



Düzce Üniversitesi Bilim ve Teknoloji Dergisi

Araştırma Makalesi

K-Ortalama Kümelerinin Sınıf Bilgisi Olarak Karar Ağacı Oluşturmada Kullanılması ve Glokom Çoklu Sınıflandırılmasında Başarıma Etkisi

Sait Can YÜCEBAŞ*, Ahmet Cumhuri KINACI

Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, TÜRKİYE

* Sorumlu yazarın e-posta adresi: can@comu.edu.tr

ÖZET

Bu çalışma çoklu sınıflandırmada performans artırımı için K-Ortalama ve Karar Ağacı yöntemlerinden oluşan bir model sunmaktadır. Model glukom veri kümesi üzerinde test edilmiş kesinlik ölçütü 0,808, ROC alanı 0,839 bulunmuştur

Anahtar Kelimeler: Veri Madenciliği, Sınıflama, Kümeleme, K-Ortalama, Karar Ağacı.

Usage Of K-Means Clusters as Class Labes In Decision Trees and Its Effect On Multiclassification Performance Of Glaucoma

ABSTRACT

In this study a model of K-Means - Decision Tree is presented to increase the multiclassification performance. This model is tested on glaucoma dataset, the accuracy and the are under ROC curve is calculated as 0.808, 0.839 respectively.

Keywords: Data Mining, Classification, Clustering, K-Means, Decision Tree

I. GİRİŞ

Çok sayıda yüksek boyutlu veri barındıran tıp alanında makina öğrenme yöntemleri yoğunlukla kullanılmaktadır [1]. Makina öğrenme ve veri madenciliği ile tıbbi teşhis [2], tedavi planlama, ilaç etkileşimi, hatırlatıcı-uyarıcı sistemler [3] geliştirilmiştir. Bahsedilen uygulama alanları içerisinde tıbbi teşhis koyma amaçlı sınıflama, literatürdeki çalışmaların büyük çoğunluğunu oluşturmaktadır [4]. Özellikle bu konudaki ilk çalışmalarda örneklemin hasta – hasta değil şeklinde ayrıldığı ikili sınıflama çalışmaları göze çarpsa da klinik uygulamalardaki kullanım örnekleri oldukça azdır [5]. Bunun en büyük nedenlerinden birisi, teşhise yönelik klinik uygulamalarda kişilerin birden çok sınıfta yer alabilmesidir. Bu durum tıp alanında kullanılan makina öğrenme yöntemleri için çoklu sınıflama kabiliyetinin önemini göstermektedir.

Çoklu sınıflama yöntemleri, ikili sınıflama çözümlerinden türetilmiştir. Bu yöntemlerin bazıları kullandıkları algoritmalarının doğası gereği direkt olarak çoklu sınıflamaya adapte edilebilirken, diğerleri ise kullandıkları formülasyonlarda bazı karmaşık değişiklikler yapılarak kullanılmaktadır [6]. K-En Yakın Komşu, Yapay Sinir Ağları ve C4.5 gibi yöntemler çoklu sınıflama çözümü için kolay adapte edilebilir yöntemler iken [7], AdaBoost ve Destek Vektör Makinası gibi yöntemlerin adaptasyonu ise daha zordur [6]. Buna ek olarak çoklu sınıflamada eğer örnek sayısı az ise, çoğunlukla her sınıfın kendi içine düşen örnek sayısı homojen olmamakta ve bu durumda bilinen sınıflama yöntemleri veriyi tam olarak öğrenmekte zorlanabilmekte ve sınıflama performansı beklenenin altında kalabilmektedir [8,9]. Bu gibi durumlarda sınıflandırma performansını artırmak adına kullanılan yöntemler iyileştirilebilir. Bu iyileştirme için çeşitli istatistik yöntemlerin [10] yanında, hibrit modeller de tercih edilebilir.

