



**TOPLULUK ÖĞRENME YÖNTEMLERİ İLE RENAL
HÜCRELİ KARSİNOM'UN TAHMİN EDİLMESİ**

Adem DOĞANER

BİYOİSTATİSTİK VE TIP BİLİŞİMİ ANABİLİM DALI

**Tez Danışmanı
Prof.Dr. Cemil ÇOLAK**

Doktora Tezi-2020

**T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**TOPLULUK ÖĞRENME YÖNTEMLERİ İLE RENAL HÜCRELİ
KARSİNOM'UN TAHMİN EDİLMESİ**


Adem DOĞANER

**Biyoistatistik ve Tıp Bilişimi Anabilim Dalı
Doktora Tezi**

**Tez Danışmanı
Prof. Dr.Cemil ÇOLAK**

MALATYA

2020

	KABUL ONAY FORMU	Doküman No	
		Yayın Tarihi	
Revizyon No			
Revizyon Tarihi			
Sayfa No			
İNÖNÜ ÜNİVERSİTESİ			
SAĞLIK BİLİMLERİ ENSTİTÜSÜ			

**İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ**

TOPLULUK ÖĞRENME YÖNTEMLERİ İLE RENAL HÜCRELİ KARSİNOM'UN
TAHMİN EDİLMESİ
DOKTORA TEZİ

DANIŞMAN
PROF. DR. CEMİL ÇOLAK

HAZIRLAYAN
ADEM DOĞANER

Jürimiz tarafından 11 / 02 / 2020 tarihinde yapılan tez savunma sınavı sonucunda bu tez ~~oybirliği~~ ~~oyçokluğu~~ ile başarılı bulunarak Biyoistatistik ve Tıp Bilişimi Anabilim Dalı Doktora Tezi olarak kabul etmiştir.

Jüri Üyelerinin Unvanı Adı Soyadı

1. Prof. Dr. Saim YOLOĞLU
2. Prof. Dr. Cemil ÇOLAK
3. Prof. Dr. Metin Fikret GENÇ
4. Prof. Dr. Zeki AKKUŞ
5. Dr. Öğr. Üyesi Mehmet Onur KAYA

İmza


.....

.....

.....

.....

.....

O N A Y

Bu tez, İnönü Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından kabul edilmiş ve Enstitü Yönetim Kurulu'nun .../.../20... tarih ve 20.../..... sayılı Kararıyla da uygun görülmüştür.

Prof. Dr. Yusuf TÜRKÖZ
Enstitüsü Müdürü

İÇİNDEKİLER

ÖZET	vi
ABSTRACT	vii
SİMGELER VE KISALTMALAR DİZİNİ	viii
ŞEKİLLER DİZİNİ.....	ix
TABLolar DİZİNİ	x
1.GİRİŞ.....	1
2.GENEL BİLGİLER	7
2.1.Topluluk Öğrenme	7
2.1.1.Topluluk Öğrenme Yöntemlerine İstatistiksel Yaklaşım.....	8
2.1.2.Topluluk Öğrenme Stratejisi	9
2.1.3.Bagging (Torbalama) Topluluk Öğrenme Yöntemi.....	9
2.1.4.Boosting (Artırma) Topluluk Öğrenme Yöntemi.....	11
2.1.5.Voting (Oylama) Topluluk Öğrenme Yöntemi	12
2.1.6.Stacking (İstifleme) Topluluk Öğrenme Yöntemi.....	14
2.2.Sınıflayıcılar	15
2.2.1.IB1 (En Yakın Komşu) Sınıflayıcı.....	15
2.2.2.IBk (K- En Yakın Komşu) Sınıflayıcı	15
2.2.3.LWL (Lokal Ağırlıklı Öğrenme) Sınıflayıcı.....	16
2.2.4.KStar Sınıflayıcısı.....	16
2.2.5.Random Forest (Rastgele Orman) Sınıflayıcı.....	16
2.2.6.REPTree (Azaltılmış Hata Budama Ağacı) Sınıflayıcı	17
2.2.7.SMO (Ardışık Minimal Optimizasyon) Sınıflayıcı.....	17
2.3.Araştırma Konusuna İlişkin Literatür	19
3.MATERYAL VE METOT	22
3.1.Örnek Büyüklüğü.....	23
3.2.Modeldeki Verilere Veritabanı Bilgi Keşfi Sürecinin Uygulanması	24
3.2.1.Veri Seçimi.....	24
3.2.2.Veri Önışleme	24
3.2.3.Veri Dönüşümü.....	24
3.2.4.Veri Madenciliği.....	25

3.2.5. Verilerin Değerlendirilmesi ve Yorumlanması.....	26
3.3. İstatistiksel Analiz.....	27
4. BULGULAR.....	28
4.1. Temel Sınıflayıcıların Performanslarının Değerlendirilmesi.....	29
4.2. Boosting Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi	33
4.3. Bagging Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi	36
4.4. Voting Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi	40
4.5. Stacking Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi	44
4.6. Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Değerlendirilmesi	47
5. TARTIŞMA	51
6. SONUÇ VE ÖNERİLER	56
KAYNAKLAR	57
EKLER.....	68
EK-1. ÖZGEÇMİŞ.....	68
EK-2. KLİNİK ARAŞTIRMALAR ETİK KURULU ONAY FORMU	69

TEŐEKKÜR

Doktora eđitim süresince, desteđini hiçbir zaman esirgemeyen, bilgi ve birikimlerini her zaman benimle paylaŐan, tez sürecinde yol gösterici olan deđerli danıŐmanım Sayın Prof. Dr. Cemil OLAK'a, doktora eđitimi boyunca ve tez sürecinde her zaman beni destekleyen ve deneyimlerini paylaŐan Sayın Prof. Dr. Saim YOLOĐLU'na sonsuz teŐekkürlerimi sunarım.

Desteklerini hiçbir zaman esirgemeyen Sayın Dr. Öğretim Üyesi Emek GÜLDOĐAN'a, Sayın Do.Dr. Harika Gözde GÖZÜKARA BAĐ'a, ArŐ. Gör. Ahmet Kadir ARSLAN'a, Biyoistatistik ve Tıp biliŐimi Anabilimdalı'nda görev yapan öğretim elemanlarına, her zaman yanımda olan ve destekleyen eŐim Gül TUZ DOĐANER'e ve aileme sonsuz teŐekkürlerimi sunarım.

Adem DOĐANER

ÖZET

Topluluk Öğrenme Yöntemleri ile Renal Hücreli Karsinom'un Tahmin Edilmesi

Amaç: Bu çalışmada farklı makine öğrenme yöntemlerinin birleştirilmesi ile elde edilen topluluk öğrenme yöntemleri ile renal hücreli karsinom hastalığının tahmin edilmesi amaçlanmıştır. Renal hücreli karsinomların sınıflandırılmasında, farklı sınıflayıcılar ve birleştirme teknikleri, topluluk öğrenme yöntemlerine uygulanarak en yüksek sınıflandırma performansını sağlayan topluluk öğrenme modelinin oluşturulması amaçlanmıştır.

Materyal ve Metot: Çalışmanın veri seti, Kahramanmaraş Sütçü İmam Üniversitesi, Sağlık Uygulama ve Araştırma Hastanesi Üroloji Kliniği'ne gelen hastaların verilerinden retrospektif olarak elde edilmiştir. Çalışmanın örneklemini 140 renal hücreli karsinom hastası ve 140 renal hücreli karsinomu olmayan toplam 280 hastadan oluşmaktadır. Veri setinde hemogram ve laboratuvar bulgularından oluşan 28 bağımsız ve bir bağımlı değişken yer almaktadır. Renal hücreli karsinomların tahmin edilmesinde boosting (artırma), bagging (torbalama), voting (oylama) ve stacking (istifleme) topluluk öğrenme yöntemi uygulanmıştır. Topluluk öğrenme yöntemlerinde, IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcıları kullanılmıştır.

Bulgular: Renal hücreli karsinomların tahmin edilmesinde en yüksek sınıflama performansı stacking topluluk öğrenme yönteminde IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcılarının birleştirilmesinden oluşan model sağlamıştır. Modelin sınıflandırma performans ölçütleri sırasıyla doğruluk değeri 0.906, duyarlılık değeri 0.906, kesinlik değeri 0.906, özgüllük değeri 0.910 ve AUC değeri 0.944 olarak elde edilmiştir.

Sonuç: Renal hücreli karsinomların tahmin edilmesinde topluluk öğrenme yöntemleri başarılı sınıflandırma gerçekleştirmiştir. Renal hücreli karsinomları en yüksek performans ile sınıflandıran topluluk yöntemi stacking topluluk öğrenme yöntemi olmuştur. Stacking topluluk öğrenme yönteminin tıbbi verilerde bilgi keşfi sürecinde uygulanması başarılı sonuçlar vermektedir. Topluluk öğrenme yöntemlerinde, birleştirme tekniklerinin geliştirilmesi ve uygun sınıflayıcıların modele dahil edilmesi, sınıflama performansını artıracaktır.

Anahtar Kelimeler: Makine öğrenme yöntemleri, sınıflayıcılar, renal hücreli karsinom, tıbbi bilgi keşfi, stacking topluluk öğrenme yöntemi, yapay zeka.

ABSTRACT

Ensemble Learning Methods based Prediction of Renal Cell Carcinoma

Objective: The aim of this research was to predict RCC disease by using ensemble learning methods obtained by combining different machine learning methods. It was aimed to create an ensemble learning model that provides the highest classification performance by applying different classifiers and combination of techniques to ensemble learning methods in the classification of RCC.

Material and Methods: The data set was obtained retrospectively from Urology Clinic of KSU, Health Practice-Research Hospital. The sample size consisted of 280 patients. The data set includes 28 predictors and one dependent variable. In the prediction of RCC ensemble learning methods of boosting, bagging, voting and stacking were applied. In the ensemble learning methods, IB1, IBk, KStar, LWL, random forest, REPTree and SMO classifiers were used.

Results: The model combining IB1, IBk, KStar, LWL, random forest and REPTree classifiers provided the highest classification performance in predicting RCC in the stacking ensemble learning method. The classification performance criteria of the model were obtained as: accuracy value 0.906, sensitivity value 0.906, precision value 0.906, specificity value 0.910 and AUC value 0.944.

Conclusion: Ensemble learning methods have been successful in predicting RCC. The ensemble method that classifies RCC with the highest performance has been the Stacking ensemble learning method. The implementation of the stacking ensemble learning method in the process of knowledge discovery in medical data yields successful results. In ensemble learning methods, the development of combination of techniques and the inclusion of appropriate classifiers in the model will improve classification performance.

Key Words: Artificial intelligence, classifier, machine learning, medical knowledge discovery, renal cell carcinoma, stacking ensemble learning methods.

SİMGELER VE KISALTMALAR DİZİNİ

Adaboost	:Adaptive boosting
AUC	:Eğri altında kalan alan
ALT	:Alanin aminotransferaz
AST	:Aspartat aminotransferaz
BT	:Bilgisayarlı tomografi
BUN	:Üre azotu
Hb	:Hemoglobin
Hct	:Hemotokrit
IB1	:En yakın komşu sınıflayıcısı
IBk	:K-en yakın komşu sınıflayıcısı
KStar	:Örnek tabanlı sınıflayıcı
KSU	:Kahramanmaraş Sütçü İmam University
LOF	:Lokal aykırı faktör
LWL	:Lokal ağırlıklı öğrenme sınıflayıcısı
MCV	:Ortalama korpuskuler hacim
MPV	:Ortalama trombosit hacmi
MRG	:Manyetik rezonans görüntüleme
Rbc	:Eritrosit kan hücresi
RCC	:Renal cell carcinoma
REPTree	:Azaltılmış hata budama ağacı sınıflayıcı
SMO	:Ardışık minimal optimizasyon sınıflayıcı
SS	:Standart sapma
SVM	:Destek vektör makinesi
TNM	:Tümör nodül metastaz
VTBK	:Veri tabanlarında bilgi keşfi süreci
WBC	:Beyaz kan hücresi

ŞEKİLLER DİZİNİ

Şekil No		Sayfa No
Şekil 1.1.	Veri tabanlarında bilgi keşfi süreci işlem aşamaları	4
Şekil 2.1.	Bagging topluluk öğrenme yöntemi işlem aşamaları.....	10
Şekil 2.2.	Boosting (artırma) topluluk öğrenme yöntemi işlem aşamaları .	11
Şekil 2.3.	Voting (oylama) topluluk öğrenme yöntemi işlem aşamaları.....	14
Şekil 2.4.	Stacking topluluk öğrenme yönteminin işlem aşamaları.....	14
Şekil 2.5.	K-en yakın komşu algoritması sınıflandırma örneği.....	15
Şekil 2.6.	Random Forest sınıflayıcısının işlem aşamaları	17
Şekil 2.7.	Destek vektör makinesinin yapısı	18
Şekil 2.8.	Doğrusal destek vektör makinesi ve doğrusal olmayan destek vektör makinesi	19
Şekil 3.2.	Güç analizine göre gruplar için önerilen örnek büyüklüğü.....	23
Şekil 3.3.	Özelliklerin seçimi.....	25
Şekil 4.1.	Temel sınıflayıcıların sınıflandırma performanslarının değerlendirilmesi.....	32
Şekil 4.2.	Boosting topluluk öğrenme yöntemlerinde sınıflayıcıların performansları	36
Şekil 4.3.	Bagging topluluk öğrenme yönteminde sınıflayıcıların performansları	40
Şekil 4.4.	Voting topluluk öğrenme yönteminde modellerin sınıflandırma performansları	43
Şekil 4.5.	Stacking topluluk öğrenme yönteminde modellerin sınıflandırma performansları	47
Şekil 4.6.	Topluluk öğrenme yöntemlerinin sınıflandırma performansları	49

TABLULAR DİZİNİ

Tablo No		Sayfa No
Tablo 2.1.	Bagging (torbalama) topluluk öğrenme algoritması.....	10
Tablo 2.2.	Adaboost topluluk öğrenme algoritması	12
Tablo 3.1.	Modeldeki değişkenlerin özellikleri	22
Tablo 3.2.	Sınıflayıcılar için parametre optimizasyon değerleri.....	26
Tablo 3.3.	Karmaşıklık matrisi.....	26
Tablo 4.1.	Renal hücreli karsinom varlığının dağılımı	28
Tablo 4.2.	Cinsiyet değişkeni açısından renal hücreli karsinom varlığının dağılımı	28
Tablo 4.3.	Laboratuar değerlerinin renal hücreli karsinom var olan ve olmayan hastalara göre karşılaştırılması	29
Tablo 4.4.	IB1 sınıflayıcısının performanslarının değerlendirilmesi	30
Tablo 4.5.	IBk sınıflayıcısının performanslarının değerlendirilmesi	30
Tablo 4.6.	KStar sınıflayıcısının performanslarının değerlendirilmesi.....	30
Tablo 4.7.	LWL sınıflayıcısının performanslarının değerlendirilmesi	31
Tablo 4.8.	Random Forest sınıflayıcısının performanslarının değerlendirilmesi	31
Tablo 4.9.	REPTree sınıflayıcısının performanslarının değerlendirilmesi	31
Tablo 4.10.	SMO sınıflayıcısının performanslarının değerlendirilmesi	32
Tablo 4.11.	Boosting topluluk öğrenme yöntemi ile IB1 sınıflayıcısının performanslarının değerlendirilmesi.....	33
Tablo 4.12.	Boosting topluluk öğrenme yöntemi ile IBk sınıflayıcısının performanslarının değerlendirilmesi.....	34
Tablo 4.13.	Boosting topluluk öğrenme yöntemi ile KStar sınıflayıcısının performanslarının değerlendirilmesi.....	34
Tablo 4.14.	Boosting topluluk öğrenme yöntemi ile LWL sınıflayıcısının performanslarının değerlendirilmesi.....	34
Tablo 4.15.	Boosting topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının performanslarının değerlendirilmesi.....	35

Tablo No		Sayfa No
Tablo 4.16.	Boosting topluluk öğrenme yöntemi ile REPTree sınıflayıcısının performanslarının değerlendirilmesi.....	35
Tablo 4.17.	Boosting topluluk öğrenme yöntemi ile SMO sınıflayıcısının performanslarının değerlendirilmesi.....	35
Tablo 4.18.	Bagging topluluk öğrenme yöntemi ile IB1 sınıflayıcısının performanslarının değerlendirilmesi.....	37
Tablo 4.19.	Bagging topluluk öğrenme yöntemi ile IBk sınıflayıcısının performanslarının değerlendirilmesi.....	37
Tablo 4.20.	Bagging topluluk öğrenme yöntemi ile KStar sınıflayıcısının performanslarının değerlendirilmesi.....	37
Tablo 4.21.	Bagging topluluk öğrenme yöntemi ile LWL sınıflayıcısının performanslarının değerlendirilmesi.....	38
Tablo 4.22.	Bagging topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının performanslarının değerlendirilmesi.....	38
Tablo 4.23.	Bagging topluluk öğrenme yöntemi ile REPTree sınıflayıcısının performanslarının değerlendirilmesi.....	39
Tablo 4.24.	Bagging topluluk öğrenme yöntemi ile SMO sınıflayıcısının performanslarının değerlendirilmesi.....	39
Tablo 4.25.	Voting topluluk öğrenme yöntemi ile model 1'in sınıflandırma performanslarının değerlendirilmesi.....	41
Tablo 4.26.	Voting topluluk öğrenme yöntemi ile model 2'nin sınıflandırma performanslarının değerlendirilmesi.....	41
Tablo 4.27.	Voting topluluk öğrenme yöntemi ile model 3'ün sınıflandırma performanslarının değerlendirilmesi.....	41
Tablo 4.28.	Voting topluluk öğrenme yöntemi ile model 4'ün sınıflandırma performanslarının değerlendirilmesi.....	42
Tablo 4.29.	Voting topluluk öğrenme yöntemi ile model 5'in sınıflandırma performanslarının değerlendirilmesi.....	42
Tablo 4.30.	Voting topluluk öğrenme yöntemi ile model 6'nın sınıflandırma performanslarının değerlendirilmesi.....	43

Tablo No		Sayfa No
Tablo 4.31.	Voting topluluk öğrenme yöntemi ile model 7'nin sınıflandırma performanslarının değerlendirilmesi.....	43
Tablo 4.32.	Stacking topluluk öğrenme yöntemi ile model 1'in sınıflandırma performanslarının değerlendirilmesi.....	44
Tablo 4.33.	Stacking topluluk öğrenme yöntemi ile model 2'nin sınıflandırma performanslarının değerlendirilmesi.....	45
Tablo 4.34.	Stacking topluluk öğrenme yöntemi ile model 3'ün sınıflandırma performanslarının değerlendirilmesi.....	45
Tablo 4.35.	Stacking topluluk öğrenme yöntemi ile model 4'ün sınıflandırma performanslarının değerlendirilmesi.....	46
Tablo 4.36.	Stacking topluluk öğrenme yöntemi ile model 5'in sınıflandırma performanslarının değerlendirilmesi.....	46
Tablo 4.37.	Stacking topluluk öğrenme yöntemi ile model 6'nın sınıflandırma performanslarının değerlendirilmesi.....	46
Tablo 4.38.	Stacking topluluk öğrenme yöntemi ile model 7'nin sınıflandırma performanslarının değerlendirilmesi.....	47
Tablo 4.39.	Topluluk öğrenme yöntemlerine ilişkin sınıflandırma performanslarının karşılaştırılması	48

1.GİRİŞ

Renal hücreli karsinom, böbrek korteksinden kaynaklanan ve böbrek tübülüslerinin iç yüzeyinden başlayarak büyüyen bir malignite tümör türüdür (1). Renal hücreli karsinom, tüm kanser türlerinin yaklaşık %2-3'ünü oluşturmaktadır. Son yıllarda renal hücreli karsinomların görülme hızı artış göstermektedir. Hastalığın artış insidansı yaklaşık %2 oranındadır (2). Hastalığın dünyadaki insidansı 4.4 (yüzbinde), kümülatif riski 0.51 (yüzbinde) olarak belirlenmiştir. Hastalığın Türkiye'deki mortalitesi 4.7 (yüzbinde) olarak belirlenmiştir (3). Böbrek malignitelerinin yaklaşık %85'ini oluşturan renal hücreli karsinomlar, genitoüriner hastalıklar arasında görülme sıklığı bakımından 3. sıradadır (1). Renal hücreli karsinomun etiolojisinde birincil risk faktörü sigara kullanımı olarak bilinmektedir (4). Sigara kullanımının yanı sıra, obezite, hipertansiyon, diyabet, böbrek taşları, kimyasal maruziyet, aile geçmişi ve böbrek travmaları renal hücreli karsinom riskini artıran diğer etiolojik faktörler olarak sıralanmaktadır (1, 5, 6).

