Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2016 - Cilt: 17 Sayı: 1
Sayfa: 159 - 166
DOI: 10.18038/btda.59175
Geliş: 10 Şubat 2016 Düzeltme: 01 Mart 2016 Kabul: 13 Mart 2016

OLASILIKSAL KABA KÜMELER TEORİSİ YAKLAŞIMI İLE EKG VERİLERİNİN SINIFLANDIRILMASI

Burak MAĞDEN[1], Sedat TELÇEKEN[1, *]

[1]Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, İkieylül Kampüsü, Anadolu Üniversitesi, 26555, Eskişehir

ÖZET

Sınıflandırma, bilgisayar mühendisliğinde bir veri kümesinin, uzmanlar yerine bilgisayarlar tarafından, özellikleri aracılığı ile gruplanması işlemidir. Kaba kümeler teorisi son yıllarda sınıflandırma problemlerinde gerekli karar kurallarının belirlenmesinde etkili bir araç olarak kullanılmaktadır. Bu çalışmada Eskişehir Osmangazi Üniversitesi Tıp Fakültesi'ndeki kalp hastalarının EKG verileri Kaba Kümeler Teorisi (KKT) ve Olasılıksal Kaba Kümeler Teorisi (OKKT) yaklaşımı ile sınıflandırılmıştır. Sınıflandırmada elde edilen sonuçlar değerlendirilirken, doğruluk ve genellik kriterlerinden yararlanılmıştır. OKKT'de bu kriterler, bir nesnenin karar kümesi içerisinde olup olmadığının koşullu olasılığını temsil eden (α, β) eşik değerleriyle yakından ilişkili olduğu görülmüş ve ortalamaya bağlı OKKT'nin genellik değerinde %49 oranında iyileşme elde edilmiştir.

Anahtar Kelimeler: Kaba kümeler teorisi, Olasılıksal kaba kümeler teorisi, EKG, Sınıflandırma, Uzman sistemler, Veri madenciliği

CLASSIFICATION OF ECG SIGNALS BY APPROACHING PROBABILISTIC ROUGH SETS THEORY

ABSTRACT

In this study, surface pressure distributions on the two low-rise building models having different Classification in computer engineering is a process of grouping a data set with respect to its properties by computers rather than experts. Rough sets theory has been used as an effective tool for determining decision rules in classification problems in the recent years. In this study, ECG data of the cardiacs in Medical School at Eskisehir Osmangazi University is classified using rough sets theory (RST) and probabilistic rough sets theory (PRST). In order to evaluate the outcomes in the classification, accuracy and generality are used as evaluation criteria. In PRST, it is observed that these criteria are closely related with the threshold (α, β), which represents the conditional probability of an object being a member of a decision cluster or not. Depending on the average, 49% of improvement is observed in generality value of PRST.

Keywords: Rough sets theory, Probabilistic rough sets theory, ECG, Classification, Expert systems, Data mining

1. GİRİŞ

Günümüzde bilişim çağının etkisi sonucunda çok hızlı bir biçimde büyük veri yığınları ortaya çıkmaktadır. Bu verilerin anlamlı olabilmesi için, gereksiz, tekrarlı ve ilişkisiz bir takım verilerin ayıklanması gereklidir. Ortaya çıkan temiz verinin ise, farklı amaçlar doğrultusunda sınıflandırması verinin takibi ve anlamlı hale gelmesine yardım etmektedir. Eldeki verilen sınıflandırması farklı branşlarda bulunan birçok bilim dalı için önemli bir yer tutmaktadır. Sınıflandırma işlemi için; karar ağaçları, destek vektör makinesi, k-yakın komşuluklar, doğrusal ve mantıksal regresyon, Naïve Bayes, C4.5, doğrusal sınıflandırıcı ve kaba kümeler gibi pek çok sınıflandırma metodu kullanılmaktadır [1].

Karar ağaçlarında yapraklar sınıfları belirtirken diğer düğümler testleri, dallar ise bu testlerin sonuçlarını belirtmektedir. Özellikler belirlendikten sonra çeşitli yöntemlerle nesnelerin birbirinden ayrılması için uygun özellikler tespit edilir. Sonrasında, doğruluğu daha yüksek bir sınıflandırma elde etmek için oluşturulan ağaçtan gereksiz dalların kaldırıldığı bir budama işlemi uygulanır. C4.5 ise bellek kullanımı

* Sorumlu Yazar: E-posta: stelceken@anadolu.edu.tr

ve sınıflandırma hızı açısından kazanç sağlamak için daha küçük bir ağaç oluşturmayı amaçlayan özelleşmiş bir karar ağacı yapısıdır [2].