Makaleye konu olan bu çalışmada veri kümesindeki örnek sayısının az olduğu çoklu sınıflama probleminde sınıflandırma performansını artırmak için K-Ortalama ve Karar Ağacı yöntemlerinden oluşan bir model tasarlanmıştır. İlgili modelde K-Ortalama yönteminden elde edilen kümeler sınıf bilgisi olarak kullanılmış, veri örneklerinin elde edilen bu sınıflara atanması için Karar Ağacı kullanılmıştır. İlgili modelin başarımı Glokom veri kümesinde ölçülmüş ve elde edilen sonuçlar diğer bağımsız makina öğrenme yöntemleri ile karşılaştırılmıştır

II. YÖNTEM

Makalenin yöntem bölümü, ilgili çalışmada kullanılan veri kümesinin özelliklerinin ve önerilen modelin detaylarının anlatıldığı iki alt bölümden oluşmaktadır.

A. MATERYALLER

İlgili çalışmada kullanılan veri seti glokom hastalığı şüphesiyle incelenen 131 kişiden oluşmaktadır. Bu kişiler yapılan klinik kontrollerinde sağlıklı, hafif hasta ve orta hasta olarak sınıflandırılmıştır. Veri kümesindeki kişilerin bu sınıflara dağılımı Tablo-1’de verilmiştir.

Tablo 1. Veri kümesinde kişilerin sınıflara dağılımı

Sağlıklı Sınıf	Hafif Hasta Sınıfı	Orta Hasta Sınıfı
26	81	24

Söz konusu hastalık dereceleri ve sağlıklı olma durumu kategorisel olarak kodlanmıştır. İlgili hastalığın kontrolünde birden çok parametre için ölçüm yapabilen optik koherens tomografi (OCT) [11] cihazının üç farklı tarama modülü ile Humphrey görme alanı ölçümü [12] kullanılmıştır. OCT cihazında kullanılan her bir tarama modülünün kendi içerisinde alt ölçüm parametreleri yer almaktadır. Yapılan ölçüm sonuçlarının tümü reel sayıdır. Makalenin bundan sonraki bölümlerinde OCT cihazının tarama modülleri Cihaz 1-2-3 ve Humphrey görme alanı ölçümü de Cihaz-4 şeklinde adlandırılmıştır. Buna göre cihazlar ve bunlara ait ölçüm parametreleri Tablo-2’de sunulmuştur

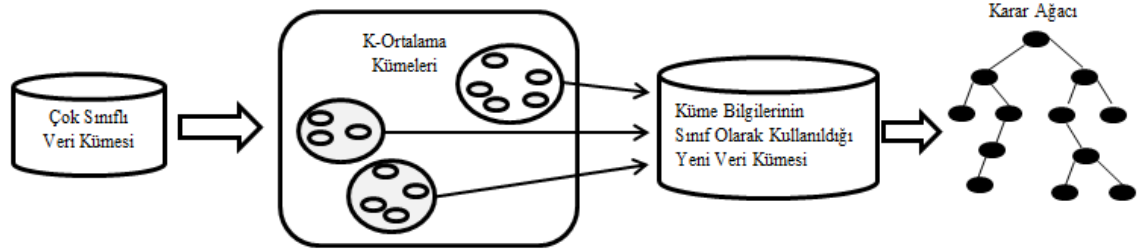
Tablo 2. Ölçüm cihazları ve parametreleri

Ölçüm Cihazı-1 (Optik Sinir Başı Analizi)	Ölçüm Cihazı-2 (Retina SinirLifi analizi)	Ölçüm Cihazı-3 (Makula Bölgesi Analizi)	Ölçüm Cihazı-4 (Humphrey Görme Alanı)
RIM AREA (optik sinir basısı)	AVG RNFL (Sinir lifi katmanı ortalama ölçüsü)	AVG GCL (Gangliyon ortalama kalınlığı)	MD (Ortalama sapma)
DISK AREA (optik sinir basısı yüzey alanı)	SUP (Süperiyör RNFL kalınlığı)	MIN GCL (Min. Gangliyon kalınlığı)	PSD (Desen standart sapma)
AVG C/D RATIO (disk çukuru – disk çapı oranı)	NASAL (Nazal RNFL kalınlığı)	1 – 6. Kadran GCL (Gangliyon kadran kalınlıkları)	
VERT C/D (dikey eksende çukur-çap oranı)	INF (İnferiyör RNFL kalınlığı)		
CUP VOLUME (disk çukuru alanı)	TEMP (Temporal RNFL kalınlığı)		