Renal hücreli karsinomların sınıflandırılmasında güncel olarak Dünya Sağlık Örgütü'nün 2016 yılında yenilemiş olduğu liste dikkate alınmaktadır. Berrak renal hücreli karsinomlar, en sık gözlenen renal hücreli karsinom türüdür. Berrak renal hücreli karsinomlar tüm renal hücreli karsinomların yaklaşık %75'ini oluşturmaktadır. Papiller renal hücreli karsinomlar, berrak renal hücreli karsinomlardan sonra en çok gözlenen diğer bir karsinom türüdür. Gözlenme sıklığı açısından papiller renal hücreli karsinomu kromofob renal hücreli karsinomlar takip etmektedir (2).

Renal hücreli karsinomların evrelendirilmesinde tümör nodül metastaz (TNM) sistemi uygulanmaktadır. TNM sistemine göre evrelendirmede daha önceki dönemlerde farklı TNM evreleri kullanılmaktadır. En son kullanılan TNM evrelendirme sisteminde 2017 yılında değişikliğe gidilmiştir. Güncel olarak uygulanan evrelendirme, 2017 yılı TNM evrelendirmesidir. Primer tümörün tutulumu (TX, T0, T1a, T1b, T2a, T2b, T3a, T3b, T3c, T4), bölgesel lenf nodülleri (NX, N0, N1) ve uzak metastaz (M0, M1) gelişimine göre evrelendirme gerçekleştirilmektedir. Evre 1 aşamasında (T1, N0, M0) tümör çapı 7 cm'nin altında ve böbrek ile sınırlıdır. Herhangi lenf tutulumu veya uzak metastaz gelişmemiştir. Evre 2 aşamasında (T2, N0, M0) tümörün çapı 7 cm'yi aşmıştır ve böbrek ile sınırlıdır. Herhangi bir lenf tutulumu ve uzak metastaz yoktur. Evre 3 aşamasında tümör ana vane ve perirenal dokuya yayılmış olabilir veya lenf nodüllerinde tutulum gözlenebilmektedir. Evre 4 aşamasında ise tümör gerota kapsülüne dışına

yayılmış olabilir, lenf nodu tutulumu gerçekleşebilir veya uzak metastaz gerçekleşebilmektedir (2, 7, 8).

Renal hücreli karsinomlar uzak organlara metastaz eğilimindedirler. En sık görülen metastaz akciğerlerdedir. Akciğeri sırasıyla, karaciğer, kemik ve beyin takip etmektedir (9).

Renal hücreli karsinomlar, ilerleyen evrelerde dahi asemptomatik seyretmektedir. Bu nedenle hastalığın tanısı çoğunlukla ileri evrelerde gerçekleşmektedir. Böbrek bölgesi ile yan bölgelerde ağrılar, hematüri ve kitle, renal hücreli karsinom hastalarında ilk gözlenen belirtilerdir. Kilo kaybı, yüksek ateş, terleme, halsizlik gibi belirtilerin gözlenebildiği renal hücreli karsinomlarda metastaz gelişen ve tutulum gösteren organlara ilişkin şikayetler de gözlenebilmektedir (10-13).

Renal hücreli karsinomun tedavideki başarısı erken tanı ile artmaktadır. Renal hücreli karsinomların fiziki muayene ile erken tanı konma olasılığı düşüktür. Fiziki muayene sırasında el ile hissedilen kitle, lenfadenopati, ödem ve varikosel varlığı sonrasında radyolojik muayene, renal hücreli karsinomların tanısında gerekli uygulamalardır (2).

Renal hücreli karsinomların tanısında laboratuvar parametreleri oldukça önemlidir. İdrar analizleri, serum kreatinin, karaciğer fonksiyon testleri, hemoglobin, koagülasyon testleri, laktat dehidrogenaz, glomerüler filtrasyon oranı renal hücreli karsinomun ön tanısında yaygın olarak kullanılan laboratuvar testleridir (2,14). Laboratuvar testlerinden elde edilen hemogram ve biyokimyasal değerler hastalığın ön tanısında bilgi sağlarken, kesin tanı ultrasonografik ve bilgisayarlı tomografi görüntüleri ile alınabilmektedir. Tümörlerin karakterize edilmesinde MRG (manyetik rezonans görüntüleme) ve BT (bilgisayarlı tomografi) daha çok tercih edilen görüntüleme teknikleridir. İleri boyuttaki malignitelerin belirlenmesinde manyetik rezonans görüntüleme daha başarılıdır. Renal hücreli karsinomların evrelendirilmesinde yine manyetik rezonans görüntüleme daha çok tercih edilmektedir (2,15). Hastalığın metastatik durumu söz konusu olduğunda bilgisayarlı tomografi çoğunlukla tercih edilmektedir. Takipli hastalarda bir diğer tanı yöntemi renal biyopsidir. Her tedavi planında renal biyopsi gerekmemesine karşın, metastatik hastalarda tedavinin planlanması için genellikle istenmektedir (2,16).

Renal hücreli karsinomların tedavisinde en sık uygulanan tedavi yöntemi cerrahi yöntemlerdir. Metastaz gelişmemiş vakalarda en sık uygulanan tedavi yöntemi parsiyel veya total nefrektomidir. Metastaz gelişen renal hücreli karsinomlarda nefrektominin

yanısıra farklı tedavi yöntemlerinin içinde bulunduğu kombine tedaviler uygulanabilmektedir. Renal hücreli karsinomların birçok antikanser ilaçlarına karşı dirençli olması uygulanacak tedavi yöntemlerini sınırlandırmaktadır (13,17). Uygulanan tedavi, hastalığın evresine, metastaz durumuna, prognostik faktörlere, hastalığın histolojik ve klinik özelliklerine göre değişiklik gösterebilmektedir (2).

Kanser hastalıklarının erken teşhis edilmesi, tedavideki başarı oranını artırmaktadır. Aynı zamanda sağkalım sürelerinde önemli artışlar sağlanmasına ve tam kür elde edilebilmesine olanak tanımaktadır. Teknolojinin ilerlemesi ile görüntüleme tekniklerinde ciddi gelişmeler sağlanması, hastalıkların teşhisi için önemli bir aşama kaydedilmesini sağlamıştır. Sağlık alanında teknolojinin geliştiği bir diğer alan ise yapay zeka ve makine öğrenme olmuştur.

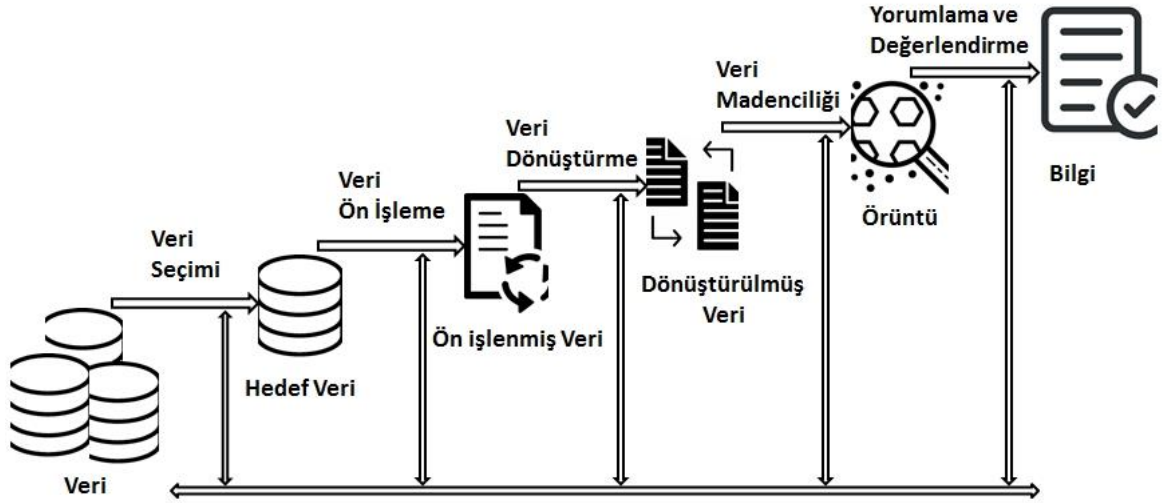
Makine öğrenme yöntemleri, hastalıkların teşhisinde ve klinik karar destek sistemlerinde son yıllarda kullanılan teknolojilerden biridir. Sağlık alanında geniş bir uygulama sahası bulan makine öğrenme, genetik hastalıkların belirlenmesinde, kanser hastalıklarının erken teşhisinde, tıbbi görüntüleme örüntülerinin tanımlanmasında, hasta takip ve teletıp gibi yapay zeka uygulamalarının temel altyapısını oluşturmaktadır.

Makine öğrenme yöntemleri, veri yığınındaki örüntüyü öğrenerek sınıflandırma ve çıkarım gerçekleştirmektedir. Makine öğrenme son yıllarda hızlı bir gelişme göstermiştir. Makine öğrenmenin bu denli hızlı gelişme göstermesi, bilgisayar ve internet ortamında veri yığınlarının hızla artması ile bu verilerden bilgi çıkarımı yapabilecek istatistiksel algoritmaların geliştirilmesine bağlı olmuştur. İnternet ortamında ve veri tabanlarında artan veri yığınları, bu verilerin işlenmesi ve bilgi sağlanması ihtiyacını doğurmuştur.

Veri yığınlarından çıkarım yapılmasını ve bilgi üretilmesini sağlayan yöntemler veri madenciliği disiplini altında değerlendirilmektedir. Veri madenciliği kısaca verideki örüntüleri keşfetme ve bilgi üretme süreci olarak ifade edilmektedir (18). Veri madenciliğinde veri yığınlarından otomatik olarak bilgi çıkarımı sağlanabilmektedir. Veri madenciliği yöntemleri bilgi çıkarımı işlemlerini gerçekleştirirken farklı istatistiksel yöntem ve makine öğrenme algoritmalarından yararlanmaktadır. Veri yığınlarından bilgi çıkarımı, esasında birçok aşamadan oluşan işlemler sürecidir.

Bu süreç VTBK (veri tabanlarında bilgi keşfi süreci) olarak ifade edilmektedir. Veri madenciliği, bilgi keşfi sürecinin sadece bir aşamasını oluşturmaktadır. Veri tabanlarında bilgi keşfi süreci, kısaca verilerdeki örüntülerden kullanışlı, anlaşılabilir, yeni ve geçerli bilgi sağlama süreci olarak ifade edilmiştir (19). Veri tabanlarında bilgi

keşfi süreci, veri yığınları içerisinde bulunan veriler arasındaki örüntüyü ortaya çıkararak yeni bilgiler elde edilmesi aşamalarını içermektedir. Veri tabanlarında bilgi keşfi sürecindeki aşamalar şekil 1.1.'de belirtilmiştir (19).



Şekil 1.1. Veri tabanlarında bilgi keşfi süreci işlem aşamaları

Veri seçimi aşamasında bilgi keşfi gerçekleştirilerek alana ilişkin ilgili veritabanlarından ham veriler elde edilmektedir.

Veri ön işleme aşamasında veritabanlarından elde edilen ham veriler içerisindeki uç değerler ve gürültülü veriler veri seti içerisinden çıkarılmaktadır. Benzer şekilde ham veri seti içerisinde yer alan eksik veya kayıp veriler yerine de uygun imputasyon yöntemlerinden yararlanılarak veri ataması gerçekleştirilmektedir. Veri tabanlarında bilgi keşfi sürecinde veri ön işleme için farklı algoritma ve tekniklerden yararlanılmaktadır.

Veri dönüşümü aşamasında veri tabanlarında bilgi keşfi için farklı değişkenlere ait veriler arasında tutarlılık sağlanması için verilere normalizasyon veya standardizasyon işlemleri uygulanmaktadır.

Bir diğer işlem aşaması olan veri madenciliğinde veri yığınlarında bulunan veriler arasındaki örüntüyü öğrenerek, yeni veriler için başarılı bir sınıflandırma veya çıkarım işlemi gerçekleştirilmek istenmektedir. Veri madenciliği birçok disiplinin birlikte uygulandığı bir alandır. Sınıflandırma ve çıkarım işlemleri, istatistik ve makine öğrenme tabanlı algoritma ve yöntemler aracılığı ile gerçekleştirilmektedir. Makine öğrenme algoritmaları veriden öğrenme yeteneklerine sahiptirler. Veri madenciliğinde

öğrenme yeteneklerine göre makine öğrenme yöntemleri denetimli öğrenme ve denetimsiz öğrenme olmak üzere iki farklı kategoride değerlendirilmektedir (20).

Değerlendirme ve yorumlama aşamasında, uygulanan makine öğrenme yöntemlerinin öğrenme performansları değerlendirilmektedir. Modelin performansının değerlendirilmesi için veri seti, eğitim veri seti ve test veri seti olmak üzere ikiye ayrılır. Eğitim veri setinde modeldeki makine öğrenme yöntemlerinin verideki örüntüyü öğrenmesi için eğitilir. Test veri setinde ise, örüntüyü öğrenen makine öğrenme yöntemlerinin yeni verilerde sınıflandırma performansı değerlendirilir. Değerlendirme işlemi modeldeki yanlılığı azaltmak adına farklı yöntemlerde geliştirilmiştir. K-katlı çapraz geçerlilik ve bootstrap örnekleme teknikleri öğrenme sırasındaki yanlılığı azaltmada sıklıkla tercih edilen tekniklerdir. Öğrenme performanslarının değerlendirilmesinde farklı metrikler kullanılabilir. Bu aşamada modelde kullanılan makine öğrenme yöntemlerinin bilgi keşfi için verideki örüntüyü öğrenme yeteneği ve öğrendikleri örüntü ile yeni verilerde sınıflandırma başarıları değerlendirilmektedir (21).

Makine öğrenme yöntemleri, sağlık alanında hastalıkların teşhisinde ve tahmin edilmesinde giderek artan bir hızla kullanılmaktadır. Hastalıkların tahmin edilmesinde makine öğrenme yöntemleri, genellikle sınıflandırma işlemi gerçekleştirilmektedir. Sınıflandırma, eğitim veri setindeki verilerin sınıfının bilinmesine bağlı olarak test veri setindeki yeni verilerin sınıflarının tahmin edilmesi işlemlerini içermektedir (22). Makine öğrenme yöntemleri çoğunlukla sınıflandırma işlemlerinde yüksek doğruluk performansına sahip olmasına karşın, bazı veri setlerinde istenilen performans düzeyine ulaşamamaktadır. Bu durumun farklı nedenleri vardır. Bu nedenlerden biri aşırı uyum (overfitting) problemidir. Aşırı uyumda eğitim veri setinde yüksek performans sağlanmasına karşın, test veri setinde düşük performans elde edilmektedir. Aşırı uyum sorunu modelin eğitim aşamasında veri setindeki örüntüyü öğrenmesinden ziyade veri setini ezberlemesinden kaynaklanmaktadır (23). Buda yüksek varyans problemine neden olmaktadır. Bir diğer neden ise yetersiz uyum (underfitting) problemidir. Öğrenmede yetersiz uyum gözlemlenmesinin temel nedeni yüksek yanlılıktır, buda veri seti için doğru modelin seçilmediği durumlarda ortaya çıkmaktadır (24). Veri setindeki sınıf dengesizliği, kayıp gözlemlerin yerine hatalı imputasyon gerçekleştirilmesi, gürültülü verilerin bulunması, yeterli örnek hacmine ulaşılmaması gibi nedenler makine öğrenme yöntemlerinin performanslarında ciddi kayıplara neden olabilmektedir (25). Öğrenme sırasında yaşanan bu performans kayıplarının iyileştirilmesine yönelik farklı

çözümler sunulmuştur. Bu yöntemlerden biride topluluk öğrenme yöntemleridir. Topluluk öğrenme yöntemleri veri setini bir makine öğrenme algoritmasının sınıflandırmasından ziyade, birden fazla makine öğrenme algoritmasının veriyi ayrı ayrı sınıflandırarak, her sınıflandırıcının tahminlerinden ortak bir sınıflandırma sonucu elde edilmesini sağlamaktadır. Böylelikle bir makine öğrenme yönteminin tahmin sonuçlarına göre, birden fazla makine öğrenme yönteminden elde edilen ortak tahmin sonuçları daha doğru, daha güvenilir ve daha yüksek performans sunmaktadır (26). Topluluk öğrenme yöntemleri birden çok sınıflayıcının, tek bir sınıflayıcının tahminlerine göre daha yüksek doğruluğa sahip sınıflandırma gerçekleştirebileceği esasına dayanmaktadır. Topluluk öğrenme yöntemleri elde ettiği başarılı sonuçlar ile son yıllarda geniş bir uygulama sahası bulmuştur. Yaygın bir şekilde kullanılan topluluk öğrenme yöntemleri birçok tanı ve teşhis çalışması içerisinde başarılı bir şekilde uygulanmıştır (27, 28).

Bu çalışmada birincil amaç olarak renal hücreli karsinomların erken teşhis ve tanı işlemlerine öncülük edecek yapay zeka uygulamaları için topluluk öğrenme yöntemleri kullanılarak renal hücreli karsinomların yüksek doğruluk performansları ile tahmin edilmesi amaçlanmıştır. Çalışmanın ikincil amacı ise, topluluk öğrenme yöntemlerinde model performansını en yüksek düzeye taşımak için, topluluk öğrenme yöntemleri olan stacking, voting, boosting ve bagging yöntemlerinde en uygun modelin belirlenmesi için sınıflayıcıların seçimi ve birleştirilmesi amaçlanmıştır.