K-yakın komşuluklar yöntemi, bütün özellikleri bir boyut olarak ve bütün nesneleri de, "n" özelliklerin sayısı olmak üzere; "n" boyutlu bir uzaydaki noktalar olarak kabul eder. Yeni bir nesnenin eksik bir özelliğini belirlemek için bu nesnenin diğer bütün nesnelerle olan uzaklığını hesaplanarak bunların içerisinden Öklid metriğine göre en yakın k tanesinin oluşturduğu küme seçilir. Son olarak da nesnenin eksik özelliğini, nesneyi bu kümeye dâhil edecek şekilde belirlemektedir [3].

Doğrusal ve mantıksal regresyon sınıflandırma yöntemi, bir bakıma tahmin yöntemi olarak da düşünülebilir. Özetle bu yöntemin amacı, bir nesnenin eksik bir özelliğini, veri uzayında diğer özellikleri bu nesne ile aynı olan nesnelerin ilgili özelliğine bakarak tahminde bulunmaya yöneliktir [2].

Destek vektör makinesi yöntemi sınıfları birbirinden ayıran hiperdüzlemler bulmaya çalışır. Bulunan uygun hiperdüzlemlerin kendisine en yakın objeye maksimum uzaklıkta olması beklenir [2].

Bayes sınıflandırma yöntemi ve türevleri nesnelerin belirli sınıflara üye olma olasılıklarını tahmin etmeye çalışmaktadır. Burada öncelikli olarak veri kümesinde bütün özelliklerin dağılımlarını hesaplanır. Yeni bir nesne geldiğinde ise bu nesnenin bütün özelliklerini incelenerek çalışma veri kümesindeki diğer nesnelerle karşılaştırılır. Bu işlemin ardından yeni nesne için her bir sınıfa üye olma olasılığını hesaplanmaktadır [4].

Kaba kümeler teorisi (KKT), özellikle büyük, gereksiz, kesin olmayan, tutarsız veya eksik bilgi içeren veri kümeleri üzerinde sınıflandırma yapma için kullanışlı bir yöntem olarak 1980'lerin başlarında bilim dünyasına tanıtıldı. Çoğunlukla yardımcı bir işlem olarak kullanılmasına rağmen; kümelerde boyut indirgeme gibi özellikleri ile birlikte KKT, tek olarak sınıflandırma yapabilecek kabiliyetleri olan bir yöntemdir. KKT'yi anlayabilmek için öncelikli olarak ayırt edilemezlik tanımının bilinmesi gerekmektedir [5]. Ayırt edilemez küme, bütün özellikleri ve sınıfları aynı olan nesnelerden oluşan kümelerdir. Başka bir deyişle ayırt edilemez kümeler içindeki nesneler, haklarında daha fazla veri olmadan sınıflandırılamayan nesnelerdir. Ayırt edilemezlik, sınıflandırma için gerekli olmayan özellikleri ortaya çıkaran önemli bir kavramdır [6]. KKT bu sayede verilerdeki bazı eksiklerin giderilmesini sağlar. Ayrıca KKT, ayırt edilemez kümelerden karar kuralları üreterek sınıflandırma yapmaktadır [7]. Ayırt edilemez kümelerde kullanılan boyut indirgeme işlemine rağmen veri kümesinde eksik veriler bulunabilir. Bu eksik veriler, uygulama alanında sorunlara yol açabilmektedir ve bu sorunları gidermek için ortaya konulan çözümlerden birisi ise Olasılıksal Kaba Küme Teorisidir (OKKT) [6,8]. OKKT'deki temel farklılık, bilinen KKT'deki kabaca tanımlı bölge tanımının, (α, β) eşik değerleri kullanılarak daha duyarlı bir biçimde kategorize etmeye çalışmak olarak özetlenebilir. Bu eşik değerlerinin belirlenmesi ise OKKT'nin önemli noktalarından biridir.