B. ÖNERİLEN MODEL

Veri kümesinin ikiden fazla sınıfa ayrıldığı çoklu sınıflama problemlerinde eğer örnek sayısı az ise, çoğunlukla her sınıfın kendi içine düşen örnek sayısı homojen olmamakta ve bu durumda bilinen sınıflama yöntemleri veriyi tam olarak öğrenmekte zorlanabilmekte ve sınıflama performansı beklenenin altında kalabilmektedir [8,9]. Bu gibi durumlarda veriyi daha homojen hale getirmek ve dolayısıyla sınıflama performansını artırmak için kümeleme yöntemlerinden yararlanılabilir [13].

Yapılan bu çalışmada örneklem boyutunun küçük ve verilerin çoklu sınıflara dağılımının homojen olmadığı durumlarda tıbbi teşhis performansını artırmak adına K-Ortalama yönteminden elde edilen kümelerin yeni sınıf bilgisi olarak Karar Ağacı'na verildiği bir model tasarlanmıştır. İlgili model Şekil-1'de sunulmuştur.



Şekil 1. K-Ortalama kümelerinin sınıf bilgisi olarak Karar Ağacı'nda kullanılması

Şekil-1'de sunulan modelin testi için glockom veri kümesi kullanılmıştır. 131 kişiden oluşan bu veri kümesi örneklem sayısı olarak büyük değildir. Bunun yanında Tablo-1'de belirtildiği üzere örneklemedeki kişilerin sınıflara dağılımının da homojen olmayışı ilgili veri kümesini bu çalışma için uygun kılmaktadır.

Sunulan bu modelin ilk adımında eldeki veriler sınıf etiketine göre değil, ölçüm cihazlarından elde edilen sonuçlardaki benzerliklerine göre kümelendirilmiştir. İlgili kümeleme için, uygulama kolaylığı, güvenilir performansı [14-16] ve tıp alanındaki kümeleme problemlerinde de sıkça tercih edilmesi [17-19] nedeniyle K-Ortalama yöntemi kullanılmıştır.

K-Ortalama yöntemini veri kümesi üzerinde uygulamak için açık kaynak kodlu Orange yazılımı [20] kullanılmıştır. Küme sayısının belirlenmesinde en düşük küme sayısı 3, en yüksek küme sayısı 7 olarak verilmiş, en uygun küme sayısının bulunmasında ilgili araç tarafından sağlanan ve bulgusal bir optimizasyon yöntemi olan silhouette [21] optimizasyonu kullanılmıştır. Optimizasyon sonucuna göre en uygun küme sayısı 3 olarak belirlenmiştir. Küme sayısının belirlenmesinden ardından, veri kümesindeki örnekler ölçüm cihazlarının verdiği sonuçlara göre üç kümeye ayrılmıştır. Veri noktalarının ve kümelerin uzaklık ölçütü olarak Öklit uzaklığı kullanılmıştır. Veri kümesindeki sınıfların K-Ortalama yönteminde oluşturulan üç kümeye dağılımı Tablo-3'de sunulmuştur.

Tablo 3. Veri kümesinde sınıfların K-Ortalama kümelerine dağılımı

Sınıflar	Küme 1	Küme 2	Küme 3
Sağlıklı Sınıf	1	23	2
Hafif Hasta Sınıfı	29	45	7
Orta Hasta Sınıfı	12	3	9

Oluşturulan modelin ikinci adımında K-Ortalama'dan elde edilen kümeler, veri kümesine yeni sınıf bilgisi olarak atanmıştır. Daha sonra K-Ortalama küme bilgileriyle oluşturulan yeni veri glokom teşhisi için bilgi kazancı tabanlı Karar Ağacına verilmiştir. Bu yöntem uygulama kolaylığı, elde edilen sonuçların kolay yorumlanabilirliği, görsel desteği ve tıptaki uygulamalarda da kullanılması nedeniyle tercih edilmiştir [22]. Karar ağacı gerçekleştirimi için Orange aracı kullanılmıştır. Ağacın oluşturulmasında bilgi kazancı oranı kullanılmış, ön budama için yapraktaki en düşük örnek sayısı 2 olarak belirlenmiştir.