2. GENEL BİLGİLER

2.1. Topluluk Öğrenme

Makine öğrenme yöntemleri, son yıllarda yapay zeka alanında ve hesaplama dayalı uygulamalı bilimlerde içerisindedir oldukça yaygın bir kullanım sahası elde etmiştir. Makine öğrenme yöntemlerinin bu başarısı, bünyesindeki algoritmaların başarılı tahmin ve yüksek doğruluğa sahip sınıflandırma işlemleri gerçekleştirmesine bağlıdır. Makine öğrenme yöntemleri güçlü algoritmalar ile birçok karmaşık veri setinde yüksek doğruluk performansı sağlarken, bazı veri setlerinde yüksek varyans ve düşük doğruluk değerine sahip sınıflandırmalar gerçekleştirmektedir. Sınıflandırma ve tahmin işlemlerindeki bu performans kaybını önlemek adına farklı metodlar öne sürülmüştür. Bu metodlardan biride topluluk öğrenme yöntemleridir. Topluluk öğrenme yöntemleri öne sürüldüğü ilk yıllarda makine öğrenme yöntemlerindeki yüksek varyansı azaltmak ve doğruluk oranını yükseltmek için uygulanmasına karşın kısa zamanda makine öğrenme yöntemlerinde sıklıkla karşılaşılan özellik seçimi, kayıp özelliklerin belirlenmesi, hata düzeltme, güven aralığının tahmini, dengesiz veri ve sınıflar gibi problemlerin çözümünde başarılı sonuçlar elde etmiştir (29). Topluluk öğrenme yöntemlerinin temel mantığı birçok noktadan alınan kararların, tek bir noktadan alınan karardan daha sağlıklı sonuçlar içereceği esasına dayanmaktadır. Bir uzmanın alacağı kararın hatalı olma riski, birden fazla uzmanın alacağı ortak kararın hatalı olma riskinden daha yüksektir. Bu bağlamda birden çok uzmanın ortak kararı, bir uzmanın kararından daha güvenilir ve daha yüksek doğruluğa sahip olmaktadır. Doğadan ve insanlar arası sosyolojik ilişkilerden esinlenerek geliştirilen topluluk öğrenme yöntemi, makine öğrenme yöntemlerinde yakın bir zamandan bu yana uygulanmasına karşın hesaplamalı bilimlerde yaygın bir kullanım alanına sahiptir (29).

İstatistik biliminde, 1977 yılında Tukey'in veri setine ikili lineer regresyon modelini önermesi ile topluluk çalışmalarına adım atılmıştır (30, 31). Bu günkü topluluk öğrenme mantığına çokta yakın olmayan çalışmalardan sonra doksanlı yıllarda Hansen ve Salamon çalışmalarında tahmin performansını geliştirmek üzere sinir ağlarından oluşan topluluk mantığını önermişlerdir (32). Bugün yine yaygın olarak kullanılan topluluk yöntemlerinden olan 'Adaboost'un temeli Schapire tarafından atılmıştır ve Freund ve Schapire zayıf sınıflandırıcıları birleştirerek güçlü bir topluluk öğrenme modeli elde etmişlerdir (31, 33, 34).

Topluluk öğrenme yöntemleri genel anlamda, bir makine öğrenme yönteminin tahmin performansından ziyade, birden fazla makine öğrenme yönteminin tahminlerinin birleştirilerek daha yüksek doğruluğa sahip ve daha yüksek performans sunan öğrenme yöntemleridir. Verilerin eğitimi birden fazla makine öğrenme yönteminde gerçekleştirilerek bir makine öğrenme yönteminden daha yüksek performansa sahip tahminler sağlanabilmektedir (35). Modelde, sınıflandırıcılar birleştirilmesinden ziyade sınıflandırıcıların elde ettikleri tahminler birleştirilecek ortak bir tahmin üretilmesi esasına dayanmaktadır (35).

2.1.1. Topluluk Öğrenme Yöntemlerine İstatistiksel Yaklaşım

Topluluk öğrenme yöntemleri, birden fazla makine öğrenme yönteminin tahminlerini birleştirerek daha yüksek performansa sahip bir sonuç üretmeyi amaçlamaktadır. Makine öğrenme yöntemlerinin tahmin performansını yükseltebilmesinin temelinde iki yolu bulunmaktadır. Bu durum yanlılığın azaltılması ve varyansın düşürülmesi ile gerçekleşebilmektedir (29). Topluluk öğrenme yöntemlerinin temel sınıflayıcıların genelleme hatasının azaltılmasını sağladığına ilişkin çok sayıda çalışma gerçekleştirilmiştir (31, 36-38). Topluluk öğrenme yöntemleri, yüksek varyansın azaltılmasında etkilidir. Varyansın azaltılmasında her bir sınıflayıcının tahminlerinin birleştirilerek ortalama bir tahmin elde edilmesi önemli bir etkidir. Sınıflayıcılar veri setlerinin sınıflandırılmasında farklı hatalar yaparken, başarılı sınıflandırmalarda benzer tahminler gerçekleştirmektedir. Bu durumda sınıflayıcıların tahminlerini birleştirerek ortalamasını almak, hata bileşenlerinin ortalama hatasının azalmasını sağlamaktadır (29). Topluluk öğrenme yöntemleri bazı durumlarda yüksek varyansın azaltılmasının yanısıra, yanlılığında azaltarak performansın artırılmasına katkı sunmaktadır (39). Topluluk öğrenme yöntemlerinde sınıflayıcı tahminlerinin birleştirilmesi her koşulda en yüksek doğruluk tahmini sunmayı garanti etmemektedir. Sınıflayıcıların tahminlerinin ortalamasını alarak birleştirme işlemi yapan bir topluluk öğrenme yöntemi, doğru birleştirme tekniği ve uygun sınıflayıcılar kullanılmadığında en yüksek sınıflandırma performansını sunan temel sınıflayıcıdan bazen daha düşük doğruluk değeri sağlayabilmektedir. Bu amaçla topluluk öğrenme yönteminde temel sınıflayıcılardan daha yüksek performans elde edilmesi için veri setine uygun birleştirme tekniği ve uygun sınıflayıcıların modele dahil edilmesi gerekmektedir. Doğru birleştirme tekniğinin uygulanması ile en yüksek performansın sağlanmasının yanı sıra, aşırı uyum problemini azaltılmasına katkı sağlanmaktadır (40, 41).

2.1.2. Topluluk Öğrenme Stratejisi

Topluluk öğrenme yöntemlerinde, tahmin performansının artmasını sağlayan bir diğer husus ise veri setinin yeterli örnek sayısına sahip olmasıdır. Veri setindeki tahminci değişken sayısı ve örnek sayısının yeterli sayıya sahip olması tahmin performansına olumlu katkı sağlamaktadır.

Topluluk öğrenme yönteminin başarısı, modeldeki sınıflayıcıların elde ettiği farklı hatalara ve tahminlere dayanmaktadır. Sınıflayıcılar veri setinden birbirlerine benzer hata ve tahminler elde etmeleri halinde topluluk yönteminin bu tahminleri birleştirmesi sonucu daha yüksek performanslı bir tahmin elde edemeyecektir. Bu amaçla sınıflayıcıların hatalarının birbirlerinden farklı olması topluluk sınıflandırma performansının artmasını sağlamaktadır. Sınıflayıcıların tahminlerinde farklılıkların elde edilmesi için eğitim veri setinin farklı alt kümeleri kullanılmaktadır. Eğitim veri setinin alt kümelerinin oluşturulmasında ve eğitilmesinde bootstrap yöntemi kullanılmaktadır (29, 42).

Topluluk öğrenme yönteminde sınıflandırma performansını etkileyen diğer bir husus ise uygun birleştirme yönteminin seçimidir. Birleştirme tekniğinin belirlenmesinde, sınıflayıcılar için uygun birleştirme tekniğinin seçimine dikkat edilmesi gerekmektedir. Birleştirme tekniklerine göre, eğitim veri seti için örneklem seçimine göre ve işlem adımlarına göre farklı topluluk öğrenme yöntemleri mevcuttur. Bu yöntemler, bagging (torbalama) topluluk öğrenme yöntemi, boosting (artırma) topluluk öğrenme yöntemi, voting (oylama) topluluk öğrenme yöntemi ve stacking (istifleme) topluluk öğrenme yöntemidir (29, 31, 35).

2.1.3. Bagging (Torbalama) Topluluk Öğrenme Yöntemi

Bootstrap aggregating kısaltması olarak ifade edilen bagging (torbalama) topluluk öğrenme yöntemi bootstrap örnekleme yöntemini esas almaktadır. Breiman tarafından geliştirilen yöntem eğitim veri setinin bootstrap örnekleme ile farklı alt boyutlarının eğitilmesi esasına dayanmaktadır (43). Bootstrap örnekleme yönteminde eğitim veri seti içerisinde her seferinde örneklerin yer değiştirmesi ile farklı alt örnekler oluşturulmaktadır. Oluşturulan her alt eğitim seti bir sınıflandırıcı ile eğitilmektedir. Aynı anda tüm sınıflayıcılar farklı alt eğitim setlerini sınıflandırmaktadır. Sınıflandırıcıların tahminlerini birleştirmek üzere bagging yöntemi çoğunluk oyu tekniğinden yararlanmaktadır. Bu teknikte tüm tahmincilerin sınıflandırma tahmini arasından sınıflandırıcıların verdikleri çoğunluk tahmini, topluluk

yönteminin sınıflandırma tahmini olarak kabul edilmektedir. Bagging algoritması tablo 2.1.'de belirtilmiştir (29).

Tablo 2.1. Bagging (torbalama) topluluk öğrenme algoritması

Girdi: Veri seti $D = \{d_1, d_2, \dots, d_n\}$ ve $d_i = (x_i, y_i)$ $x_i \in X$ ve $y_i \in \{-1, +1\}$

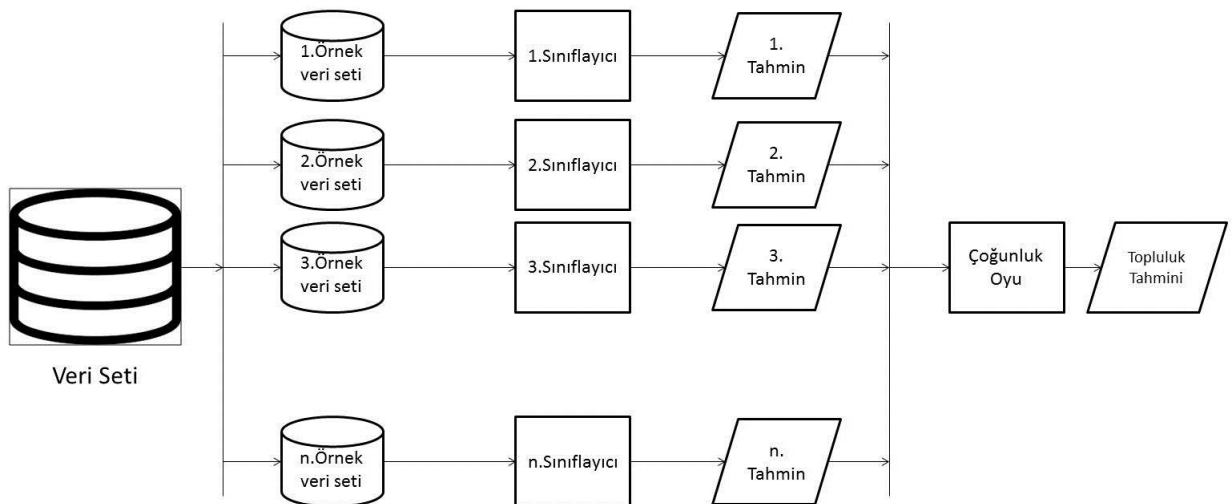
B : Bootstrap örneklem sayısı

Çıktı: $H: X \rightarrow \{-1, +1\}$ son sınıflayıcı.

İşlem Aşamaları

1. for b =1 to B do
2. b. Bootstrap örnek D_b olmak üzere, D kümesinden N örnek çekilerek yer değiştirilir.
3. Her bir D_b Bootstrap örneğinden, H_b sınıflayıcısı öğrenir.
4. end for
5. H_1, H_2, \dots, H_n sınıflayıcılardan çoğunluk oyu yöntemiyle son sınıflama sonucu elde edilir. $H(x) = \text{sign}(\sum_{b=1}^B H_b(x))$

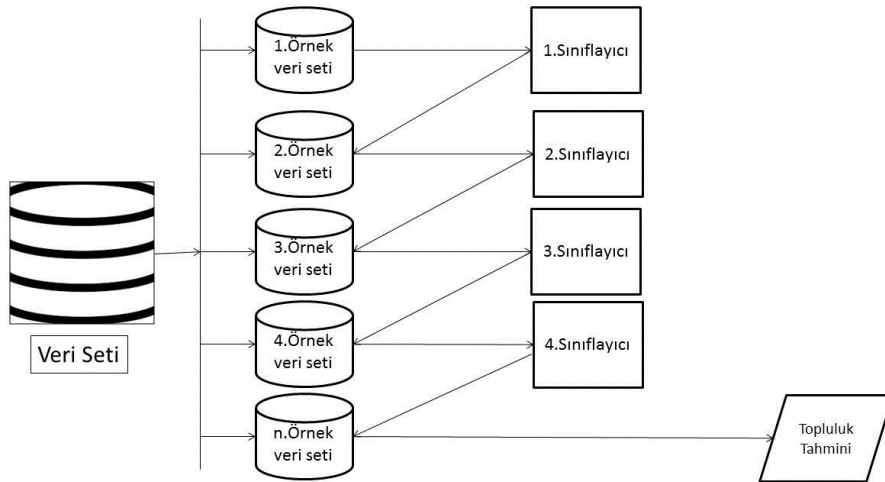
Bagging topluluk öğrenme yönteminde farklı türde sınıflayıcılardan ziyade tek türde sınıflayıcı, veri eğitiminde kullanılmaktadır. Tek tür sınıflayıcıdan bootstrap alt örnek eğitim veri seti sayısı kadar sınıflayıcı eş anlı olarak farklı alt veri setlerini eğitmektedir. Her sınıflayıcının tahminleri birleştirilmektedir. Bagging topluluk öğrenme yönteminin işlem aşamaları şekil 2.1.'de belirtilmiştir.



Şekil 2.1. Bagging topluluk öğrenme yöntemi işlem aşamaları

2.1.4.Boosting (Artırma) Topluluk Öğrenme Yöntemi

Boosting (artırma) topluluk öğrenme yöntemi zayıf öğreniciler için sınıflayıcıları birleştirilerek, tek bir sınıflayıcıya göre daha güçlü bir sınıflandırıcı elde etmeyi amaçlamaktadır. Zayıf öğrenici, rastlantısal tahminlerden daha az hata yapma olasılığına sahip sınıflayıcılar oluşturabilen öğrenme algoritmalarını ifade etmektedir. Boosting topluluk yöntemi, sınıflandırma işlemlerinde güçlü bir sınıflayıcı yerine çok sayıda zayıf sınıflayıcının birleştirilmesini esas almaktadır. Güçlü sınıflayıcıların kompleks yapıya sahip olmaları nedeniyle bu sınıflayıcıların eğitilmesi zor olmaktadır. Zayıf sınıflayıcıların eğitilmesinin kolay olması ile birden fazla basit sınıflayıcının tahminlerinin birleştirilmesi daha pratik bir sonuç sağlamaktadır. Boosting topluluk öğrenme yönteminde bagging topluluk öğrenme yönteminden farklı olarak, çok sayıda sınıflayıcı eş zamanlı olarak eğitim veri setlerini eğitmektedir. Boosting yönteminde iteratif işlemler uygulanmaktadır (31). Her iteratif işlemde bir önceki tahminde oluşan hatalar dikkate alınarak aynı hataların gerçekleşmemesini önlemek için boosting yöntemi ile güçlü bir sınıflayıcı elde edilmektedir (35). Boosting algoritmasında üç sınıflayıcı oluşturulmaktadır. İlk sınıflayıcı, eğitim veri seti içerisinde rastgele seçilen S1 veri setini sınıflandırmaktadır. İkinci sınıflayıcı, yarısı birinci sınıflayıcı tarafından yanlış sınıflandırılmış, diğer yarısı doğru sınıflandırılmış veri seti üzerine eğitilmektedir. Son sınıflayıcı daha önce iki sınıflayıcının uyum göstermediği veriler üzerine eğitilmiştir. Her sınıflayıcı tahminler üretmektedir. Boosting topluluk öğrenme yönteminde çoğunluk oyunu alan sınıflandırıcının tahminleri, topluluk tahmini olarak kabul edilmektedir (31). Boosting topluluk öğrenme yöntemine ilişkin işlem aşamaları şekil 2.2.'de belirtilmiştir.



Şekil 2.2. Boosting (artırma) topluluk öğrenme yöntemi işlem aşamaları

Adaboost algoritması, boosting topluluk öğrenme yöntemleri içerisinde güçlü ve yaygın şekilde uygulanan topluluk yöntemlerinden biridir. Freund ve Shapire'nin önerdiği adaboost yönteminde eğitim veri setinin alt kümelerinin eğitilmesi yerine, eğitim veri setinin farklı ağırlıkları kullanılmaktadır (34). İteratif işlemler ile öğrenen adaboost algoritması, sınıflama başarısını, ardışık öğrenme işlemlerinde bir sınıflayıcının ağırlığının bir önceki sınıflayıcının doğruluğuna bağlı olarak yeniden ağırlıklama ile elde etmektedir. Adaboost algoritmasında iteratif işlemlerde sınıflayıcılar, kendinden önceki sınıflayıcıların hatalarına ve yanlış sınıflandırmalarına odaklanmaktadır. Böylelikle her aşamada hataların azaltılması ve doğru sınıflandırmaların artırılması için olanak tanınmaktadır (29). Adaboost algoritması tablo 2.2.'de belirtilmiştir (29).

Tablo 2.2. Adaboost topluluk öğrenme algoritması

Girdi: Veri seti $D = \{d_1, d_2, \dots, d_n\}$ ve $d_i = (x_i, y_i)$ $x_i \in X$ ve $y_i \in \{-1, +1\}$

M :Maksimum sınıflayıcı sayısı

Çıktı: H sınıflayıcı olmak üzere $H: X \rightarrow \{-1, +1\}$.

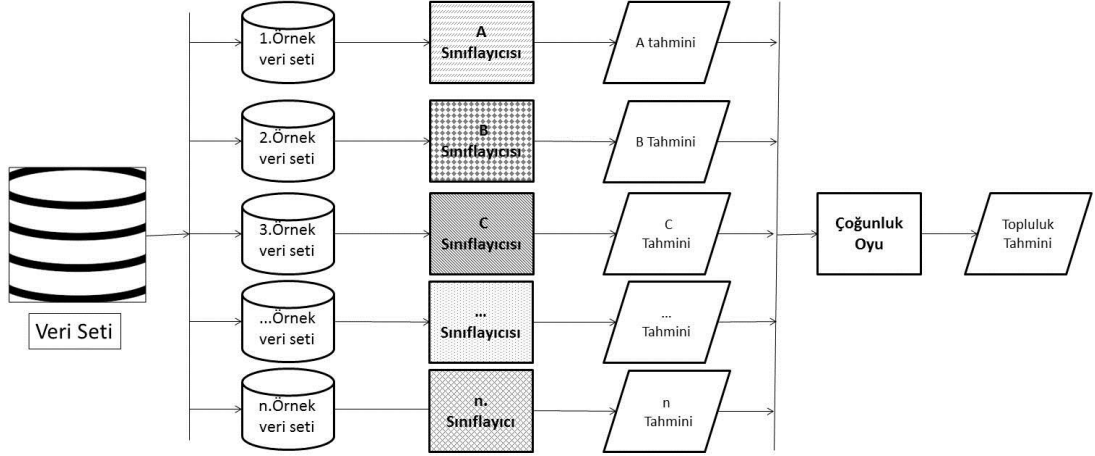
İşlem Aşamaları

1. Ağırlıklandırmayı başlatma $w_i^1 = \frac{1}{N}$ $i \in \{1, \dots, N\}$ ve $m = 1$ olmak üzere,
2. while $m \leq M$ do
3. Zayıf sınıflayıcılar D üzerinde çalıştırılır.
4. $hata_m = \sum_{i=1}^N w_i^m h(-y_i H_m(x_i))$ H_m in ağırlıklı hatası hesaplanır.
5. $\alpha_m = \frac{1}{2} \log \frac{1-err_m}{err_m}$. Zayıf sınıflayıcının ağırlığı hesaplanır.
6. $v_i^m = w_i^m \exp(-\alpha_m y_i H_m(x_i))$. Her i örneği için ağırlıklar güncellenir.
7. $S_m = \sum_{j=1}^N v_j$ ve $w_i^{(m+1)} = v_i^m / S_m$. ağırlıklar yeniden normalize edilir.
8. İterasyon sayacı artırılır: $m \leftarrow m + 1$
9. end while
10. Son sınıflayıcı: $H(x) = \text{sign}(\sum_{j=1}^M \alpha_j H_j(x))$.