OKKT'de kabul, ret ve bekleme bölgeleri olarak adlandırılan pozitif, negatif ve sınır bölgeleri ile tanımlanan üçlü-karar durumu elde edilmektedir [9]. Etkili bir OKKT uygulamalarının önündeki en büyük engel, bu bölgelerin α ve β değerleri ile sınıflandırmanın doğruluğunu ve genelliğini dengeleyecek şekilde belirlenmesidir. Yapılan çalışmada farklı (α, β) eşik değerleri denenerek seçilen veri kümesi için uygun olan (α, β) eşik değerleri elde edilmeye çalışılmıştır.

Bölüm 2'de, KKT ve OKKT'nin genel yapısı üzerinde durulmuş, Bölüm 3'te sınıflandırma işlemlerinin değerlendirilme kriterleri, doğruluk ile genellik arasındaki ilişki tanıtılmıştır. Bölüm 4'te farklı sınıflandırma yöntemleri ile KKT ve OKKT ile yapılan deneysel çalışmalar verilerek Bölüm 5'te sonuçlar ortaya konulmuştur.

2. KABA KÜMELER TEORİSİ VE OLASILIKSAL KABA KÜMELER TEORİSİ'NİN YAPISI

Bu bölümde kullanılan OKKT ve KKT'nin daha iyi anlaşılması için temel bilgileri özetlemek gereklidir. Bilinen KKT'de sınıflandırılan bir kümenin alttan ve üstten sınırlı yaklaşım kümelerin denklemleri (1) ve (2)'de gösterilmiştir [6,10-12].

$$\underline{yakınsak}(C) = \{x \in U | [x] \subseteq C\} \quad (1)$$

$$\overline{yakınsak}(C) = \{x \in U | [x] \cap C \neq \emptyset\} \quad (2)$$

Bu denklemlerde U evrensel kümeyi, $[x]$ ise denklik bağıntısı $E \subseteq U \times U$ olacak şekilde x'in denklik sınıfını temsil eder. C ise U'nun işlem sonucunda istenilen alt kümesidir. Denklem (1) ve (2), C altkümesi ile $[x]$ denklik sınıfının arasındaki ilişkiyi gösterir. Denklem (1), tamamı ile C tarafından kapsanan denklik sınıflarını; denklem (2) ise en az bir elemanı C ile ortak olan denklik sınıflarını ifade eder. KKT'den farklı olarak OKKT, (4) ve (5) denkleminde verildiği üzere denklik sınıfı ile C arasındaki kesişim yüzdesi ile ilgilidir [6, 9].

$$\left(P(C|[x])\right) = \frac{|C \cap [x]|}{|[x]|} \quad (3)$$

$$\underline{yakınsak}_{(\propto,\beta)}(C) = \{x \in U | P(C|[x]) \geq \alpha\} \quad (4)$$

$$\overline{yakınsak}_{(\propto,\beta)}(C) = \{x \in U | P(C|[x]) \geq \beta\} \quad (5)$$

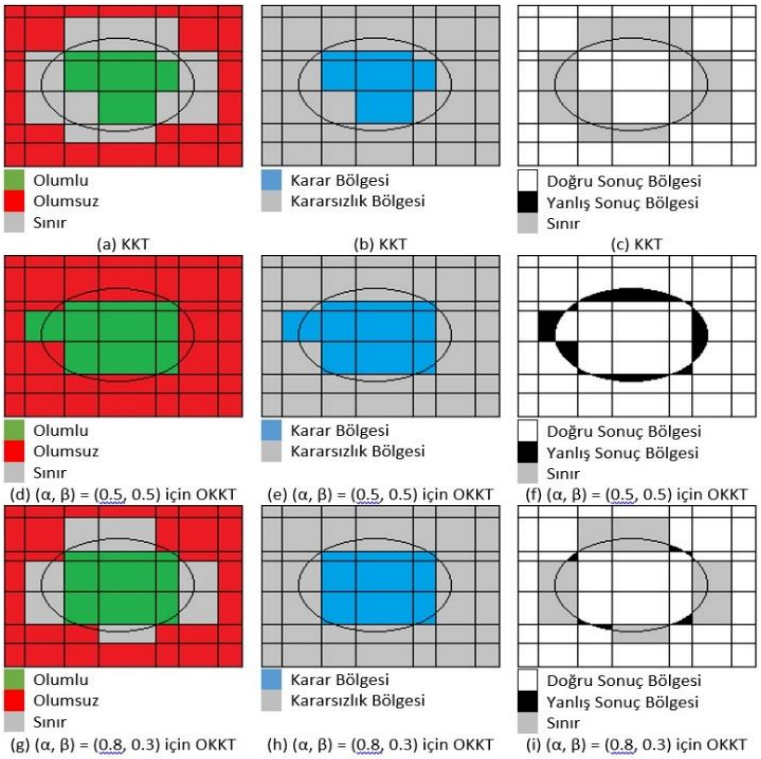
Bu denklemlerde $P(C|[x])$, x'in koşullu olasılığı olarak adlandırılır ve x'in denklik sınıfındaki nesnelerinin aynı zamanda C kümesinin elemanı olmasının olasılığını temsil eder. α ve β değerleri $[0, 1]$ aralığında değerler alabilir [13]. Burada OKKT'de, eşik değerleri (6) ve (7) nolu denklemlerde $\alpha = 1$ ve $\beta = 0$ olarak belirlendiğinde elde edilen alt ve üst yaklaşım kümeleri KKT yaklaşımına dönüşür.