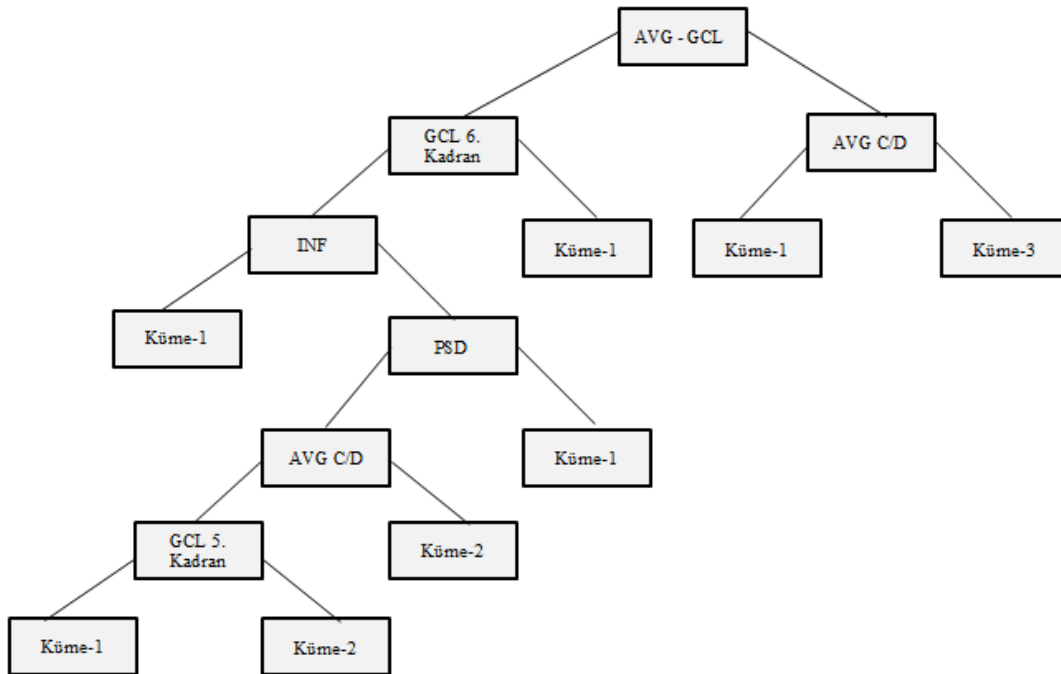
III. BULGULAR ve TARTIŞMA

Bu bölüm iki alt başlığa ayrılmaktadır. İlk alt başlıkta önerilen modelin glokom hastalığının teşhisindeki başarımı ve oluşan karar ağacının yapısı sunulmuştur. İkinci alt başlıkta ise modelin başarımı, makina öğrenmesinde sıkça kullanılan yöntemler olan Karar Ağacı, Yapay Sinir Ağları, Destek Vektör Makinaları, K-En Yakın Komşu yöntemleri ile karşılaştırılmıştır.

A. MODELİN BAŞARIMI ve KARAR AĞACI YAPISI

Çalışmaya konu olan modelin başarımlarını ölçümü için veri kümesi tasarlanan modele verilerek üzerinde k-katmanlı çarpaz geçerlilik testi yapılmıştır. Katman sayısı beş olarak belirlenmiştir. Elde edilen başarımların sonucu kesinlik¹ ve ROC eğrisi altında kalan alan ile ölçülmüştür. Buna göre modelin kesinlik ölçütü 0,808 ve ROC eğrisi altında kalan alan 0,839 olarak hesaplanmıştır.