2.1.5. Voting (Oylama) Topluluk Öğrenme Yöntemi

Voting (oylama) topluluk öğrenme yönteminde, farklı türdeki sınıflayıcıların tahminlerinin birleştirilerek çoğunluğun tahmini, topluluk tahmini olarak kabul edilmektedir. Esasında voting yöntemi bir birleştirme tekniği olmasına karşın farklı türdeki sınıflayıcıların birleştirilmesinde oldukça yaygın bir kullanım alanı

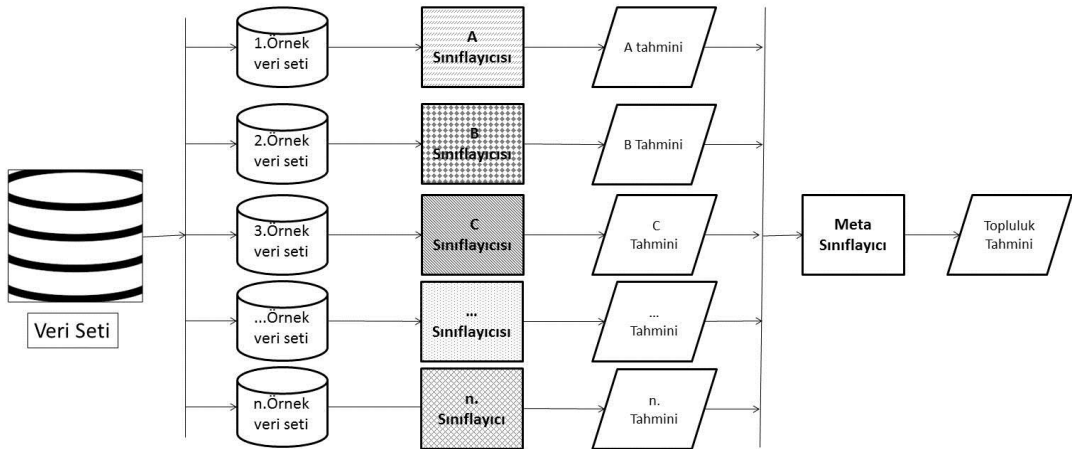
oluşturmuştur. Aynı veri setini farklı türdeki sınıflayıcıların sınıflandırması tahminlerde çeşitlilik sağlamaktadır. Tahminlerde çeşitliliğin oluşması, topluluk yönteminde doğruluk performansını artıran bir unsurdur. Bu nedenle voting topluluk yöntemlerinde genelde çeşitliliğin sağlanması durumunda yüksek performanslı tahminler elde edilebilmektedir (29). Boosting ve bagging topluluk yöntemlerinde aynı sınıflayıcıdan çok sayıda sınıflayıcı, veri setinin farklı alt kümelerini eğitmektedir. Voting yönteminde ise aynı veri setinin, farklı türdeki sınıflayıcılar tarafından sınıflandırılması söz konusudur. Farklı türdeki sınıflayıcılar, tahminlerdeki çeşitliliği artıracakları için topluluk tahminindeki hatanın azalmasını sağlayarak performans artışı sağlayabilmektedir. Öte yandan voting yönteminde eğitim veri setinin alt boyutundan ziyade, eğitim veri setinin tamamının modele dahil edilmesi, yine performansı artıran önemli bir etkidir. Voting topluluk öğrenme yöntemi birleştirme işleminde farklı tekniklerden yararlanmaktadır (31). Bu tekniklerden biri de majority voting (çoğunluk oyu) yöntemidir. Çoğunluk oyu yönteminde her sınıflayıcı bir tahmin gerçekleştirmektedir. Tahminlerdeki herhangi bir kararın sınıflayıcıların yarısından daha fazlası tarafından alınması durumunda topluluk kararı olarak geçerli olmaktadır. Çoğunluk oyu tekniğinde sınıflayıcıların tahmin kararlarında yarısından daha fazlası elde edilemediği durumlarda topluluk tahmini üretilmeyebilmektedir. Bu gibi durumlarda bir başka birleştirme tekniği olan plurality voting (oy çokluğu) tekniğinin kullanılması daha uygun olmaktadır. Oy çokluğunda sınıflayıcı tahminlerinde yarısından fazlasının kararı çoğunluğu aranmaksızın, en fazla tahmin edilen karar topluluk tahmini olarak kabul edilmektedir. Voting topluluk öğrenme yöntemindeki bireysel sınıflayıcıların performanslarının eşit olmaması durumunda, sınıflayıcıların performans ağırlıklarına göre birleştirme işlemi yapan weighted voting (ağırlıklı oylama) tekniğinin kullanılması daha uygundur. Voting topluluk öğrenme yöntemindeki işlem aşamaları şekil 2.3.'de belirtilmiştir (35).



Şekil 2.3. Voting (oylama) topluluk öğrenme yöntemi işlem aşamaları

2.1.6. Stacking (İstifleme) Topluluk Öğrenme Yöntemi

Wolpert tarafından geliştirilen stacking (istifleme) topluluk öğrenme yöntemi, farklı tür sınıflayıcıların tahminlerini meta sınıflayıcı için girdi kabul ederek bu tahminlerden daha yüksek performanslı bir tahmin üretme esasına dayanmaktadır (44). Stacking topluluk öğrenme yöntemi bu anlamda iki aşamalı bir öğrenme süreci sunmaktadır. İlk aşamada aynı eğitim veri setinden farklı tür sınıflayıcılar ile tahminler elde edilmektedir. İkinci aşamada ise ilk aşamadan elde edilen tahminler meta sınıflayıcıda işlenerek topluluk öğrenme modelinin tahmini elde edilmektedir. Tahminlerden meta sınıflayıcı ile topluluk tahmini oluşturan bu yöntem daha yüksek performans sağlamak üzere geliştirilmiştir. Stacking topluluk öğrenme yöntemine ilişkin işlem aşamaları şekil 2.4.'de belirtilmiştir.



Şekil 2.4. Stacking topluluk öğrenme yönteminin işlem aşamaları

2.2. Sınıflayıcılar

2.2.1. IB1 (En Yakın Komşu) Sınıflayıcı

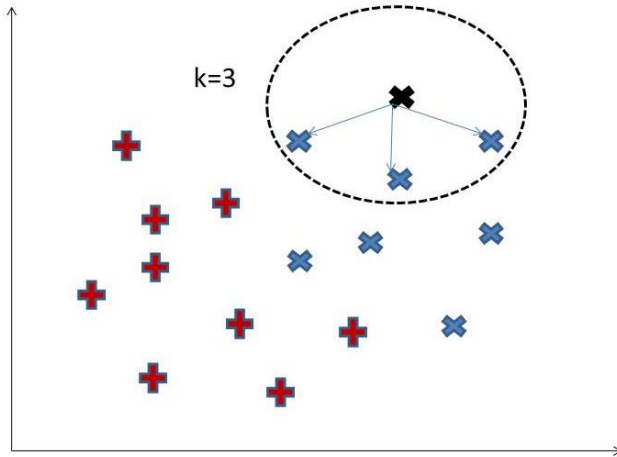
IB1 (Instance Based-1) sınıflayıcıları örnek tabanlı sınıflayıcılardan biridir. En yakın komşu algoritmasını esas alarak geliştirilmiştir. En yakın örneği bulmada normalleştirilmiş öklid uzaklığını kullanmaktadır. İki örnek arasındaki uzaklığı bulmada benzerlik fonksiyonundan yararlanmaktadır. Benzerlik Fonksiyonu,

$$D(x, y) = \sqrt{\sum_{i=1}^n f(x_i, y_i)}$$

eşitliği ile ifade edilmektedir. IB1 sınıflayıcısı kayıp verilere karşı toleransı, örneklerin aşamalı olarak işleme alınması ve özniteliklerin normalize edilmesi işlemleri dışında en yakın komşu algoritmasından farklı değildir (45-47).

2.2.2. IBk (K- En Yakın Komşu) Sınıflayıcı

IBk (Instance Based-k) sınıflayıcısı k en yakın komşu algoritması (k-NN) olarak ifade edilmektedir. Örnek tabanlı sınıflayıcılardan biri olan IBk çıktı uzayındaki nesnelere girdi uzayındaki nesnelere benzer olmasını dikkate almaktadır. Test örneğine en yakın olan k adet eğitim örneğini belirlemektedir (35). Denetimli öğrenme yöntemlerinden biri olan k- en yakın komşu algoritması benzerlik ölçülerine dayanarak örnekler arasındaki uzaklık ölçülerini belirlemektedir. Öklid, Minkowski gibi uzaklık ölçülerinin kullanıldığı k en yakın komşu algoritmasında, örneklerin birbirlerine olan uzaklıkları belirlenmektedir. İdeal bir k sayısı belirlenerek örneklerin sınıflandırılması sağlanmaktadır. Şekil 2.5.'te k en yakın komşu algoritmasına ilişkin sınıflandırma örneği belirtilmiştir (45, 48, 49).



Şekil 2.5. K-en yakın komşu algoritması sınıflandırma örneği

2.2.3. LWL (Lokal Ağırlıklı Öğrenme) Sınıflayıcı

LWL (Locally Weighted Learning) lokal ağırlıklı öğrenme algoritması, sınıflandırma işleminde örneklere ağırlıklar atamaktadır. LWL, test örneğine yakın konumda bulunan eğitim örneklerinin yakınlık konumuna göre ağırlıklar atamaktadır. En yakın örneğe en yüksek ağırlığı atayan LWL sınıflayıcısı, tahmin işleminde tüm veri seti yerine bir alt kümesini kullanmaktadır. Tüm veri setinden genel bir tahmin elde etmek yerine, test örneğine komşu eğitim örneklerinin yer aldığı lokal modeller elde etmektedir. Esnek bir kullanıma sahip olan LWL, örnek tabanlı sınıflayıcılar içerisinde etkin bir kullanıma sahiptir (50, 51).

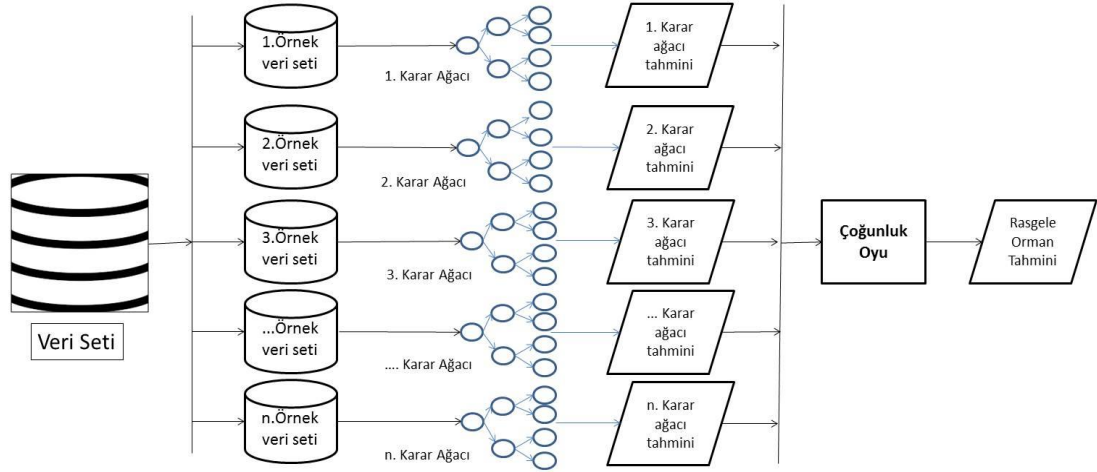
2.2.4. KStar Sınıflayıcısı

KStar (K^*) sınıflayıcısı, uzaklık ölçütü olarak entropiyi dikkate alan örnek tabanlı sınıflayıcıdır. En yakın komşu sınıflayıcısına benzerlik göstermesine rağmen KStar algoritması işlem aşamaları bakımından farklılıklar göstermektedir. Entropi uzaklığını dikkate almanın yanısıra, olasılık toplamlarında çözüme dahil etmektedir. Hem gerçek özellikler hemde sembolik özelliklerin entegrasyonu sağlanarak eksik veri problemi için çözüm sağlanmaktadır (52, 53).

2.2.5. Random Forest (Rastgele Orman) Sınıflayıcı

Breiman tarafından geliştirilen Random Forest (rastgele orman) sınıflayıcısı, oluşturduğu çok sayıda karar ağacı ile topluluk oluşturmaktadır (43). Bootstrap örnekleme yöntemi ile veri setinden farklı alt kümeler oluşturularak karar ağacı topluluğundaki her bir karar ağacı tarafından eğitilmektedir. Her bir düğüm için rastgele olarak belirlenen değişkenlerden en iyisi dikkate alınır ve düğümler dallara ayrılmaktadır (54). Random Forest algoritmasında 2 temel parametre, kullanıcı tarafından belirlenerek sınıflandırma başlatılmaktadır. Bu parametrelerden biri oluşturulması planlanan karar ağacı sayısı, diğer parametre ise bölünmeyi gerçekleştiren değişken sayısıdır. Random Forest sınıflayıcısı ağaç üretmek için sınıflandırma ve regresyon ağaçları (CART) yönteminden yararlanmaktadır. Random Forest yönteminin değerlendirme kriterlerinden biri GINI indeksidir. Sınıflandırma işleminde her bir ağacın tahmini alınarak oy çokluğu yöntemiyle sınıflayıcı kararı belirlenmektedir (43, 55). Random Forest sınıflayıcıları over fitting (aşırı uyum) sorununa karşı dirençli olmasının yanısıra hızlı çalışan bir sınıflayıcıdır. Veri setinin üçte ikisi eğitim veri seti, kalan kısmı ise test verisi olarak bölünmektedir. Çok sayıda karar ağacının birer

sınıflayıcı olarak kullanılması ve her sınıflayıcıdan elde edilen tahminlerden çoğunluk oyu ile genel bir tahmin üretilmesi, tahminlerdeki hata oranını düşürürken yanlışlııda azaltmaktadır. Güçlü bir sınıflayıcı olan Random Forest, sınıflandırma ve regresyon tahminlerinde yüksek performans sunmaktadır. Kayıp verilere karşı dirençli olan Random Forest sınıflayıcısı eksik verilerde de güçlü bir sınıflandırma gerçekleştirebilmektedir. Random Forest sınıflayıcısının işlem aşamaları şekil 2.6.'da belirtilmiştir (43, 55, 56).



Şekil 2.6. Random Forest sınıflayıcısının işlem aşamaları

2.2.6. REPTree (Azaltılmış Hata Budama Ağacı) Sınıflayıcı

REPTree (Reduced Error Pruning Tree) sınıflayıcı, karar ağacı tabanlı bir sınıflayıcı yöntemidir. Bilgi kazancını kriter olarak değerlendirir ve azaltılmış budama hatası ile karar ağacı oluşturmaktadır (57). Bilgi kazancının elde edilmesinde entropiden yararlanılmaktadır. Hızlı bir karar ağacı olan REPTree, hata azaltmada varyanstaki farklılıklardan yararlanmaktadır. REPTree sınıflayıcısı sayısal değerlerde sonuç üretmektedir (58-60).

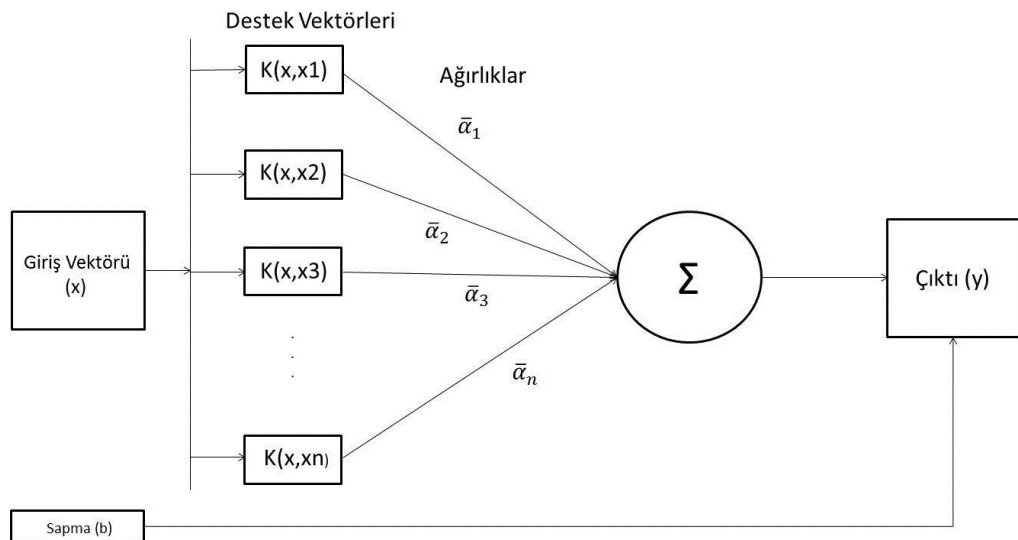
2.2.7. SMO (Ardışık Minimal Optimizasyon) Sınıflayıcı

SMO (Sequential Minimal Optimization) sınıflayıcı, esasen destek vektör makinesini temel alan bir sınıflayıcıdır. Platt tarafından geliştirilen SMO sınıflayıcısı, nümerik kuadratik programlama yerine analitik kuadratik programlama tekniklerini kullanmaktadır (61). SMO sınıflayıcısı, kuadratik optimizasyon problemlerini hızlı bir şekilde çözecek algoritmaya sahiptir. SMO sınıflayıcısı, içerisinde iki Lagrange çarpanının yer aldığı en küçük optimizasyon problemi ile ilgilenmektedir. Bu işlemi

iteratif olarak gerçekleştirerek her aşamada destek vektör makinesini güncellemektedir (61-64).