$$\underline{yakınsak}_{(\propto,\beta)}(C) = \{x \in U | P(C|[x]) \geq \alpha\} \quad (6)$$

$$\overline{yakınsak}_{(\propto,\beta)}(C) = \{x \in U | P(C|[x]) \geq \beta\} \quad (7)$$

Dolayısı ile KKT, OKKT'nin eşik değerleri $(\alpha, \beta) = (1, 0)$ olacak şekilde bir alt tanımı olarak kabul edilebilir. KKT'nin bölgeleri x'in C kümesinde olup olmamasına göre net bir şekilde belirlendiği (6) ve (7) denklemlerinden görülebilir. Diğer bir yandan OKKT bu bölgelerin sınırlarını α ve β eşik değerlerini kullanarak esnetir. OKKT'de bir x nesnesi için; eğer x'in denklik sınıfı ile bu denklik sınıfının C kümesi ile kesişme oranı 1 yerine α'dan büyük ise x nesnesi istenilen C kümesinin içinde kabul edilir. Benzer şekilde bu oran β'dan küçük ise x nesnesi istenilen C kümesinin dışında kabul edilir. Sınıflandırma bu bölgeler üzerinden yapılmaktadır. Başka bir deyişle x nesnesinin C kümesinin/sınıfının içinde olup olmadığı $\beta \leq P(C|[x]) \leq \alpha$ önermesinin doğru olup olmaması ile belirlenmektedir.

OKKT'de belirleyici konu kabul edilebilir oranda doğruluk kaybı ile etkinliği arttırmak için bu eşik değerlerinin belirlenmesidir. Bir nesnenin hangi sınıfta olduğu bu değerlere göre belirlendiği için bu işlem büyük önem taşımaktadır.



Şekil 1. Farklı eşik değerleri için doğruluk ve genellik bölgeleri

3. DEĞERLENDİRME KRİTERLERİ

Şekil (1a,b,c)'de eş olmayan alanlarla oluşturulmuş dikdörtgensel bölgeye çizilen bir elipsin iç ve dış bölgesinin KKT yöntemiyle belirlenmesi gösterilmektedir. Burada yeşil ile belirlenen bölge elipsin kesinlikle içerisinde kalan taralı alan, kırmızı bölge ise elipsin kesinlikle dışında kalan bölgedir. Gri ile belirlenmiş bölge için ise elipsin sınır bölgelerini belirleyen alan denilebilir. OKKT ise, sınır alanlarının seçilen eşik değerlerine göre elipsin iç veya dış bölgesine dâhil edilmesini belirlemeye çalışmaktır. İkinci üç şekilde ise (Şekil 1d,e,f) $(\alpha, \beta) = (0.5, 0.5)$ eşik değerleri için OKKT tarafından elde edilen iç ve dış bölgeler gösterilmektedir. Şekil 1d'de görüldüğü gibi bütün denklik kümeleri için bir sınıf elde edilmiş olmasına karşın, Şekil 1(f) de halen yanlış sınıflandırılmış büyük bir alan vardır. Son üç şekilde ise (Şekil 1g,h,i) $\beta < 0.5 < \alpha$ değerleri için OKKT el edilen iç ve dış bölgeler resmedilmiştir. Şekil 1i'de son durumda yanlış sonuç bölgesinin minime edilmiş olduğu görülmektedir. Buradan hareketle seçilen eşik değerinin doğruluğuna göre sınır bölgesinde kalan ihtilaflı bölgelerin elipsin içine ya da dışına dahil edilmesinde OKKT'nin üreteceği sonuçlar, KKT'ye göre daha doğru olacaktır.

