınıflandırma metodlarının doğruluk ve çalıştırma zamanlarındaki değişimler gösterilmiştir.

Tablo 5. Model Olmadan Ve Model İçin Doğruluk/ Zaman Değişimi

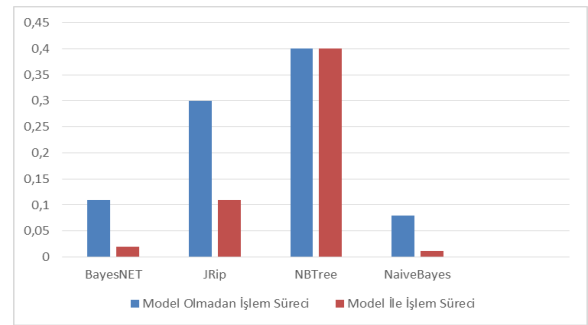
Algoritma	Model Olmadan Doğruluk (%)	Model ile Doğruluk (%)	Model Olmadan Zaman (sn)	Model ile Zaman (sn)	Eğitim Kümesi (%)
NaiveBayes	82.2464	82.2464	0.080	0.012	60
BayesNET	84.058	84.058	0.110	0.020	60
IB1	76.087	76.087	0.020	0.010	60
IBK	83.3333	83.3333	0.010	0.001	60
KStar	83.3333	83.3333	0.010	0.001	60
LWL	60.1449	60.1449	0.010	0.001	60
JRip	81.1594	83.3333	0.300	0.110	60
NBTree	84.058	84.058	1.500	1.500	60
VFI	67.3913	67.3913	0.001	0.001	60

Tablo 5'e bakıldığında JRip algoritması dışında algoritmaların doğruluklarında değişme olmadığı, ancak JRip algoritmasının doğruluğunda da bir iyileşme olduğu görülüyor. Aynı zamanda hemen hemen tüm algoritmaların işlem sürecinde kayda değer bir iyileşmenin olduğu anlaşılıyor. Örneğin, NaiveBayes

sınıflandırma için, veri kümesi tam veri kümesine dönüştürülmeden NaiveBayes sınıflandırma uygulandığındaki işlem süreci 0.80 sn iken önerilen model uygulandığında NaiveBayes sınıflandırma için işlem süreci 0.20 sn'ye düşmektedir. Bu da NaiveBayes sınıflandırma için işlem süreci yönünde %75'lik bir iyileşmeye denk gelmektedir. Bu iyileşmenin nedeni, nesnelere (hastalara) ait koşul nitelikleri ile karar niteliği (hastalıklar) arasındaki bağıntının karmaşıklığı, belli bir belirsizlik yaratmıştır. KKT bu karmaşıklığı, belirsizliği belli bir ölçüde azaltılmıştır. Bu da sınıflandırma doğrulukları sabit iken, performansta önemli bir iyileşmeyi sağlanmıştır. İyileşme sonuçları Şekil 1 üzerinde gösterilmiştir. Şekil 1 (a) ve (b) olarak iki parçaya bölünmüştür. Çünkü tek parça üzerinde yürütme zaman değeri çok küçük olan algoritmadaki iyileşmeleri gözlemlemek zorlaşır. Çözüm olarak Şekil 1 iki ayrı parçaya bölünerek değerlerin açık olarak gözlemlenmesi sağlanmıştır.



Şekil 2. Zaman Bazında İyileşme Sonuçları (a)



(b)

Bu çalışmada diğer önemli bir konu, herhangi bir insana ait EKG'sinin %85'e yakın bir doğruluk oranı ile doğru sınıflandırılma işlemidir. Aşağıda ROSE programı ortamında kaba kümeler yardımı ile elde edilmiş bazı kural çıkarımları verilmiştir. Bu kurallar, koşul nitelikleri ile karar niteliği arasındaki bağlantıya bağlı olarak nesnenin sınıfını belirten kurallardır.