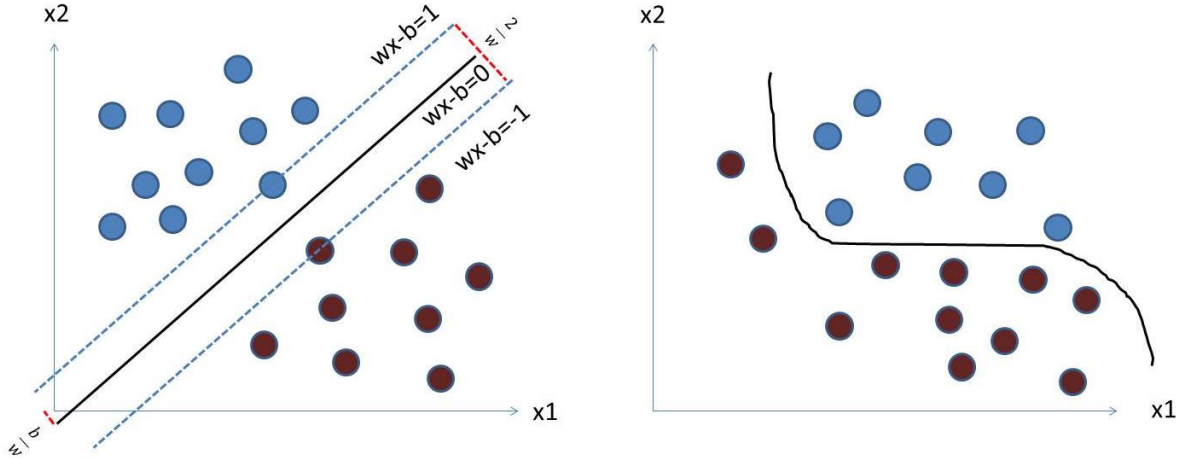
Destek vektör makinesi ise, sınıflandırılmak veya tahminde bulunulmak istenen değişkene ilişkin sınıfları birbirinden ayırabilecek en iyi düzlemleri belirlemeyi amaçlamaktadır (21, 65). Sınıflandırma ve regresyon modelleri için oldukça başarılı tahminler gerçekleştirebilen destek vektör modelleri doğrusal ve doğrusal olmayan modellerde uygulanabilmektedir. Destek vektör makinesi sınıflandırma problemlerini kareli optimizasyon formunda çözerek işlem basamaklarını azaltmaktadır. Destek vektör makinelerinin optimizasyon tekniğini esas alması, birçok karmaşık ve büyük veri setlerinde işlem kolaylığı ve hızlı çözüm olanağı sağlamaktadır (66, 67). Destek vektör makineleri dağılımdan bağımsız öğrenebilen, yapısal riski minimize etmek üzere tasarlanmış optimizasyon temelli denetimli öğrenme sınıflayıcılarıdır (68). Destek vektör makinesinin yapısına ilişkin grafik, şekil 2.7.'de belirtilmiştir.

Destek vektör makineleri farklı çekirdek fonksiyonları ile sınıflama ve tahmin işlemlerini gerçekleştirmektedir. Farklı veri setlerinde en uygun çekirdek fonksiyonunun belirlenmesi destek vektör makinesinin sınıflandırma ve tahmin performansını doğrudan etkilemektedir. En sık kullanılan çekirdek fonksiyonlarından biri radyal tabanlı çekirdek fonksiyonudur. Doğrusal çekirdek fonksiyonu, Laplace çekirdek fonksiyonu, polinomiyal çekirdek fonksiyonu, bessel çekirdek fonksiyonu ve hiperbolik tanjant çekirdek fonksiyonu destek vektör makinelerinde kullanılan diğer çekirdek fonksiyonlarından bazılarıdır (69, 70).



Şekil 2.7. Destek vektör makinesinin yapısı (71)

Destek vektör makineleri doğrusal ve doğrusal olmayan modeller için iki farklı şekilde değerlendirilmektedir. Doğrusal destek vektör makinesi ve doğrusal olmayan destek vektör makinesinin grafiksel gösterimi şekil 2.8.'de belirtilmiştir (72).



Şekil 2.8. Doğrusal destek vektör makinesi ve doğrusal olmayan destek vektör makinesi

2.3. Araştırma Konusuna İlişkin Literatür

Renal hücreli karsinomların makine öğrenme yöntemleri ile sınıflandırılmasına ilişkin çalışmaların başlangıcı yaklaşık on beş yıl öncelerine dayanmaktadır. Makine öğrenme ve topluluk öğrenme yöntemlerinin son yıllarda gelişme göstermiş olması nedeniyle çalışmaların başlangıcı yakın döneme aittir.

Won ve arkadaşları çalışmalarında örüntü analizlerinden yararlanarak renal hücreli karsinomu olan hastalar ile diğer ürolojik hastalıklara sahip hastaların doku örüntülerini makine öğrenme yöntemleri ile sınıflandırmışlardır. Sınıflandırma ve tahmin işlemlerinde bir karar ağacı yöntemi olan C4.5 sınıflayıcısı uygulanmıştır. Makine öğrenme yöntemlerinin hastalıkların erken teşhisinde önemli bir etken olabileceği ortaya konulmuştur (73).

Liu ve arkadaşları çalışmalarında renal hücreli karsinomların sınıflandırılmasında k-en yakın komşu algoritması ve genetik algoritmadan yararlanmışlardır. Çalışmada veri seti gen ekspresyonlarından elde edilmiştir. Yöntemin sınıflandırmada başarılı olduğu gözlemlenmiştir (74).

Fuchs ve arkadaşları çalışmalarında renal hücreli karsinomu olan hastaların patoloji bulgularından elde ettiği mikrodizi verilerini hesaplayarak hastaların sağkalım sürelerini tahmin edilebileceğini belirtmişlerdir. Çalışmada Random Forest

sınıflayıcısını kullanan arařtırmacılar, hastalık tahmininde uygulanan yöntemin bulguları ile patoloğların bulgularının örtüřtüğü sonucuna varmışlardır (75).

Gönen ve arkadaşları çalışmalarında renal hücreli karsinomların sınıflandırılmasında çoklu çekirdek öğrenme yöntemi ve destek vektör makinesi yöntemlerinden yararlanmışlardır. Çalışmada veri kaynaklarının doğrusal biçimde birleştirilmesini esas alan modellerinde mikro dizi görüntülerin sınıflandırılmasında başarı sağlanmışdır (76).

Lee ve arkadaşları çalışmalarında anjiyomiyolipom ile renal hücreli karsinomların sınıflandırılmasında k-en yakın komşu sınıflayıcısı, Random Forest sınıflayıcısı ve destek vektör makinelerinden yararlanmışdır. Modelin veri seti MDCT görüntülerinden elde edilmiştir (77).

Lin ve arkadaşları çalışmalarında bilgisayarlı tomografi görüntülerinden yararlanarak berrak hücreli renal karsinomun Fuhrman sınıflandırmasını tahmin etmişlerdir. Tahmin işleminde gradient boosting tekniğinden yararlanmışlardır (78).

Tabibu ve arkadaşları çalışmalarında Pan-renal hücreli karsinomların sınıflandırılmasına ilişkin çalışmalarında histopatolojik görüntülerin makine öğrenme yöntemleri ve konvolüsyonel sinir ağıları yöntemlerini uygulamışlardır. Çalışmalarında görüntülerin tanınmasında konvolüsyonel sinir ağıları kullanmalarının yanısıra sınıflandırma performansının artırılması için yeni bir destek vektör makinesi modeli öne sürmüşlerdir (79).

Topluluk öğrenme yönteminin klinik çalışmalarda sınıflandırma için kullanımı, makine öğrenme yöntemlerinin klinik çalışmalarda kullanılmaya başlaması ile ortaya çıkmıştır. Topluluk öğrenme yöntemlerinin yakın zamanda ortaya çıkması ve etkili bir yöntem olması, halen güncelliğini korumasını sağlamaktadır.

Tan ve Gilbert çalışmalarında topluluk öğrenme yöntemlerinden yararlanarak gen ekspresyonlarından kanser sınıflandırma işlemi gerçekleştirmişlerdir. Çalışmada C4.5 karar ağacını kullanan arařtırmacılar, bagging ve boosting topluluk öğrenme yöntemlerini modele dahil etmişlerdir (80).

Luo ve Cheng çalışmalarında göğüs kanserlerinin mamografi ile teşhisinde topluluk öğrenme yöntemlerinden yararlanmışlardır. Modelde karar ağacı ve destek vektör makinesi sınıflayıcıların birleşiminden oluşan modeli önermişlerdir. Topluluk modelin tekli sınıflayıcılara göre daha yüksek performans sağladığını gözlemlemişlerdir (81).

Nagi ve Bhattacharyya çalışmalarında kanserli vakalarda mikrodizi verilerinin sınıflandırılmasında topluluk öğrenme yöntemlerinden yararlanmışlardır. Modelde Naive Bayes, k- en yakın komşu ve j48 sınıflayıcıları yer alırken, boosting ve bagging topluluk yöntemleri uygulanmıştır (82).

Onan çalışmasında göğüs kanserlerinin sınıflandırılmasında topluluk öğrenme yöntemlerinden yararlanmıştır. Çalışmada farklı sınıflayıcılar ile birlikte farklı topluluk yöntemleri modele dahil edilmiştir. Topluluk yöntemlerinin temel sınıflayıcılardan daha iyi performans sağladıkları gözlemlenmiştir (83).

Wang ve arkadaşları çalışmalarında prostat kanserlerinin tespitinde istifleme tabanlı topluluk yöntemlerinde karar ağaçları sınıflayıcılarını uygulamışlardır. Model, kanser tespitinde başarılı sonuçlar elde etmiştir (84).



3.MATERYAL VE METOT

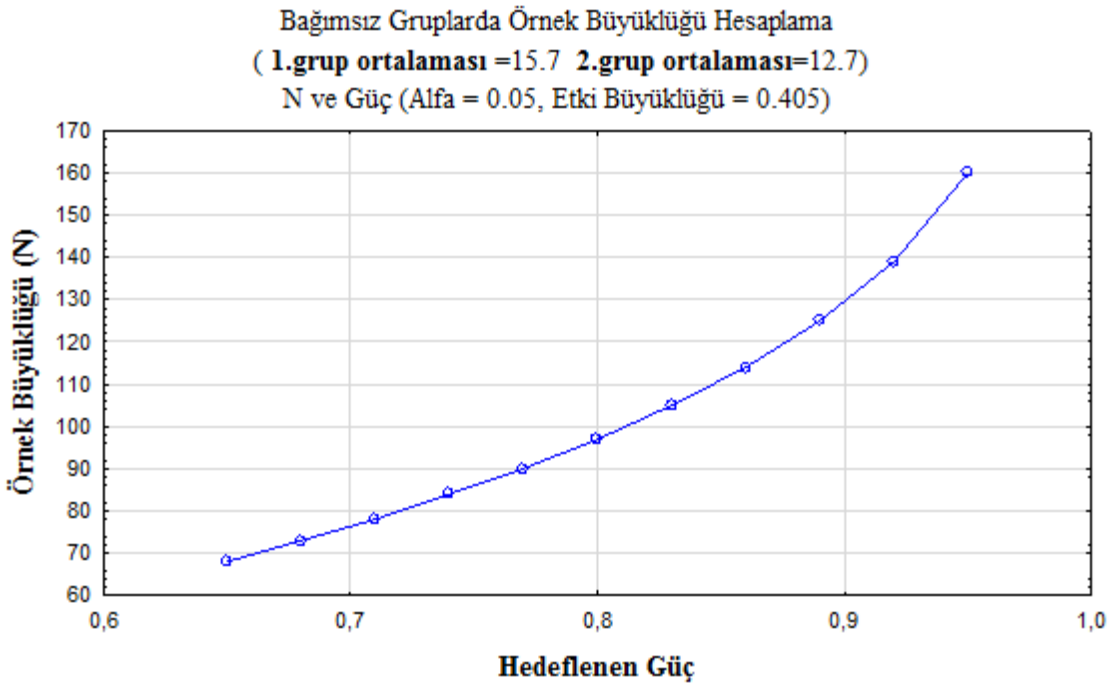
Bu çalışmanın gerçekleştirilmesinde, Kahramanmaraş Sütçü İmam Üniversitesi Klinik Araştırmalar Etik Kurulu 2018/07-21 sayılı protokolü ile izin onayı alınmıştır (Ek.2). Çalışmanın verileri Kahramanmaraş Sütçü İmam Üniversitesi Üroloji Servisi'ne gelen renal hücreli karsinom hastalığı teşhisi konmuş 140 hasta ile üroloji servisine gelen renal hücreli karsinom dışında farklı teşhis konulmuş 140 hastadan oluşmaktadır. Hastaların yaş, cinsiyet, hemogram değerleri ve biyokimyasal laboratuvar değerleri çalışmanın veri setini oluşturmaktadır. Çalışmada renal hücreli karsinomun sınıflandırılması tablo 3.1.'deki değişkenlere göre gerçekleştirilmiştir.

Tablo 3.1. Modeldeki değişkenlerin özellikleri

Öznitelikler	Değişken tipi	Değişken Açıklaması	Değişken Rolü
Renal hücreli karsinom	Kategorik	Var/ Yok	Bağımlı/Hedef
Cinsiyet	Nominal	Erkek/Kadın	Bağımsız/Tahminleyici
Patoloji	Nominal	Pozitif/Negatif	Bağımsız/Tahminleyici
Yaş	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
WBC (Beyaz kan hücresi)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Hb (Hemoglobin)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Hct (Hemotokrit)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
MCV(Ortalama corpuscular hacim)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Nötrofil	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Lenfosit	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Monosit	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Eozinofil	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Bazofil	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Rbc (Eritrosit kan hücresi)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Platelet	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
PLR	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
MPV (Ortalama trombosit hacmi)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
BUN (Üre azotu)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Kreatinin	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Sodyum	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
AST (Aspartat aminotransferaz)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
ALT (Alanin aminotransferaz)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Protrombin zamanı	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
INR(international normalized ratio)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
PTT (Parsiyel tromboplastin zamanı)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Potasyum	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
MCHC (ortalama hemoglobin yoğunluğu)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
NLR (nötrofil lenfosit oranı)	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici
Glukoz	Nümerik	Pozitif reel sayı	Bağımsız/Tahminleyici

3.1.Örnek Büyüklüğü

Çalışmanın örnek büyüklüğünün belirlenmesinde güç analizinden yararlanılmıştır. Referans çalışmalardaki bilgiler eşliğinde Ortalama±SS 12.7±7.2 ve Ortalama±SS 15.7±7.6 değerleri dikkate alınarak tip I hata (alfa) 0.05 düzeyinde, tip II hata (beta) 0.20 düzeyinde, 0.80 testin gücünde çalışmaya her grup için n:97 birey olmak üzere toplam iki grup için n:194 birey alınması gerektiği hesaplanmıştır. Şekil 3.2.'de güç analizi ile örnek sayısının belirlenmesine ilişkin bulgular grafikte belirtilmiştir. Veri madenciliği çalışmalarında çok değişkenli istatistiksel yöntemler uygulanmaktadır. Sadece güç analizi ile elde edilen örnek sayısı veri madenciliği çalışmaları için yeterli olmayabilmektedir. Çok değişkenli istatistiksel yöntemlerde önerilen örnek sayısı modeldeki değişken sayısının 5 katı ve üzeri olması yönündedir. Farklı kaynaklarda çok değişkenli istatistiksel yöntemlerde örneklem sayısının hesaplanmasında, değişken sayısı 6 ve üzeri olduğu durumlarda $n=104+k$ (k: tahminleyici değişken sayısı) ve her değişken için 10 kat örnek sayısı seçilmesi gerektiği belirtilmiştir (85, 86). Bu bilgiler eşliğinde modelde bulunan 28 bağımsız tahminleyici değişken için önerilen örnek büyüklükleri, değişken sayısının 5 katı için $5 \times 28=140$, $n=104+k$ eşitliği için $104+28=132$, değişken sayısının 10 katı için $n=10 \times 28=280$, güç analizi için $n=194$ birey olarak belirlenmiştir. Çalışmada önerilen en yüksek örnek sayısı $n=280$ dikkate alınarak 280 bireye ait veriler çalışmaya dahil edilmiştir.



Şekil 3.2. Güç analizine göre gruplar için önerilen örnek büyüklüğü

3.2.Modeldeki Verilere Veritabanı Bilgi Keşfi Sürecinin Uygulanması

3.2.1.Veri Seçimi

Veri seti, bir bağımlı değişken, 28 bağımsız değişken ve hasta veritabanından elde edilen toplam 280 hastanın verilerinden oluşmaktadır. Veri setinde, renal hücreli karsinom bağımlı değişkeni, hastaların yaşı, cinsiyeti, patoloji, beyaz kan hücresi, hemoglobin, hemotokrit, ortalama korpuskuler hacim, nötrofil, lenfosit, monosit eozinofil, bazofil, eritrosit kan hücresi, platelet, platelet lenfosit oranı, ortalama trombosit hacmi, üre azotu, kreatinin, sodyum, aspartat aminotransferaz, alanin aminotransferaz, protrombin zamanı, uluslararası normalleştirilmiş oran, parsiyel tromboplastin zamanı, potasyum, hemoglobin değeri, ortalama hemoglobin yoğunluğu miktarı, nötrofil lenfosit oranı ve glukoz değerleri ise bağımsız (tahminleyici) değişkenleri oluşturmaktadır.

3.2.2.Veri Ön İşleme

Veri setindeki değişkenlerde aykırı (uç) değer analizi, lokal aykırı faktörü (LOF) ile gerçekleştirilmiştir. Değerlendirme sonucunda uç değerler çalışma veri setinden çıkarılmıştır. Lokal aykırı faktör algoritması verilerin yerel yoğunluklarına bağlı olarak her gözlem için kendisine yakın nesnelerin uzaklıklarını ölçeklendirmektedir. Lokal aykırı faktör, uç değer ve aykırı gözlemlerin tespitinde kullanılan güçlü bir yöntemdir (87, 88).

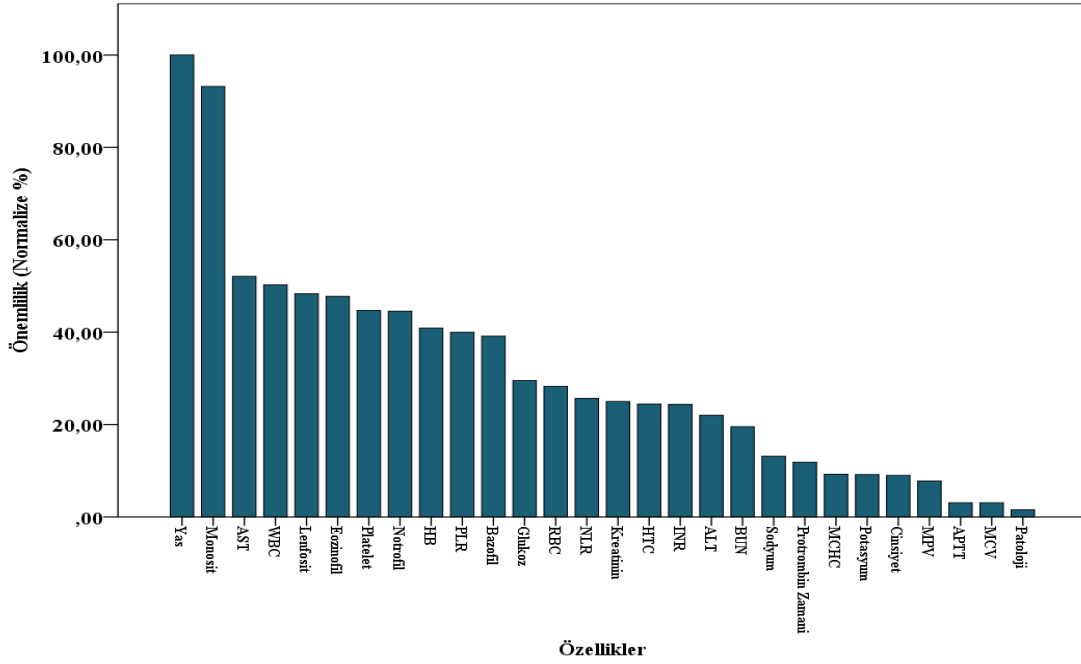
3.2.3.Veri Dönüşümü

Veri setinde bulunan nicel değişkenler için Z dönüşümü (standardizasyon) uygulanmıştır. Z dönüşümü aşağıdaki eşitlikte belirtilmiştir.

$$z = \frac{x - \mu}{\sigma}$$

μ , örneklem için aritmetik ortalamayı ve σ , örneklem için standart sapmayı ifade etmektedir. Veri madenciliği çalışmalarında sınıflandırma performansına katkısı olmayan özellikler çalışmadan çıkarılmaktadır. Katkısı olmayan özellikler çalışmadan çıkarılarak boyut indirgeme işlemi gerçekleştirilmektedir. Bu işlemlere özellik seçimi adı verilmektedir. Çalışmada bağımsız değişkenlerin veri madenciliği sürecine katkısını değerlendirmek üzere özellik seçimi işlemi uygulanmıştır. Özellik seçimi ve bilgi keşfi sürecinde sınıflandırma performansına katkılarının önemlilikleri yüzde ifadeler ile şekil 3.3.'te belirtilmiştir. Bu değerlere göre “patoloji” değişkeninin bilgi keşfi sürecine

katkısı çok düşük düzeyde ve önemsiz bulunmuştur. Bu amaçla “patoloji” değişkeni çalışmadan çıkarılmıştır.