Sınıflandırma işleminin performansını değerlendirirken iki kriterden yararlanılmıştır. Bu kriterlerden ilki yapılan sınıflandırmanın gerçeği ne kadar yansıttığını temsil eden ve yayınlanmış çoğu araştırmanın odaklandığı doğruluk değeridir. Doğruluğu hesaplamak için çeşitli yöntemler öne sürülmüştür [14]. Bu çalışmada doğruluk; doğru sınıflandırılmış nesnelerin sayısının, sınıflandırılmış tüm nesnelerin sayısına oranı olarak ele alınmıştır. Klasik KKT bir uzman tarafından belirlenmiş, çalışılan veri kümesindeki bütün özellikleri ve sınıfları aynı olan nesnelere benzer nesneler için karar üretebilmektedir. Bu KKT'nin maksimum doğruluğa sahip olduğu anlamına gelmektedir (Şekil 1c). Doğruluk kriteri denklem (8) de olduğu gibi ifade edilir.

$$D(\alpha, \beta) = \frac{\left|\underline{yakınsak}_{(\alpha,\beta)}(C) \cap C\right|}{\left|\underline{yakınsak}_{(\alpha,\beta)}(C)\right|} \qquad (8)$$

Burada $D(\alpha,\beta)$ değeri α ve β eşik değerleri için doğruluğu temsil etmektedir. İkinci kriter ise sınıflandırmanın ne kadar kullanılabilir olduğunu temsil eden genelliktir [15]. Nesneler uzayındaki elemanların yüzde kaçının sınıflandırılabildiğini ifade eden bir kriter olarak kabul edilebilir. Bu çalışmada sınıflandırma işlemi sonrası sınıflandırılmış nesnelerin sayısı, nesne uzayının eleman sayısına bölünmesiyle hesaplanmaktadır. Doğruluğun kriterinin aksine; KKT, doğruluğunun maksimum olmasını sağlayan aynı sebeple olabilecek en küçük genellik değerine sahiptir (Şekil 1(b)). Genellik denklem (9)'da verilen denklem ile hesaplanabilir.

$$G(\alpha,\beta)=\frac{\left|\underline{yakınsak}_{(\alpha,\beta)}(C)\right|}{|U|} \qquad (9)$$

Burada α değeri 0.5 olarak belirlendiğinde genellik artarken (Şekil 1(e)), doğruluk değerinde düşme olacaktır (Şekil 1(f)). α değeri, doğruluk değeriyle doğru, genellik değeriyle ise ters orantılı olarak değişmektedir.

4. DENEYSEL SONUÇLAR

Çalışmada kullanılan veri kümesi Eskişehir Osmangazi Üniversitesi Tıp Fakültesi'nin Kardiyoloji Bölümü'nden toplanmıştır. Veri kümesi 145 sol ventrikül genişlemesi, 95 atriyal genişlemesi, 211 miyokardiyal iskemi, 118 miyokard enfarktüsü ve 120 sağlıklı olmak üzere 689 hastanın bilgisini içermektedir. Hastalar 14 belirtiye sahiptir. Veri kümesi doktorlar tarafından doldurulmuş teşhis sütunu dâhil olmak üzere 15 sütun ve 689 satırdan oluşmaktadır.

Öncelikle daha hızlı ve bellek kullanımı daha az olan bir sınıflandırma işlemi gerçekleştirmek için boyut indirgeme işlemi uygulanmıştır. Boyut indirgeme işlemi içerisinde hasta bilgileri teşhis sütunlarına göre gruplanmış ve bu grupların birbirlerinden ayrılması için kullanılabilecek belirti sütunları tespit edilmiştir. Çıkan sonuçta 3 belirtinin teşhis yapılırken ele alınan hastalıklar için ayırıcı olmadığı belirlenmiştir. Belirlenen 3 sütun doktorlarla değerlendirilip onay alınarak sınıflandırma işlemi için etkisiz olmaları sebebiyle veri kümesinden çıkartılmıştır. Bu işlemin ardından veri kümesi; 15 sütundan 12 sütuna indirgenerek 689x12 boyutunda bir matris halini almıştır.