K-Ortalama yönteminden elde edilen kümeler sınıf bilgisi olarak kullanılarak karar ağacına verildiğinde oluşan karar ağacı yapısı Şekil-2'de sunulmuştur.



Şekil 2. K-Ortalama - Karar Ağacı modelinin ağaç yapısı

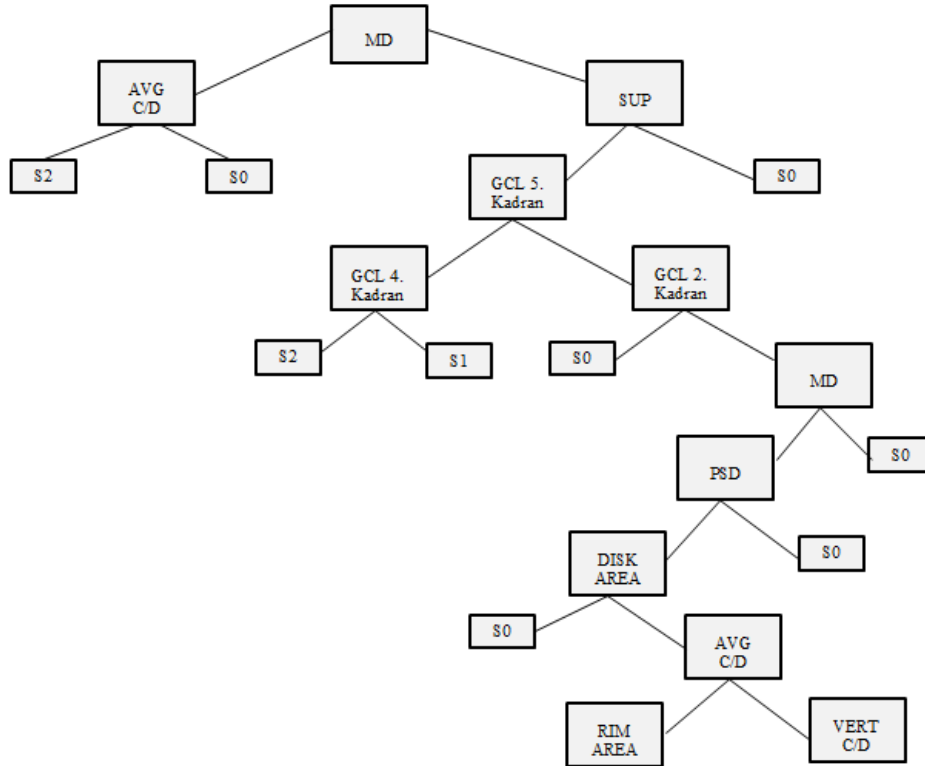
¹ Accuracy

Şekil-2’de görüldüğü üzere ilgili modelin glokom teşhisindeki ayırt ediciliği en yüksek parametre Cihaz-3’e ait gangliyon ortalama kalınlığıdır (AVG – GCL). İlgili ölçüm son yıllarda glokom teşhisi için üzerinde çalışılan önemli parametreler arasındadır [23,24]. Ağacın ikinci seviyesinde ise Cihaz-3’e ait başka bir parametre olan 6. Kadran GCL ve Cihaz-1 tarafından ölçülen optik sinir başı ölçümü olan disk çukurunun disk çapına oranı (AVG C/D) yer almıştır. Daha alt seviyelerde ise İnteriyor Sinir Lifi Kalınlığı (INF), Humphrey görme alanı için desen standart sapması (PSD), 5. Kadran GCL yer almıştır. Ağacın farklı seviyelerinde yer alan bu parametrelerin tümü glokom teşhisi için değerli parametreler olarak literatürde geçmektedir [25,26].

B. KARŞILAŞTIRMALI SONUÇLAR

Önerilen modelin başarımının çok sınıflı glokom teşhisinde avantaj sağlayıp sağlamadığını göstermek adına ilgili modelin başarımı bağımsız makina öğrenme yöntemleri ile karşılaştırılmıştır. Bu karşılaştırma için hem farklı tıp alanlarında kullanılan makina öğrenme yöntemleri hem de glokom teşhisi için kullanılan karar destek sistemlerinde sıklıkla tercih edilen yöntemler olan Karar Ağacı, Destek Vektör Makinası, K-En Yakın Komşu ve Yapay Sinir Ağı seçilmiştir [5,27]. Bu bağımsız yöntemler Orange aracı ile gerçekleştirilmiş ve başarım testi için 5 katmanlı çarpraz geçerlilik testi uygulanmıştır.