Şekil 3.3. Özelliklerin seçimi

3.2.4. Veri Madenciliği

Renal hücreli karsinomların tahmin edilmesi için veri setine birçok makine öğrenme yöntemi uygulanmıştır. Uygulanan bu makine öğrenme yöntemlerinin temel sınıflayıcı performanslarının yanısıra, topluluk öğrenme yöntemleri içerisindeki performansları modelde değerlendirilmiştir. Modelde en iyi performansı sağlayan en yakın komşu yöntemi algoritmalarından olan IB1, IBK, Kstar, LWL, karar ağaçları algoritmalarından olan REPTree ve Random Forest ve fonksiyon tabanlı sınıflayıcılardan SMO sınıflayıcıları uygulanmıştır. Sınıflayıcıların hiperparametre optimizasyonları gridsearch ve multiseach yöntemleri ile gerçekleştirilmiştir. Sınıflayıcıların parametreleri ve en ideal parametre değerleri tablo 3.2.'de belirtilmiştir. Bu yöntemlerde temel sınıflayıcı performanslarının yanısıra bagging, boosting, voting ve stacking topluluk yöntemleri içerisindeki performanslarını değerlendirmek üzere modele dahil edilmiştir.

Tablo 3.2. Sınıflayıcılar için parametre optimizasyon değerleri

Sınıflayıcı	Parametre	İdeal parametre değeri
IBk	K, W, Uzaklık Ölçütü	3, 0, Öklit Uzaklığı
KStar	B(Karıştırma Parametresi), M(Kayıp değer parametresi)	20, Ortalama entropi eğrisi
LWL	K, U (Ağırlıklandırma Fonksiyonu)	2, Lineer
Random Forest	P(Oran), İterasyon, Özellik Sayısı, Maksimum Derinlik	100, 100, 0(Sınırsız), -1(Sınırsız)
REPTree	Yapraktaki toplam ağırlık, Minimum varyans, Maksimum Derinlik	2, 0.001, -1(Sınırsız)
SMO	C (karmaşıklık parametresi), ε, Tolerans parametresi	1, 1e-12, 0.001

3.2.5. Verilerin Değerlendirilmesi ve Yorumlanması

Veri setinde sınıflayıcıların ve topluluk yöntemlerinin performanslarının değerlendirilmesinde 10-katlı çapraz geçerlilik yönteminden yararlanılmıştır. Sınıflayıcıların ve topluluk yöntemlerinin sınıflandırma performanslarının değerlendirilmesinde doğruluk, duyarlılık, özgüllük, AUC (ROC eğrisi altında kalan alan) ve kesinlik metrikleri ile değerlendirilmiştir. Performans değerlendirmesinde kullanılan metriklerin hesaplanması tablo 3.3.'deki karmaşıklık matrisindeki bilgiler eşliğinde gerçekleştirilmektedir.

Tablo 3.3. Karmaşıklık Matrisi

		Referans		
		Pozitif	Negatif	Toplam
Tahmin	Pozitif	Doğru Pozitif (DP)	Yanlış Negatif (YN)	DP+YN
	Negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)	YP+DN
	Toplam	DP+YP	YN+DN	DP+YN+YP+DN

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YN + YP}$$

$$\text{Duyarlılık} = \frac{DP}{DP + YP}$$

$$\text{Özgüllük} = \frac{DN}{DN + YN}$$

$$\text{Kesinlik} = \frac{DP}{DP + YN}$$

3.3. İstatistiksel Analiz

Veri setine bilgi keşfi süreci uygulanmasında veri madenciliği yöntemlerinden IB1, IBK, Kstar, LWL, REPTree ve Random Forest ve SMO sınıflayıcıları kullanılmıştır. Topluluk öğrenme yöntemlerinden bagging, boosting, voting ve stacking yöntemleri farklı sınıflayıcılar ile birlikte uygulanmıştır. Stacking topluluk yönteminde meta sınıflayıcı olarak “logistic” sınıflayıcıdan yararlanılmıştır. Boosting topluluk yönteminin uygulanmasında “Adaboost” algoritmasından yararlanılmıştır. Hiperparametre optimizasyonu için gridsearch ve multiseach algoritmalarından yararlanılmıştır (89, 90, 91). Veri setinde bilgi keşfi sürecinin işletilmesinde R 3.6.0 yazılımı ve Weka 3.9 (Waikato Environmental Knowledge Analysis) açık kaynak kodlu yazılımlarından yararlanılmıştır (92, 93).

Verilerin normal dağılıma uygunluğu Kolmogorov-Smirnov testi ile değerlendirilmiştir. Normal dağılım gösteren değişkenlerde gruplar arasındaki farklılık bağımsız örneklerde t testi, homojenlik Levene testi ile değerlendirilmiştir. Normal dağılım göstermeyen değişkenlerde grup karşılaştırmaları için Mann-Whitney U testi kullanılmıştır. Kategorik değişkenler arasındaki frekans dağılımı farkı Pearson’s Ki-kare testi ve Fisher kesin Ki-kare testi ile değerlendirilmiştir. İstatistiksel anlamlılık $p < 0.05$ olarak kabul edilmiştir. Verilerin değerlendirilmesinde Statistica 13.3 Akademik versiyon ve IBM SPSS versiyon 22 programından yararlanılmıştır (94, 95).

4.BULGULAR

Veri seti, 140 (%50) renal hücreli karsinom hastası ile 140 (%50) renal hücreli karsinom olmayan hastaların verilerinden oluşmaktadır. Grup dağılımları tablo 4.1.' de belirtilmiştir.

Tablo 4.1. Renal hücreli karsinom varlığının dağılımı

Renal hücreli karsinom	Sayı	Yüzde (%)
Var	140	50.0
Yok	140	50.0
Toplam	280	100.0

Cinsiyet değişkeni açısından renal hücreli karsinom varlığının dağılımı tablo 4.2.'de belirtilmiştir. Renal hücreli karsinom hastalığına sahip olan bireylerin % 51.4'ü erkeklerden oluşmaktadır. Renal hücreli karsinoma sahip kadınların oranı ise % 48.6 olarak elde edilmiştir. Renal hücreli karsinomu olmayan bireylerde erkeklerin oranı % 59.3 iken kadınların oranı % 40.7 olarak belirlenmiştir. Cinsiyet açısından renal hücreli karsinom varlığı incelendiğinde dağılımsal bir farklılık saptanmamıştır ($p=0.186$).

Tablo 4.2. Cinsiyet değişkeni açısından renal hücreli karsinom varlığının dağılımı

		Renal hücreli karsinom				p
		Var		Yok		
		Sayı	Yüzde (%)	Sayı	Yüzde (%)	
Cinsiyet	Erkek	72	51.4	83	59.3	0.186
	Kadın	68	48.6	57	40.7	

Renal hücreli karsinom hastalarında ve renal hücreli karsinom olmayan hastalarda laboratuvar değerlerine ilişkin karşılaştırma bulguları tablo 4.3.'te belirtilmiştir. Karşılaştırma sonuçlarına göre renal hücreli karsinomu olan hastalar ile renal hücreli karsinomu olmayan hastalar arasında yaş, hemoglobin, monosit, eozinofil, PLR, MPV ve INR değişkenleri açısından farklılık istatistiksel olarak anlamlı bulunmuştur ($p<0.05$). Yaş, eozinofil, PLR ve INR değerleri renal hücreli karsinomu olan hastalarda daha yüksek değerlerde görülürken hemoglobin, monosit ve MPV değerleri renal hücreli karsinomu olmayan hastalarda daha yüksek değerlerde olduğu gözlemlenmiştir. Diğer laboratuvar değerleri açısından renal hücreli karsinom hastaları ile renal hücreli karsinomu olmayan hastalar arasında istatistiksel olarak farklılık saptanmamıştır ($p>0.05$).

Tablo 4.3. Laboratuvar değerlerinin renal hücreli karsinom var olan ve olmayan hastalara göre karşılaştırılması

	Renal hücreli karsinom		p
	Var	Yok	
	$\bar{X}\pm SD$	$\bar{X}\pm SD$	
Yaş	58.61±13.17	46.10±14.53	<0.001
WBC	7.97±2.57	8.20±2.66	0.452
HB	13.47±2.42	14.06±2.15	0.033
HTC	40.97±6.38	41.65±5.50	0.344
MCV	82.27±12.28	83.15±5.52	0.442
MCHC	33.10±5.58	33.13±2.68	0.945
Nötrofil	5.01±2.09	5.28±2.51	0.332
Lenfosit	2.05±0.76	2.12±0.63	0.435
Monosit	0.50±0.22	0.56±0.28	0.035
Eozinofil	0.21±0.20	0.16±0.13	0.018
Bazofil	0.04±0.06	0.04±0.02	0.824
RBC	4.81±0.85	5.40±4.55	0.133
Platelet	284.41±90.42	277.16±77.56	0.472
NLR	2.58±1.20	2.52±1.38	0.720
PLR	159.18±86.98	140.99±57.67	0.040
MPV	9.21±1.28	10.09±0.99	<0.001
BUN	17.08±9.47	15.51±6.93	0.114
Kreatinin	0.91±0.63	0.87±0.41	0.527
Sodyum	141.39±4.78	140.66±3.50	0.147
Potasyum	4.58±0.60	4.47±0.43	0.076
AST	22.80±10.96	25.25±9.95	0.051
ALT	23.53±12.32	24.82±14.83	0.430
Glukoz	112.73±42.54	109.00±46.19	0.482
INR	1.04±0.12	0.97±0.10	0.001
Protrombin Zamanı	12.01±2.84	11.96±0.97	0.860
APTT	26.41±2.97	27.01±3.17	0.107

$\bar{X}\pm SD$: Ortalama±Standart Sapma; Bağımsız örneklerde t testi: gruplar arasındaki farklılık istatistiksel olarak anlamlıdır (p<0,05)

4.1. Temel Sınıflayıcıların Performanslarının Değerlendirilmesi

Renal hücreli karsinomların tahmin edilmesi için bilgi keşfi sürecinde öncelikle sınıflayıcıların topluluk öğrenme yöntemine bağlı olmaksızın bireysel sınıflandırma performansları değerlendirilmiştir. Modelde çok sayıda sınıflayıcı uygulanmış ve bunlar arasından veri setine uygun ve yüksek performans sağlayan sınıflayıcılar modele dahil edilmiştir. Yedi farklı sınıflayıcının modele dahil edildiği bilgi keşfi sürecinde IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcıları uygulanmıştır.

Tablo 4.4.'te IB1 sınıflayıcısının renal hücreli karsinomları sınıflandırma performansları değerlendirilmiştir. IB1 sınıflayıcısının doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.892 ve AUC değeri 0.849 olarak elde edilmiştir. IB1 sınıflayıcısı en yüksek performansı özgüllük metriğinde sağlamıştır.

Tablo 4.4. IB1 sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.892
AUC	0.849

IBk sınıflayıcısının sınıflandırma performansları tablo 4.5.'te belirtilmiştir. Tablodaki bilgilere göre IBk sınıflayıcısının doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.892 ve AUC değeri 0.849 olarak elde edilmiştir. IBk, en yüksek performansı özgüllük metriğinde sağlamıştır.

Tablo 4.5. IBk sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.892
AUC	0.849

KStar sınıflayıcısının sınıflandırma performansları tablo 4.6.'da belirtilmiştir. Tablodaki bilgilere göre KStar sınıflayıcısının doğruluk değeri 0.808, kesinlik değeri, 0.814 duyarlılık değeri 0.808, özgüllük değeri 0.862 ve AUC değeri 0.932 olarak elde edilmiştir. KStar sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

Tablo 4.6. KStar sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.808
Kesinlik	0.814
Duyarlılık	0.808
Özgüllük	0.862
AUC	0.932

LWL sınıflayıcısının sınıflandırma performansları tablo 4.7.'de belirtilmiştir. Tablodaki bilgilere göre LWL sınıflayıcısının doğruluk değeri 0.682, kesinlik değeri 0.684, duyarlılık değeri 0.684, özgüllük değeri 0.671 ve AUC değeri 0.726 olarak elde edilmiştir. LWL sınıflayıcısı en yüksek performansı AUC metriğinde elde edilmiştir.

Tablo 4.7. LWL sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.682
Kesinlik	0.684
Duyarlılık	0.684
Özgüllük	0.671
AUC	0.726

Random Forest sınıflayıcısının sınıflandırma performansları tablo 4.8.'de belirtilmiştir. Tablodaki bilgilere göre Random Forest sınıflayıcısının doğruluk değeri 0.867, kesinlik değeri, 0.867 duyarlılık değeri, 0.867, özgüllük değeri 0.861 ve AUC değeri 0.932 olarak elde edilmiştir. Random Forest sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

Tablo 4.8. Random Forest sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.867
Kesinlik	0.867
Duyarlılık	0.867
Özgüllük	0.861
AUC	0.932

REPTree sınıflayıcısının sınıflandırma performanslarına ilişkin bulgular tablo 4.9.'da belirtilmiştir. Tablodaki bulgulara göre REPTree sınıflayıcısının doğruluk değeri 0.706, kesinlik değeri 0.706, duyarlılık değeri 0.706, özgüllük 0.710 ve AUC değeri 0.729 olarak belirlenmiştir. REPTree sınıflayıcısı en yüksek performansı AUC metriğinde sağlamıştır.

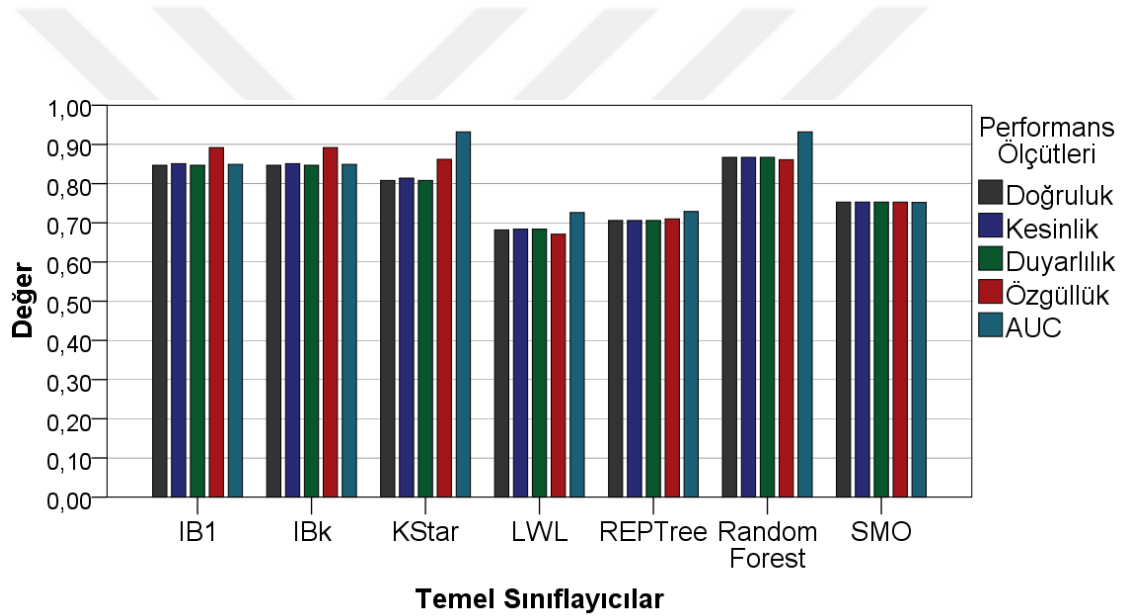
Tablo 4.9. REPTree sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.706
Kesinlik	0.706
Duyarlılık	0.706
Özgüllük	0.710
AUC	0.729

SMO sınıflayıcısının sınıflandırma performanslarına dair bulgular tablo 4.10.'da belirtilmiştir. Bulgulara göre SMO sınıflayıcısının doğruluk değeri 0.753, kesinlik değeri, 0.753, duyarlılık değeri, 0.753, özgüllük değeri 0.753 ve AUC değeri 0.752 olarak elde edilmiştir. SMO sınıflayıcısı tüm metrikler açısından yaklaşık olarak benzer performans göstermiştir.

Tablo 4.10. SMO sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.753
Kesinlik	0.753
Duyarlılık	0.753
Özgüllük	0.753
AUC	0.752



Şekil 4.1. Temel sınıflayıcıların sınıflandırma performanslarının değerlendirilmesi

Temel sınıflayıcıların performansları karşılaştırıldığında en yüksek doğruluk performansını Random Forest sınıflayıcısı sağlamıştır. Kesinlik ve duyarlılık metrikleri açısından sınıflayıcıların performansları değerlendirildiğinde Random Forest en yüksek performans sağlayan sınıflayıcı olmuştur. Özgüllük metriği açısından IB1 ve IBk sınıflayıcıları en yüksek performansı sağlamıştır. AUC değeri açısından en yüksek performansı KStar ve Random Forest sınıflayıcısı sağlamaktadır. Temel (bireysel) sınıflayıcıların sınıflandırma performanslarına ilişkin bulgular şekil 4.1.'de belirtilmiştir.

4.2.Boosting Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi

Renal hücreli karsinomların tahmin edilmesine dair bilgi keşfi sürecinde veri setine boosting topluluk öğrenme yöntemleri uygulanmıştır. Boosting topluluk öğrenme yönteminde modele IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcıları ayrı ayrı dahil edilmiş ve sınıflandırma performansları değerlendirilmiştir.

IB1 sınıflayıcısının boosting topluluk öğrenme yöntemi ile sınıflandırma performansına dair bulgular tablo 4.11.'de belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yönteminde IB1 sınıflayıcısının doğruluk değeri 0.824, kesinlik değeri 0.830, duyarlılık değeri 0.824, özgüllük değeri 0.879 ve AUC değeri 0.855 olarak elde edilmiştir. Boosting topluluk öğrenme yönteminde IB1 sınıflayıcısı en yüksek performansı özgüllük metriğinde sağlamıştır.

Tablo 4.11. Boosting topluluk öğrenme yöntemi ile IB1 sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.824
Kesinlik	0.830
Duyarlılık	0.824
Özgüllük	0.879
AUC	0.855

Tablo 4.12.'de boosting topluluk öğrenme yöntemi ile IBk sınıflayıcısının sınıflandırma performanslarına dair bulgular belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yönteminde IBk sınıflayıcısının doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.849 olarak elde edilmiştir. Boosting topluluk öğrenme yönteminde IBk sınıflayıcısı en yüksek performansı özgüllük metriğinde sağlamıştır.