Yapılan deneysel çalışmalar için, veri kümesine sırasıyla KKT ve OKKT uygulanmıştır. KKT'den elde edilen sonuçlar OKKT'nin eşik değerlerini hesaplamada kullanılmıştır. Veri kümesindeki nesneler ikiden fazla sınıfa üye olabildiği için β değeri önem arz etmemektedir; çünkü denklik sınıfının, C sınıfına göre koşullu olasılığı β'nın altında olan nesneler için C sınıfına ait olmadıkları haricinde bir karar üretilemez. Karar sütunu iki değer alabilen veri kümeleri için bu durumda incelenen nesnenin diğer sınıfa ait olduğu yorumu yapılabilir. Elde edilen veri setine KKT tabanlı sınıflandırma işlemi uygulandığında genellik değeri 0,6372 olarak bulunmuştur. Bu değerin düşüklüğü çalışmayı derinleştirme ihtiyacı doğurmuştur.

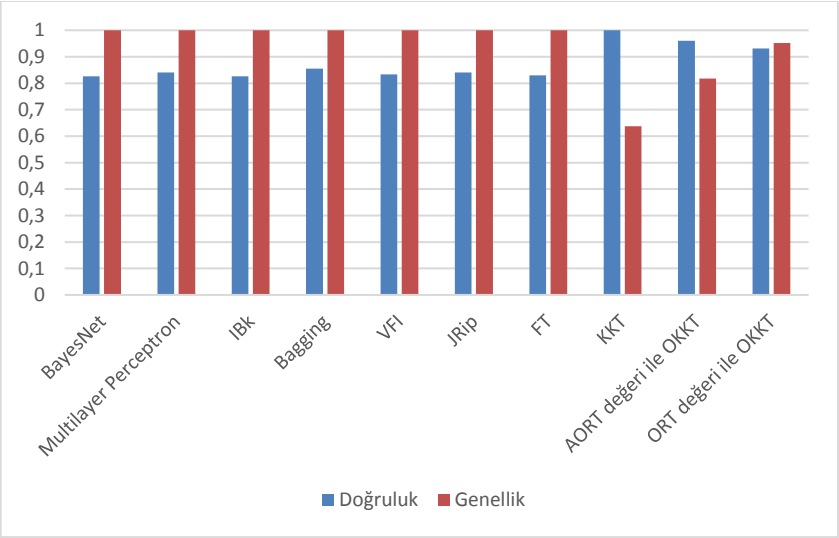
Bunun üzerine eşik değerlerinin hassas bir şekilde elde edilmesine yönelik Ortalama (ORT) ve Ağırlıklı Ortalama (AORT) değerleri ile OKKT yöntemi uygulamaya karar verilmiştir. ORT değerini hesaplamak için denklem (10)'da gösterildiği gibi her denklik sınıfındaki en yüksek değere sahip sınıfa ait nesnelerin, o denklik sınıfındaki yüzdesi (DSY) toplanılmış ve toplam denklik sınıfı sayısına (DSS) bölünmüştür. Diğer bir yandan AORT değeri için ise; her denklik sınıfının en yüksek değere sahip sınıfa ait nesnelerinin DSY ile denklik sınıfının toplam eleman sayısı (DES) çarpılmış ve bulunan değerler toplanmıştır. Bu toplamı ise veri kümesindeki toplam eleman sayısına bölünmüştür.

$$ORT=\frac{\sum DSY}{DSS} \qquad (10)$$

$$AORT=\frac{\sum DSY.DES}{|U|} \qquad (11)$$

Belirlenen ORT ve AORT değerlerinin ardından Algoritma (1)'de daha önce önerilen pseudo koda göre OKKT kullanılarak doğruluk ve genellik değerleri elde edilmiştir. Geliştirilen bu algoritma veri kümesine uygulandığında ORT değeri 0.7163 ve AORT değeri 0.7750 olarak bulunmuştur. Ayrıca bulunan sonuçlar diğer popüler sınıflandırma yöntemlerini içerisinde bulunduran WEKA veri madenciliği uygulama geliştirme ortamı kullanarak karşılaştırılmıştır [16]. WEKA ortamında kullanılan veri kümesi boyut indirgeme işleminden sonra elde edilen veri kümesidir. Tablo (1)'deki sonuçlar göstermiştir ki; ufak bir doğruluk kaybı ile ciddi miktarda genellik kazancı elde edilebilmektedir. Deneyler ayrıca uygulanan bütün KKT tabanlı yöntemlerin diğer yöntemlere göre tatmin edici doğruluk değerine sahip olduğunu göstermiştir.