Karar ağacı oluşturulurken veri setinin düğümlere atanması ve dallandırılma için bilgi kazancı oranı ve ön budama için yapraktaki en düşük örnek sayısı 2 olarak kullanılmıştır. Elde edilen karar ağacı derinliği 15 dir. Bu nedenle kolay gösterim açısından ağaç derinliği 9 olarak sınırlandırılarak Şekil-3’de verilmiştir.



Şekil 3. Karar Ağacı yapısı

Modelin Şekil-2’de verilen karar ağacı yapısı ile Şekil-3’de verilen bağımsız karar ağacı yapısı karşılaştırıldığında hem farklı seviyelerdeki düğümlerde yer alan parametreler açısından hem de ağaç derinliği açısından farklar olduğu görülmüştür. Bağımsız karar ağacı incelendiğinde glokom teşhisi için ayırt ediciliği en yüksek parametre Humphrey görme alanı için ortalama sapma ölçütü (MD) olmuştur. Bağımsız karar ağacının başarımı 5 katmanlı çarpraz geçerlilik testi ile ölçüldüğünde kesinlik ölçütü 0,648 ve ROC eğrisi altında kalan alan 0,735 olarak çıkmıştır.

Glokom veri kümesinde teşhis performansı için test edilen ikinci bağımsız yöntem K-En Yakın Komşu yöntemidir. Bu yöntemde veri noktalarının uzaklığının hesaplanmasında Öklit Uzaklığı kullanılmış ve komşu sayısı 10 olarak belirlenmiştir. Yapılan başarımlar testinde kesinlik ölçütü 0,608 ve ROC eğrisi altında kalan alan 0,798 olarak çıkmıştır.

Çok katmanlı geriye yayımlı yapay sinir ağı modelinde giriş ve gizli katman için 20 nöron ve eğitim için 300 iterasyon kullanılmıştır. Modelin başarımlar ölçütlerinden kesinlik 0,609 ve ROC eğrisi altında kalan alan 0,840 olarak ölçülmüştür.

K-Ortalama – Karar Ağacı Modeli’nin başarımının karşılaştırıldığı bir diğer bağımsız makina öğrenme yöntemi Destek Vektör Makinası’dır. Bu modelde çekirdek için radyal temelli fonksiyon, ceza parametresi (c) üst sınırı 1, nümerik tolerans 10^{-3} , karmaşıklık sınırı (ν) 0,5 olarak belirlenmiştir. Bu şekilde kurulan modelin başarımlar ölçütleri kesinlik değeri 0,664 ve ROC eğrisi altında kalan alan 0,779 olmuştur.

K-Ortalama yöntemi ile elde edilen kümelerin yeni sınıf bilgisi olarak karar ağacına verildiği modelin kesinlik ve ROC eğrisi altında kalan başarımlar ölçütleri, bağımsız makina öğrenme yöntemlerinden Karar ağacı, Yapay Sinir Ağları, Destek Vektör Makinası ve K-En Yakın Komşu ile karşılaştırılmıştır. İlgili karşılaştırma sonuçları Tablo-4’de sunulmuştur

Tablo 4. K-Ortalama – Karar Ağacı Modelinin bağımsız makina öğrenme yöntemleri ile performans karşılaştırılması

Yöntem	Kesinlik	ROC Eğrisi Altında Kalan Alan
K-Ortalama – Karar Ağacı	0,808	0,839
Karar Ağacı	0,648	0,735
K-En Yakın Komşu	0,608	0,798
Yapay Sinir Ağları	0,609	0,840
Destek Vektör Makinası	0,664	0,779

Tablo-4 incelendiğinde bağımsız makina öğrenme yöntemleri arasında kesinlik ölçütü bakımından en iyi yöntem Destek Vektör Makinası, ROC eğrisi altında kalan alan bakımından da Yapay Sinir Ağı olmuştur. K-Ortalama-Karar Ağacı Modeli 0,808’lik kesinlik ve 0,839 olan ROC eğrisi altında kalan alan ile bağımsız yöntemlerden daha iyi bir performans göstermiştir.