Tablo 4.12. Boosting topluluk öğrenme yöntemi ile IBk sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.849

KStar sınıflayıcısının boosting topluluk öğrenme yöntemi ile sınıflandırma performanslarına ilişkin bulgular tablo 4.13.'de belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yöntemi ile KStar sınıflayıcısının doğruluk değeri 0.796, kesinlik değeri 0.803, duyarlılık değeri 0.796, özgüllük değeri 0.852 ve AUC değeri 0.831 olarak belirlenmiştir. Boosting topluluk öğrenme yöntemi ile KStar sınıflayıcısı en yüksek performansı özgüllük metriğinde elde etmiştir.

Tablo 4.13. Boosting topluluk öğrenme yöntemi ile KStar sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.796
Kesinlik	0.803
Duyarlılık	0.796
Özgüllük	0.852
AUC	0.831

Tablo 4.14.'de boosting topluluk öğrenme yöntemi ile LWL sınıflayıcısının sınıflandırma performanslarına dair bulgular belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yöntemi ile LWL sınıflayıcısının doğruluk değeri 0.784, kesinlik değeri 0.784 duyarlılık değeri 0.784, özgüllük değeri 0.791 ve AUC değeri 0.847 olarak elde edilmiştir. Boosting topluluk öğrenme yöntemi ile LWL sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

Tablo 4.14. Boosting topluluk öğrenme yöntemi ile LWL sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.784
Kesinlik	0.784
Duyarlılık	0.784
Özgüllük	0.791
AUC	0.847

Boosting topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının sınıflandırma performansına ilişkin bulgular tablo 4.15.'de belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının doğruluk değeri 0.863, kesinlik değeri 0.864, duyarlılık değeri 0.863, özgüllük değeri 0.845 ve AUC değeri 0.938 olarak elde edilmiştir. Random Forest sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

Tablo 4.15. Boosting topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.863
Kesinlik	0.864
Duyarlılık	0.863
Özgüllük	0.845
AUC	0.938

Tablo 4.16.'da Boosting topluluk öğrenme yönteminde REPTree sınıflayıcısının sınıflandırma performansına dair bulgular belirtilmiştir. Tablodaki bulgulara göre Boosting topluluk öğrenme yönteminde REPTree sınıflayıcısının doğruluk değeri 0.788, kesinlik değeri 0.788, duyarlılık değeri 0.788, özgüllük değeri 0.801 ve AUC değeri 0.878 olarak elde edilmiştir. Boosting topluluk öğrenme yönteminde REPTree sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

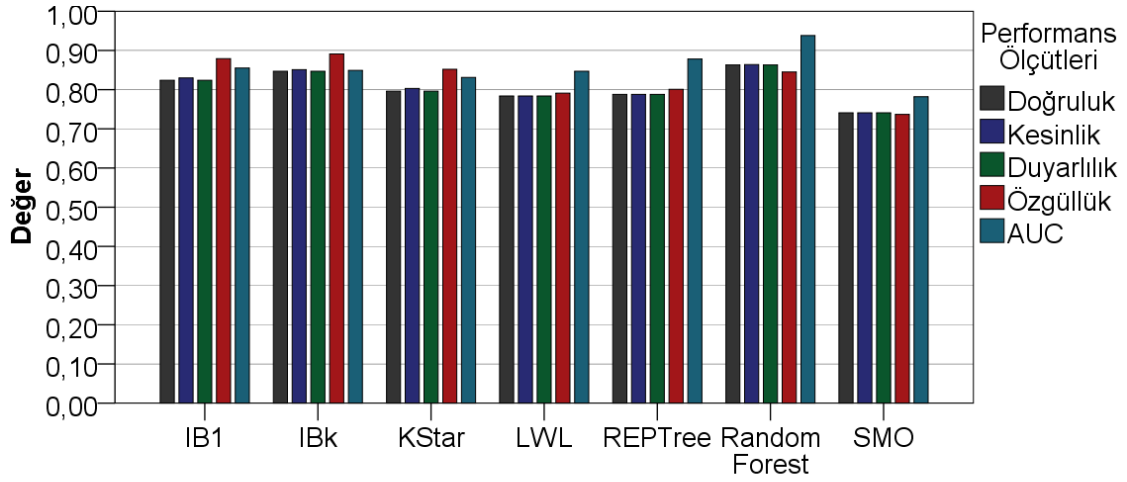
Tablo 4.16. Boosting topluluk öğrenme yöntemi ile REPTree sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.788
Kesinlik	0.788
Duyarlılık	0.788
Özgüllük	0.801
AUC	0.878

Boosting topluluk öğrenme yönteminde SMO sınıflayıcısının sınıflandırma performansına ilişkin bulgular tablo 4.17.'de belirtilmiştir. Tablodaki bulgulara göre boosting topluluk öğrenme yönteminde SMO sınıflayıcısının doğruluk değeri 0.741, kesinlik değeri 0.741, duyarlılık değeri 0.741, özgüllük değeri 0.737 ve AUC değeri 0.782 olarak elde edilmiştir. Boosting topluluk öğrenme yönteminde SMO sınıflayıcısı en yüksek performansı AUC metriğinde sağlamıştır.

Tablo 4.17. Boosting topluluk öğrenme yöntemi ile SMO sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.741
Kesinlik	0.741
Duyarlılık	0.741
Özgüllük	0.737
AUC	0.782



Boosting Topluluk Öğrenme Yöntemi ile Sınıflayıcılar

Şekil 4.2. Boosting topluluk öğrenme yöntemlerinde sınıflayıcıların performansları

Boosting topluluk öğrenme yöntemlerinde doğruluk metriğinde en yüksek sınıflandırma performansını Random Forest sınıflayıcısı elde etmiştir. Kesinlik, duyarlılık ve AUC metriklerinde yine en yüksek sınıflandırma performansını Random Forest sınıflayıcısı elde etmiştir. Özgüllük metriği açısından en yüksek sınıflandırma performansını IBk sınıflayıcısı sağlamıştır. Boosting topluluk öğrenme yöntemlerinde sınıflayıcıların performansları şekil 4.2.'de belirtilmiştir.

4.3. Bagging Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi

Bagging topluluk öğrenme yöntemi, veri setine bilgi keşfi sürecinde uygulanmıştır. Bagging topluluk öğrenme yönteminde modele dahil edilen IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcılarının sınıflandırma performansları değerlendirilmiştir.

Bagging topluluk öğrenme yönteminde IB1 sınıflayıcısının sınıflandırma performanslarına ilişkin bulgular tablo 4.18.'de belirtilmiştir. Tablodaki bulgulara göre bagging topluluk öğrenme yönteminde IB1 sınıflayıcısının doğruluk değeri 0.843, kesinlik değeri 0.847, duyarlılık değeri 0.843, özgüllük değeri 0.884 ve AUC değeri 0.894 olarak elde edilmiştir. IB1 sınıflayıcısı en yüksek performansı AUC metriğinde elde etmiştir.

Tablo 4.18. Bagging topluluk öğrenme yöntemi ile IB1 sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.843
Kesinlik	0.847
Duyarlılık	0.843
Özgüllük	0.884
AUC	0.894

Tablo 4.19.'da bagging topluluk öğrenme yönteminde IBk sınıflayıcısının sınıflandırma performansına dair bulgular belirtilmiştir. Bagging topluluk öğrenme yönteminde IBk sınıflayıcısının doğruluk değeri 0.847, kesinlik değeri 0.849, duyarlılık değeri 0.847, özgüllük değeri 0.879 ve AUC değeri 0.893 olarak elde edilmiştir. Bagging topluluk öğrenme yönteminde IBk sınıflayıcısı en yüksek performansını AUC metriğinde elde etmiştir.

Tablo 4.19. Bagging topluluk öğrenme yöntemi ile IBk sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.849
Duyarlılık	0.847
Özgüllük	0.879
AUC	0.893

Bagging topluluk öğrenme yönteminde KStar sınıflayıcısının sınıflandırma performanslarına ilişkin bulgular tablo 4.20.'de belirtilmiştir. Bulgulara göre bagging topluluk öğrenme yönteminde KStar sınıflayıcısının doğruluk değeri 0.800, kesinlik değeri 0.807, duyarlılık değeri 0.800, özgüllük değeri 0.853 ve AUC değeri 0.853 olarak elde edilmiştir. Sınıflayıcı en yüksek performansını özgüllük ve AUC metriklerinde elde etmiştir.

Tablo 4.20. Bagging topluluk öğrenme yöntemi ile KStar sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.800
Kesinlik	0.807
Duyarlılık	0.800
Özgüllük	0.853
AUC	0.853

Tablo 4.21.'de bagging topluluk öğrenme yönteminde LWL sınıflayıcısının sınıflandırma performanslarına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre bagging topluluk öğrenme yönteminde LWL sınıflayıcısının doğruluk değeri 0.717, kesinlik değeri 0.720, duyarlılık değeri 0.718, özgüllük değeri 0.702 ve AUC değeri 0.786 olarak elde edilmiştir. LWL sınıflayıcısı en yüksek performansı AUC metriğinde sağlamıştır.

Tablo 4.21. Bagging topluluk öğrenme yöntemi ile LWL sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.717
Kesinlik	0.720
Duyarlılık	0.718
Özgüllük	0.702
AUC	0.786

Bagging topluluk öğrenme yönteminde Random Forest sınıflayıcısına ilişkin sınıflandırma performansları tablo 4.22.'de belirtilmiştir. Tablodaki bulgulara göre Bagging topluluk öğrenme yönteminde Random Forest algoritmasının doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.817 ve AUC değeri 0.919 olarak elde edilmiştir. Bagging topluluk öğrenme yönteminde Random Forest sınıflayıcısı en yüksek performansını AUC metriğinde sağlamıştır.

Tablo 4.22. Bagging topluluk öğrenme yöntemi ile Random Forest sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.817
AUC	0.919

Tablo 4.23.'de bagging topluluk öğrenme yönteminde REPTree sınıflayıcısına ilişkin sınıflandırma performansları belirtilmiştir. Tablodaki bulgulara göre bagging topluluk öğrenme yönteminde REPTree sınıflayıcısının doğruluk değeri 0.757, kesinlik değeri 0.762, duyarlılık değeri 0.757, özgüllük değeri 0.732 ve AUC değeri 0.832 olarak elde edilmiştir. REPTree sınıflayıcısı en yüksek performansını AUC metriğinde sağlamıştır.

Tablo 4.23. Bagging topluluk öğrenme yöntemi ile REPTree sınıflayıcısının performanslarının değerlendirilmesi

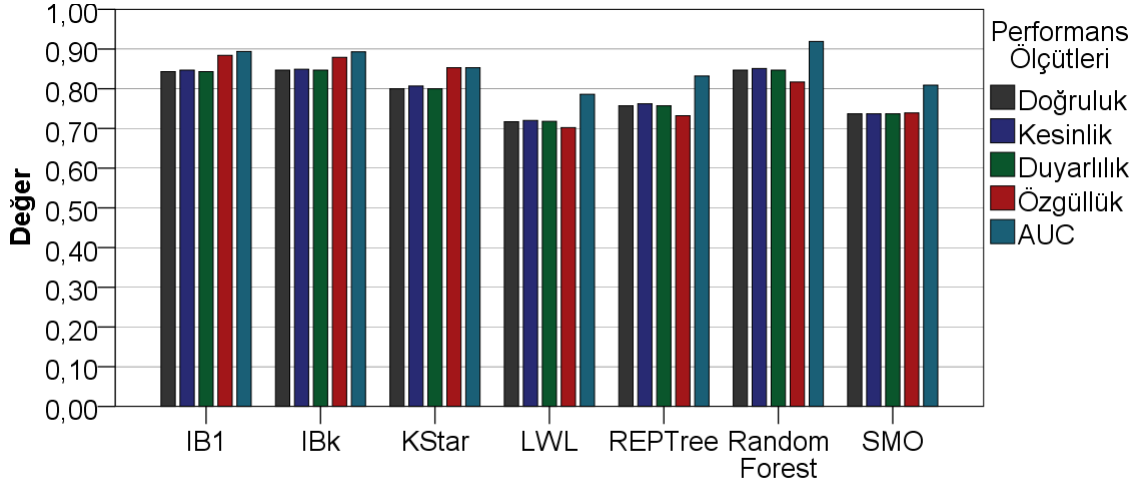
Performans ölçütü	Değer
Doğruluk	0.757
Kesinlik	0.762
Duyarlılık	0.757
Özgüllük	0.732
AUC	0.832

Bagging topluluk öğrenme yönteminde SMO sınıflayıcısına ilişkin sınıflandırma performansları tablo 4.24.'de belirtilmiştir. Tablodaki bulgulara göre SMO sınıflayıcısının doğruluk değeri 0.737, kesinlik değeri 0.737, duyarlılık değeri 0.737, özgüllük değeri 0.739 ve AUC değeri 0.809 olarak elde edilmiştir. SMO sınıflayıcısı en yüksek performansı AUC metriğinde sağlamıştır.

Tablo 4.24. Bagging topluluk öğrenme yöntemi ile SMO sınıflayıcısının performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.737
Kesinlik	0.737
Duyarlılık	0.737
Özgüllük	0.739
AUC	0.809

Bagging topluluk öğrenme yöntemlerinde doğruluk metriği açısından en yüksek sınıflandırma performansı IBk ve Random Forest sınıflayıcısı sağlamıştır. Kesinlik metriği açısından en yüksek sınıflandırma performansını Random Forest sınıflayıcısı sağlamıştır. IBk ve Random Forest sınıflayıcıları duyarlılık metriğinde yine en yüksek performansı gösteren sınıflayıcılar olduğu gözlemlenmiştir. IB1 sınıflayıcısı özgüllük metriğinde en yüksek sınıflandırma performansına sahip sınıflayıcı olduğu gözlemlenmiştir. AUC metriğinde ise en yüksek sınıflandırma performansını Random Forest sınıflayıcısı sağlamıştır. Bagging topluluk öğrenme yönteminde sınıflayıcıların performansları şekil 4.3.'de belirtilmiştir.



Bagging Topluluk Öğrenme Yöntemi ile Sınıflayıcılar

Şekil 4.3. Bagging topluluk öğrenme yönteminde sınıflayıcıların performansları

4.4. Voting Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi

Voting topluluk öğrenme yönteminde farklı sınıflayıcılardan oluşan 7 model oluşturulmuştur. Model 1, IB1 sınıflayıcısının voting topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 2, IB1 ve IBk sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 3, IB1, IBk ve KStar sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 4, IB1, IBk, KStar ve LWL sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki sınıflandırma performanslarını değerlendirmek üzere oluşturulmuştur. Model 5, IB1, IBk, KStar, LWL ve Random Forest sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki sınıflandırma performanslarını değerlendirmek üzere oluşturulmuştur. Model 6, IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki performanslarını değerlendirmek üzere oluşturulmuştur. Model 7, IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcılarının birlikte voting topluluk öğrenme yöntemindeki performanslarını değerlendirmek üzere oluşturulmuştur.

Tablo 4.25.'te voting topluluk öğrenme yönteminde model 1'in sınıflandırma performanslarına ilişkin bulgular belirtilmiştir. Model 1'in voting topluluk öğrenme yöntemindeki doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847,

özgüllük değeri 0.891 ve AUC değeri 0.849 olarak elde edilmiştir. Model 1 en yüksek performansı özgüllük metriğinde sağlamıştır.

Tablo 4.25. Voting topluluk öğrenme yöntemi ile model 1'in sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.849

Model 2'nin voting topluluk öğrenme yöntemindeki sınıflandırma performanslarına dair bulguları tablo 4.26.'da belirtilmiştir. Bulgulara göre model 2'nin doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.849 olarak elde edilmiştir. Model 2, en yüksek sınıflandırma performansını özgüllük metriğinde sağlamıştır.

Tablo 4.26. Voting topluluk öğrenme yöntemi ile model 2'nin sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.849

Voting topluluk öğrenme yönteminde model 3'ün sınıflandırma performanslarına ilişkin bulgular tablo 4.27.'de belirtilmiştir. Tablodaki bulgulara göre model 3'ün doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.942 olarak elde edilmiştir. Model 3, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.27. Voting topluluk öğrenme yöntemi ile model 3'ün sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.942

Tablo 4.28.'de voting topluluk öğrenme yönteminde model 4'ün sınıflandırma performanslarına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre model 4'ün doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.922 olarak elde edilmiştir. Model 4 en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.28. Voting topluluk öğrenme yöntemi ile model 4'ün sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.922

Voting topluluk öğrenme yöntemlerinde model 5'in sınıflandırma performanslarına ilişkin bulgular tablo 4.29.'da belirtilmiştir. Tablodaki bulgulara göre Model 5'in doğruluk değeri 0.854, kesinlik değeri 0.861, duyarlılık değeri 0.855, özgüllük değeri 0.907 ve AUC değeri 0.934 olarak elde edilmiştir. Model 5 en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.29. Voting topluluk öğrenme yöntemi ile model 5'in sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.854
Kesinlik	0.861
Duyarlılık	0.855
Özgüllük	0.907
AUC	0.934

Tablo 4.30.'da voting topluluk öğrenme yönteminde model 6'nın sınıflandırma performanslarına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre model 6'nın doğruluk değeri 0.870, kesinlik değeri 0.874, duyarlılık değeri, 0.871, özgüllük değeri 0.909 ve AUC değeri 0.928 olarak elde edilmiştir. Model 6, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.30. Voting topluluk öğrenme yöntemi ile model 6'nın sınıflandırma performanslarının değerlendirilmesi

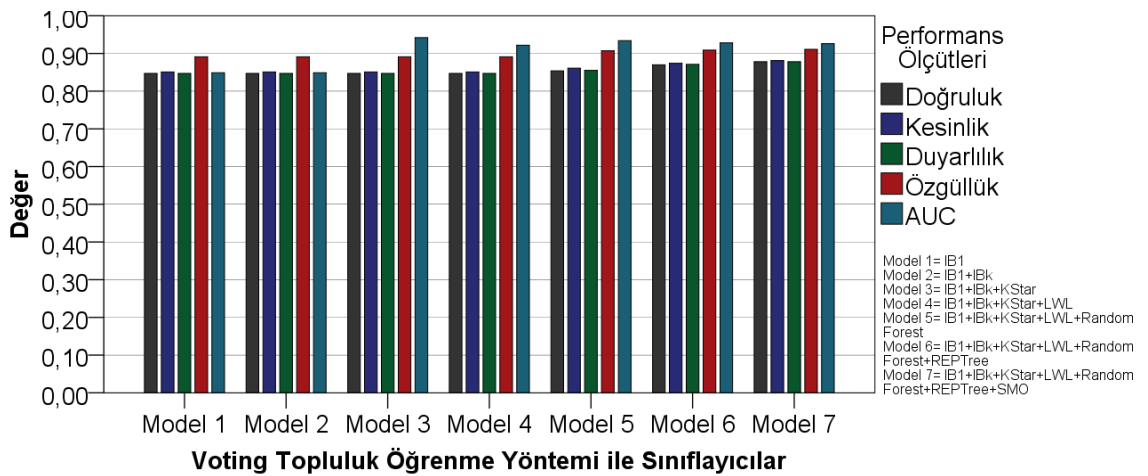
Performans ölçütü	Değer
Doğruluk	0.870
Kesinlik	0.874
Duyarlılık	0.871
Özgüllük	0.909
AUC	0.928

Voting topluluk öğrenme yönteminde model 7'nin sınıflandırma performansına ilişkin bulgular tablo 4.31.'de belirtilmiştir. Tablodaki bulgulara göre model 7'nin doğruluk değeri 0.878, kesinlik değeri 0.881, duyarlılık değeri 0.878, özgüllük değeri 0.909 ve AUC değeri 0.926 olarak elde edilmiştir. Model 7, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.31. Voting topluluk öğrenme yöntemi ile model 7'nin sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.878
Kesinlik	0.881
Duyarlılık	0.878
Özgüllük	0.909
AUC	0.926

Voting topluluk öğrenme yönteminde en yüksek doğruluk değeri, en yüksek kesinlik değeri, en yüksek duyarlılık değeri ve en yüksek özgüllük değeri model 7 tarafından sağlanmıştır. En yüksek AUC değeri ise model 3'te elde edilmiştir. Model performanslarına ilişkin bulgular şekil 4.4.'te belirtilmiştir.