Sınıflandırma yöntemlerinin genellik ve doğruluk arasındaki değişim Şekil (2)'de gösterilmiştir. Her bir sınıflandırma yönteminden en yüksek doğruluk değerine sahip olanı seçilerek bütün KKT tabanlı yöntemlerle kıyaslanmıştır. Ağırlıklı ortalama kullanılarak hesaplanan OKKT yöntemi, KKT ile karşılaştırıldığında doğruluk değeri %4 azalırken, genellik değerinde %28 artış olmuştur. Ortalama kullanılarak hesaplanan OKKT yöntemi ise, KKT ile karşılaştırıldığında doğruluk değerinde %7 azalma olurken, genellik değerinde %49 artış olmuştur.



Şekil 2. EKG veri sınıflandırmasında kullanılan yöntemlerin önerilen yöntem ile doğruluk ve genellik değeri üzerinden karşılaştırılması

Algoritma 1 Pseudo kod	
1:	*Klasik KKT'yi uygula.*
2:	*Genellik (G) ve ortalamayı (ORT) hesapla.*
3:	*Sınırdaki her denklik sınıfı için en büyük yüzdeye sahip kararı al, bu karara sahip denklik sınıfındaki nesneleri listele.*
4:	*Ortalamayı hesapla ve normal ortalama (N_ORT) olarak kaydet.*
5:	*Denklem (9) daki gibi ağırlıklı ortalama değerini hesaplayara AORT olarak kaydet.*
6:	*OKKT'yi (α, β) = (ORT, 0) olarak uygula.*
7:	*Genellik ve Doğruluk kriterlerini (α, β) = (ORT, 0) için hesapla.*
8:	*OKKT'yi (α, β) = (AORT, 0) değeri için hesapla.*
9:	*Genellik ve Doğruluk kriterlerini (α, β) = (AORT, 0) için hesapla.*

Tablo 1. Deney sonuçları

Sınıflandırma Yöntemi		Doğruluk	Genellik
Bayes	BayesNET	0.8260	1
	NaiveBayes	0.8188	1
	NaiveBayesSimple	0.8188	1
Functions	Logistic	0.8224	1
	Multilayer Perceptron	0.8405	1
	SMO	0.8079	1
Lazy	IB1	0.7572	1
	IBk	0.8260	1
	KStar	0.8260	1
	LWL	0.5289	1
Meta	AdaBoostM1	0.3949	1
	Bagging	0.8550	1
	Decorate	0.8369	1
	LogitBoost	0.8224	1
	Vote	0.2862	1
Misc	HyperPipes	0.2028	1
	VFI	0.8333	1
Rules	JRip	0.8405	1
	OneR	0.4057	1
	PART	0.8079	1
	Ridor	0.7789	1
	ZeroR	0.2862	1
Trees	J48	0.8152	1
	SimpleCart	0.8188	1
	FT	0.8297	1
	NBTree	0.8188	1
RST	KKT	1	0.6372
	AORT değeri ile OKKT	0.9609	0.8175
	ORT değeri ile OKKT	0.9314	0.9521

5. SONUÇLAR

EKG verileri uzmanlar tarafından bilinen kalp rahatsızlıklarının teşhisinin konulmasında yardımcı bir bilgi olarak kullanılmaktadır. Yapılan çalışmada bu verilerin bazılarında bulunan eksik, gereksiz, bilgilerin OKKT yöntemi yardımıyla ortaya çıkartılmıştır. Çalışmada eldeki veriler 26 farklı sınıflandırma yöntemi ile KKT ve OKKT kullanılarak elde edilen sınıflandırmalar doğruluk ve genellik kriterleri üzerinden karşılaştırılmıştır. Karşılaştırılan yöntemlerden özellikle destek vektör makinesi, k-yakın komşuluklar ve doğrusal ve mantıksal regresyon yöntemleri EKG sinyallerinin sınıflandırılması için sıklıkla kullanılan yöntemlerdir.