Önerilen modelin başarımlar ölçütleri bakımından diğer yöntemlere göre yüzde bakımından sunduğu artış Tablo-5’de sunulmuştur.

Tablo 5. K-Ortalama – Karar Ağacı Modelinin yüzde olarak sağladığı başarımları kazancı

Yöntem	Modelin Sağladığı Kesinlik Artışı %	Modelin Sağladığı ROC Eğrisi Altında Kalan Alan Artışı %
Karar Ağacı	25	14
K-En Yakın Komşu	33	5
Yapay Sinir Ağları	32	=
Destek Vektör Makinası	22	8

Yukarıdaki tabloya bakıldığında önerilen modelin kesinlik artışı %22 - %33 aralığında değişirken, ROC eğrisi altında kalan artışı %5 - %14 arasında olmuştur.

IV. SONUÇ

Tıp alanı ürettiği yüksek miktardaki çok boyutlu veriler ile makina öğrenme yöntemlerinin odak noktalarından biri olmuştur [2]. Bu alanda yapılan ilk uygulamalar ikili sınıflamaya yönelik [5] olsa da günümüzde belirli bir hastalığın farklı seviyelerinin teşhis edilebildiği çoklu sınıflama yöntemlerine ihtiyaç duyulmaktadır. Çoklu sınıflama, eldeki veri kümesinin sayı olarak yeterli olmadığı durumlarda, farklı sınıflara düşen kişi sayısının azlığından etkilenebilmekte, bu durumda çoklu sınıflama yöntemleri yüksek teşhis performansı sunamamaktadır [8-9].

Yapılan bu çalışmada, bahsi geçen probleme çözüm olması amacı ile K-Ortalama ve Karar Ağacı yöntemlerinden oluşan bir model tasarlanmıştır. Bu modelde K-Ortalama yönteminin oluşturduğu küme bilgileri, yeni sınıf bilgileri olarak kullanılmış ve karar ağacına verilmiştir. İlgili modelin testi için, kişilerin çoklu sınıflar içine homojen olarak dağılmadığı Glokom veri kümesi kullanılmıştır. Bu veri kümesindeki kişiler, glokom teşhisi için kullanılan dört adet ölçüm (optik sinir başı analizi, retina sinir lifi analizi, makula bölgesi analizi ve Humphrey görme alanı ölçütü) ve toplam yirmi alt parametredeki benzerliklerine göre K-Ortalama yöntemi ile üç kümeye ayrılmıştır. İlgili kümeler yeni sınıf bilgisi olarak atanmış ve veri kümesi karar ağacına verilmiştir. Karar ağacından elde edilen sonuçlara göre, gangliyon ortalama kalınlığı, glokom teşhisinde ayırt ediciliği en yüksek parametre olarak çıkmıştır. Glokom teşhisi için literatürde geçen önemli diğer parametreler [23-26] de ağacın çeşitli seviyelerindeki düğümlerinde yer almıştır.

İlgili modelin sağladığı avantajı görmek adına önerilen model, Karar Ağacı, K-En Yakın Komşu, Yapay Sinir Ağları ve Destek Vektör Makinası yöntemleri ile karşılaştırılmıştır. Karşılaştırma için katman sayısının beş olarak belirlendiği çok katmanlı çarpaz geçerlilik testi kullanılmış ve başarımları ölçütü olarak kesinlik ve ROC eğrisi altında kalan alan kullanılmıştır. Model sunduğu 0,808'lik kesinlik ve 0,839 olan ROC eğrisi altında kalan alan ile karşılaştırdığı diğer yöntemlere göre daha üst bir başarımları sergilemiştir. Sunulan model karşılaştırıldığı bağımsız makina öğrenme yöntemine göre %22 - %33 kesinlik ölçütü ve %5 - %14 arası ROC eğrisi altında kalan alan artışı sağlamıştır.

Hem elde edilen başarımları sonuçları hem de ağaçta yer alan parametreler incelendiğinde K-Ortalama – Karar Ağacı Modeli'nin homojen olmayan çok sınıflı verilerin sınıflandırmasında ümit veren bir

yöntem olduğu görülmüştür. Bu modellerle ilgili kesin kaniya varmak için ise, ileriki çalışmalarda modelin benzer özellikli birçok veri kümesi üzerinde test edilmesi planlanmaktadır.

V. KAYNAKLAR

- [1] E.S. Berner, *Clinical Decision Support Systems: State of the Art*. AHRQ Publication, Rockville, MD (2009).
- [2] E. Coiera, *Clinical Decision Support Systems: Guide to Health Informatics*. 3rd Edition, CRC Press, (2003)
- [3] Sen, Arun et al. *Journal of Biomedical Informatics* **45(5)**(2012)1009–1017.
- [4] A. Satyanandam et al. *International Journal of Computer & Organization Trends*.**2(3)**(2012)53-60
- [5] I. Kononenko *Artificial Intelligence in Medicine* **23(1)**(2001)89–109
- [6] M. You et al. *Int J Data Min Bioinform.* **5(4)**(2011)383-401.
- [7] N. Mehra, S. Gupta *International Journal of Computer Science and Information Technologies* **4(4)**(2013)572 – 576
- [8] G. Tsoumakas, I. Katakis *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, 6. Baskı, IGI Global, (2008)
- [9] B.M. Shahshahani, D.A. Landgrebe. *Geoscience and Remote Sensing* **32(5)**(1994)1087 – 1095
- [10] S.W. Kim, R.P.W. Duin, *On Improving Dissimilarity-Based Classifications Using a Statistical Similarity Measure*, **15th Iberoamerican Congress on Pattern Recognition**, Sao Paulo – Brazil, (2010) 418–425
- [11] Savini et al. *Current Opinion in Ophthalmology* **22 (2)**(2011)115–123
- [12] Y. Burnstein et al. *American Journal of Ophthalmology* **129(3)**(2000)328–333
- [13] L. Churilov et al. *Journal of Management Information Systems* **21(4)**(2005)85-100
- [14] X. Wu et al. *Knowl. Inf. Syst.* **14(1)**(2008):1-37
- [15] M. Bramer, *Principles of data mining*, 1st Ed., Springer-Verlag, (2007)
- [16] Anonim,
http://www.academia.edu/4857097/Integrating_Clustering_with_Different_Data_Mining_Techniques_in_the_Diagnosis_of_Heart_Disease (Erişim tarihi: 17th of January, 2015)
- [17] N.S. Nithya et al. *International Journal of Computer Science Trends and Technology* **1(2)**(2013)17-23
- [18] P. Filipczuk et al. *Image Processing and Communications Challenges*, 3. Baskı, Springer, (2011)
- [19] U. Orhan et al. *Expert Systems with Applications* **38(10)**(2011)13475–13481.
- [20] J. Demsar et al. *Journal of Machine Learning Research* **14**(2013)2349-2353.
- [21] P. Rousseeuw et al. *Journal of Statistical Software* **1(4)**(1996)1-30.
- [22] A.T. Azar et al. *Neural Computing and Applications* **23(7)**(2013)2387-2403.
- [23] J.C. Mwanza et al. *Ophthalmology* **119(6)**(2012)1151–1158.
- [24] O. Tan et al. *Ophthalmology* **116(12)**(2009)2305–2314.
- [25] R. Sihota et al. *Invest Ophthalmol Vis Sci* **47(5)**(2006)2006-2010.
- [26] Z. Yang et al. *PLoS ONE* **10(5)**(2015)e0125957.
- [27] C. Bowd, M.H.Goldbaum *Optometry & Vision Science* **85(6)**(2008) 396–405.