KKT diğer tüm sınıflandırma yöntemlerine göre en doğru sınıflandırmayı yapmasına karşılık genellik kriteri ile kıyaslandığında dezavantajlı durumdadır. Buradaki problemi çözmek için OKKT kullanılmıştır. OKKT, doğruluk değerini dengeleyerek sınıflandırmanın genelliğini arttırmaya yönelik bir yöntemdir.

Deneylerde OKKT yönteminde kullanılması gerekli eşik değerlerinin belirlenmesine yönelik ortalama ve ağırlıklı ortalama değerleri hesaplanarak doğruluk ve genellik kriterlerinde en-iyileme gerçekleştirilmeye çalışılmıştır. Burada ortalama değerini α eşik değeri kabul eden OKKT sınıflandırma yöntemi genellik üzerinde KKT yöntemine göre %49 oranında iyileşme sağlamıştır.

Sonuç olarak üretilen sınıflandırmanın bir uzman sistem olarak, kardiyologların daha kolay teşhis koymalarında yardımcı bir araç olacağı öngörülmektedir. Sağlık alanında son karar daima bir uzman tarafından verildiği için bu sistemin yüksek doğruluk ve genellik değerleri ile uzmanların teşhis koyma sürelerini oldukça kısaltacak bir araç olacağı düşünülmektedir.

İleriki çalışmalarda farklı veri kümeleri üzerinden OKKT'nin sınıflandırmaya etkisi araştırılacak ve genellik ile doğruluk kriterlerini dengeli bir biçimde kullanmayı amaçlayan uyarlanabilir bir tespit sisteminin elde edilmesine çalışılacaktır.

KAYNAKLAR

[1] Entezari-Maleki R, Rezai A, Minaei-Bidgoli B. Comparison of classification methods based on the type of attributes and sample size. Journal of Convergence Information Technology 2009; 4.3: 94-102.

[2] Han J, Kamber M. Data mining: Consepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, San Francisco, CA, USA: Elseiver, 2006.

[3] Kantardzic M. Data mining: concepts, models, methods, and algorithms University of California, California, CA, USA: Wiley-Interscience 2003.

[4] Yang Y, Liu X. A re-examination of text categorization methods. In: ACM conference on Research and Development in Information Retrieval; 1999; Berkeley, CA, USA, pp. 42-49.

[5] Komorowski J, Pawlak Z, Polkowski L, Skowron A. Rough fuzzy hybridization: a new trend in decision-making. Rough sets: a tutorial 1999; Singapore, Springer-Verlag.

[6] Yao Y. Probabilistic rough set approximations. International Journal of Approximate Reasoning 2008; 49:2, 255-271.

[7] Çekik R, Telçeken S, EKG Sinyallerinin Kaba Kümeler Teorisi Kullanılarak Sınıflandırılması. Anadolu University Journal of Science and Technology A-Applied Science and Engineering 2014;, 15.2: 125-135.

[8] Ziarko W. Probabilistic approach to rough sets. International Journal of Approximate Reasoning 2008; 48:2, 272-284.

[9] Yao Y. Three-way decisions with probabilistic rough sets. Information Science 2010; 180:3, 341-353.

[10] Pawlak Z. Rough Sets. International Journal of Computer and Information Sciences 1982; 11: 341-356.

[11] Pawlak Z. Rough sets: Theoretical aspects of reasoning about data. Vol. 9. Springer Science & Business Media, 2012.

[12] Walczak B. Massart DL. Rough Sets theory. Chemenometrics and Intelligent Laboratory Systems 1999; 47: 1-16.

[13] Deng XF, Yao YY. A multifaceted analysis of probabilistic three-way decisions. Fundamenta Informaticae 2014; 132.3: 291-313.

[14] Hernandez-del-Olmo F, Gaudiso E. Evaluation of recommender system: a new approach. Expert Systems with Applications 2008; 35:3, 790-804.

[15] Herlocker JL, Konstan J.A, Terveen LG, Riedl JT. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 2004; 22:1, 5-53.

[16] Mark H, Eibe F, Geoffrey H, Bernhard P, Peter R, Ian HW (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.