Şekil 4.4. Voting topluluk öğrenme yönteminde modellerin sınıflandırma performansları

4.5. Stacking Topluluk Öğrenme Yöntemi ile Sınıflayıcıların Performanslarının Değerlendirilmesi

Stacking topluluk öğrenme yönteminde farklı modeller oluşturularak sınıflayıcılar modellere dahil edilmiştir. Stacking topluluk öğrenme yönteminde model 1, IB1 sınıflayıcısının stacking topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 2, IB1 ve IBk sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 3, IB1, IBk ve KStar sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki sınıflandırma performansını değerlendirmek üzere oluşturulmuştur. Model 4, IB1, IBk, KStar ve LWL sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki sınıflandırma performanslarını değerlendirmek üzere oluşturulmuştur. Model 5, IB1, IBk, KStar, LWL ve Random Forest sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki sınıflandırma performanslarını değerlendirmek üzere oluşturulmuştur. Model 6, IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki performanslarını değerlendirmek üzere oluşturulmuştur. Model 7, IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcılarının birlikte stacking topluluk öğrenme yöntemindeki performanslarını değerlendirmek üzere oluşturulmuştur. Stacking topluluk öğrenme yönteminde modeldeki sınıflayıcılardan elde edilen tahminlerden daha güçlü bir tahmin geliştirilmesi amacıyla “logistic” sınıflayıcısı meta sınıflayıcı olarak belirlenmiştir.

Tablo 4.32.’de stacking topluluk öğrenme yönteminde model 1’in sınıflandırma performanslarına ilişkin bulgular belirtilmiştir. Model 1’in sınıflandırma bulgularına göre doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.835 olarak elde edilmiştir. Model 1, en yüksek sınıflandırma performansını özgüllük metriğinde sağlamıştır.

Tablo 4.32. Stacking topluluk öğrenme yöntemi ile model 1’in sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.835

Stacking topluluk öğrenme yönteminde model 2'nin sınıflandırma performansına ilişkin bulgular tablo 4.33.'de belirtilmiştir. Tablodaki bulgulara göre model 2'nin doğruluk değeri 0.847, kesinlik değeri 0.851, duyarlılık değeri 0.847, özgüllük değeri 0.891 ve AUC değeri 0.835 olarak elde edilmiştir. Model 2, en yüksek sınıflandırma performansını özgüllük metriğinde sağlamıştır.

Tablo 4.33. Stacking topluluk öğrenme yöntemi ile model 2'nin sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.847
Kesinlik	0.851
Duyarlılık	0.847
Özgüllük	0.891
AUC	0.835

Tablo 4.34.'de stacking topluluk öğrenme yönteminde model 3'ün sınıflandırma performansına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre model 3'ün doğruluk değeri 0.858, kesinlik değeri 0.860, duyarlılık değeri 0.859, özgüllük değeri 0.881 ve AUC değeri 0.835 olarak ele edilmiştir. Model 3, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.34. Stacking topluluk öğrenme yöntemi ile model 3'ün sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.858
Kesinlik	0.860
Duyarlılık	0.859
Özgüllük	0.881
AUC	0.909

Stacking topluluk öğrenme yönteminde model 4'ün sınıflandırma performansına ilişkin bulgular tablo 4.35.'te belirtilmiştir. Tablodaki bulgulara göre model 4'ün doğruluk değeri 0.863, kesinlik değeri 0.864, duyarlılık değeri 0.863, özgüllük değeri 0.883 ve AUC değeri 0.905 olarak elde edilmiştir. Model 4, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.35. Stacking topluluk öğrenme yöntemi ile model 4'ün sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.863
Kesinlik	0.864
Duyarlılık	0.863
Özgüllük	0.883
AUC	0.905

Tablo 4.36.'da stacking topluluk öğrenme yönteminde model 5'in sınıflandırma performansına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre Model 5'in doğruluk değeri 0.902, kesinlik değeri 0.902, duyarlılık değeri 0.902, özgüllük değeri 0.909 ve AUC değeri 0.944 olarak elde edilmiştir. Model 5, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.36. Stacking topluluk öğrenme yöntemi ile model 5'in sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.902
Kesinlik	0.902
Duyarlılık	0.902
Özgüllük	0.909
AUC	0.944

Stacking topluluk öğrenme yönteminde model 6'nın sınıflandırma performanslarına ilişkin bulgular tablo 4.37.'de belirtilmiştir. Tablodaki bulgulara göre stacking topluluk öğrenme yönteminde model 6'nın doğruluk değeri 0.906, kesinlik değeri 0.906, duyarlılık değeri 0.906, özgüllük değeri 0.910 ve AUC değeri 0.944 olarak elde edilmiştir. Model 6, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.37. Stacking topluluk öğrenme yöntemi ile model 6'nın sınıflandırma performanslarının değerlendirilmesi

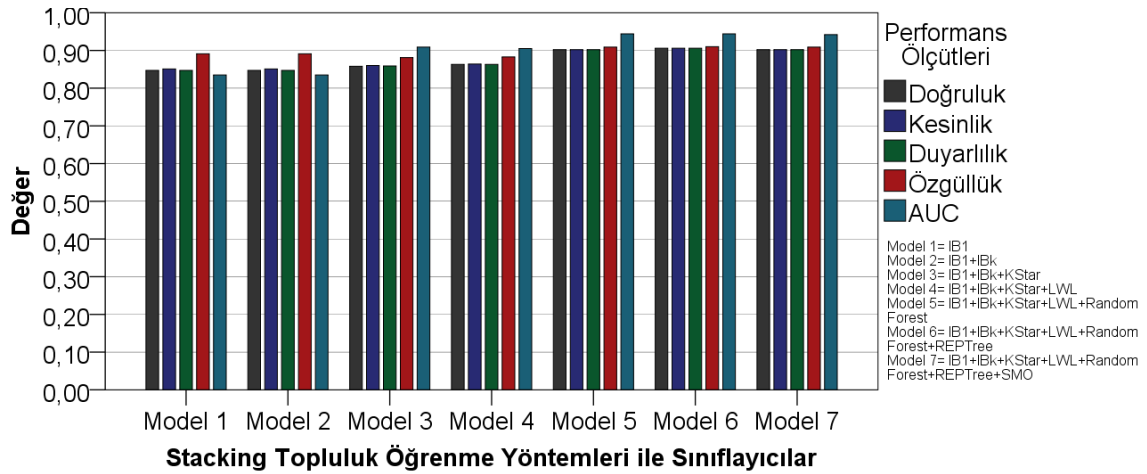
Performans ölçütü	Değer
Doğruluk	0.906
Kesinlik	0.906
Duyarlılık	0.906
Özgüllük	0.910
AUC	0.944

Tablo 4.38.'de stacking topluluk öğrenme yönteminde model 7'nin sınıflandırma performansına ilişkin bulgular belirtilmiştir. Tablodaki bulgulara göre model 7'nin doğruluk değeri 0.902, kesinlik değeri 0.902, duyarlılık değeri 0.902, özgüllük değeri 0.909 ve AUC değeri 0.942 olarak elde edilmiştir. Model 7, en yüksek sınıflandırma performansını AUC metriğinde sağlamıştır.

Tablo 4.38. Stacking topluluk öğrenme yöntemi ile model 7'nin sınıflandırma performanslarının değerlendirilmesi

Performans ölçütü	Değer
Doğruluk	0.902
Kesinlik	0.902
Duyarlılık	0.902
Özgüllük	0.909
AUC	0.942

Stacking topluluk öğrenme yönteminde en yüksek doğruluk değeri, en yüksek kesinlik değeri, en yüksek duyarlılık değeri, en yüksek özgüllük değeri model 6'daki sınıflayıcılar tarafından sağlanmıştır. AUC metriğinde ise en yüksek performansı model 5 ve model 6'daki sınıflayıcılar sağlamıştır. Stacking topluluk öğrenme yöntemindeki modellerin sınıflandırma performansları şekil 4.5.'te belirtilmiştir.



Şekil 4.5. Stacking topluluk öğrenme yönteminde modellerin sınıflandırma performansları

4.6. Topluluk Öğrenme Yöntemlerinin Sınıflandırma Performanslarının Değerlendirilmesi

Renal hücreli karsinomların sınıflandırılmasında farklı sınıflayıcılar ve farklı topluluk öğrenme yöntemleri uygulanarak en iyi performansı sağlayan model

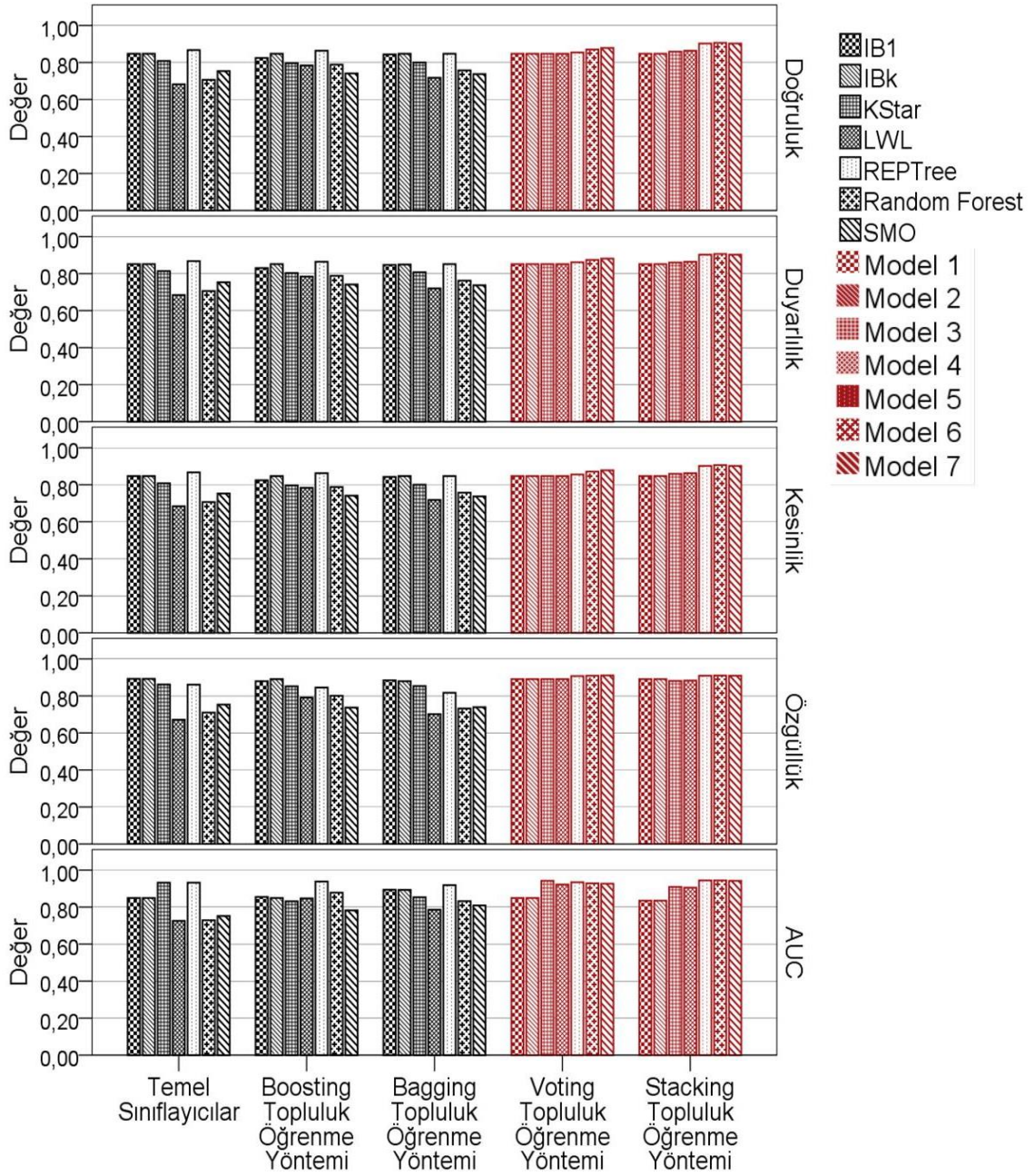
belirlenmesi sağlanmıştır. Topluluk öğrenme yöntemlerinin farklı metriklerdeki sınıflandırma performansları tablo 4.39.'da belirtilmiştir.

Tablo 4.39. Topluluk öğrenme yöntemlerine ilişkin sınıflandırma performanslarının karşılaştırılması

		Performans ölçütleri				
		Doğruluk	Duyarlılık	Kesinlik	Özgüllük	AUC
Temel sınıflayıcılar	IB1	0.847	0.851	0.847	0.892	0.849
	IBK	0.847	0.851	0.847	0.892	0.849
	KStar	0.808	0.814	0.808	0.862	0.932
	LWL	0.682	0.684	0.684	0.671	0.726
	REPTree	0.867	0.867	0.867	0.861	0.932
	Random Forest	0.706	0.706	0.706	0.710	0.729
	SMO	0.753	0.753	0.753	0.753	0.752
Boosting topluluk öğrenme yöntemi	IB1	0.824	0.830	0.824	0.879	0.855
	IBK	0.847	0.851	0.847	0.891	0.849
	KStar	0.796	0.803	0.796	0.852	0.831
	LWL	0.784	0.784	0.784	0.791	0.847
	REPTree	0.863	0.864	0.863	0.845	0.938
	Random Forest	0.788	0.788	0.788	0.801	0.878
	SMO	0.741	0.741	0.741	0.737	0.782
Bagging topluluk öğrenme yöntemi	IB1	0.843	0.847	0.843	0.884	0.894
	IBK	0.847	0.849	0.847	0.879	0.893
	KStar	0.800	0.807	0.800	0.853	0.853
	LWL	0.717	0.720	0.718	0.702	0.786
	REPTree	0.847	0.851	0.847	0.817	0.919
	Random Forest	0.757	0.762	0.757	0.732	0.832
	SMO	0.737	0.737	0.737	0.739	0.809
Voting topluluk öğrenme yöntemi	Model 1	0.847	0.851	0.847	0.891	0.849
	Model 2	0.847	0.851	0.847	0.891	0.849
	Model 3	0.847	0.851	0.847	0.891	0.942
	Model 4	0.847	0.851	0.847	0.891	0.922
	Model 5	0.854	0.861	0.855	0.907	0.934
	Model 6	0.870	0.874	0.871	0.909	0.928
	Model 7	0.878	0.881	0.878	0.911	0.926
Stacking topluluk öğrenme yöntemi	Model 1	0.847	0.851	0.847	0.891	0.835
	Model 2	0.847	0.851	0.847	0.891	0.835
	Model 3	0.858	0.860	0.859	0.881	0.909
	Model 4	0.863	0.864	0.863	0.883	0.905
	Model 5	0.902	0.902	0.902	0.909	0.944
	Model 6	0.906	0.906	0.906	0.910	0.944
	Model 7	0.902	0.902	0.902	0.909	0.942

Tabloda, IB1, IBk, KStar, LWL, Random Forest, REPTree ve SMO sınıflayıcılarının temel sınıflayıcı performansları, boosting topluluk öğrenme yöntemindeki performansları ve bagging topluluk öğrenme yöntemindeki performanslarının yanısıra voting topluluk öğrenme yönteminde ve stacking topluluk

öğrenme yönteminde uygulanan model 1 (IB1), model 2 (IB1+IBk), model 3 (IB1+IBk+Kstar), model 4 (IB1+IBk+KStar+LWL), model 5 (IB1+IBk+KStar+LWL+ Random Forest), model 6 (IB1+IBk+KStar+LWL+ Random Forest +REPTree) ve model 7'nin (IB1+ IBk+ KStar+ LWL+ Random Forest + REPTree+ SMO) farklı metriklerdeki performansları ifade edilmiştir.



Şekil 4.6. Topluluk öğrenme yöntemlerinin sınıflandırma performansları

Bulgulara göre doğruluk, duyarlılık, kesinlik, özgüllük ve AUC metriklerinde en yüksek performansı stacking topluluk öğrenme yöntemi sağlamıştır. Model 6'da en yüksek sınıflandırma performansını sağlayan stacking topluluk öğrenme yöntemi, 0.906 doğruluk değeri, 0.906 duyarlılık değeri, 0.906 kesinlik değeri, 0.910 özgüllük değeri ve 0.944 AUC değerlerini elde etmiştir. Tüm sınıflayıcı ve topluluk yöntemlerinin sınıflandırma performansları şekil 4.6.'da ifade edilmiştir.



5.TARTIŞMA

Renal hücreli karsinomlar son yıllarda insidansı artış gösteren hastalıklardan biridir. Böbrekte gözlenen malignitelerin yaklaşık % 85'inden sorumludur (96). Gelişmiş görüntüleme sistemleri kullanılarak tanı koyma ve erken teşhis ile tümör daha erken evrelerde tespit edilebilmektedir (97, 98). Teşhis yöntemlerinin gelişmesine ve kullanımının artmasına karşın mortalitede istenilen ölçüde azalma sağlanamamıştır (99, 100). Kanser hastalıklarının erken teşhis edilmesi, tam kür olasılığının artmasına katkı sağlamaktadır. Hastanın yaşam süresinin uzaması ve yaşam kalitesinin artmasında erken teşhisin katkısı önemlidir. Bu amaçla erken teşhis için, farkındalık bilincinin artırılması ve toplum sağlık taramalarının yaygınlaştırılması önerilmektedir (101, 102). Kanser hastalıklarının erken teşhisinde toplum farkındalığının yanısıra teşhiste kullanılan teknik ve cihazların geliştirilmesi de önemlidir. Bazı kanser türlerinde erken evrede fiziki muayene ve radyolojik görüntüleme kanser gelişimine ilişkin belirtiler saptanamamaktadır. Bu nedenle kanserin erken teşhisinde, sadece fiziki muayene, biyopsi, hemogram değerleri ve görüntüleme yöntemlerinin yanısıra özel tekniklerde kullanılabilir. Bu özelliklerin başında son yıllarda sağlık alanında başarı ile uygulanabilen makine öğrenme yöntemleri ve yapay zeka sistemleri gelmektedir. Kanserlerin erken teşhisinde yapay zeka veya makine öğrenme yöntemlerinin uygulandığı çok sayıda araştırma mevcuttur (103-106). Makine öğrenme yöntemleri, kanser hastalıklarında erken teşhis için verilerden özelliklerine göre sınıflandırma gerçekleştirerek, veriler arasındaki örüntüyü öğrenmektedir. Makine öğrenme yönteminin örüntüyü öğrenmesi ile yeni gelecek her veri için bir sınıflandırma tahmini gerçekleştirebilmektedir. Renal hücreli karsinomlarda erken evrelerde kanser nadiren fiziki bir belirti göstermektedir (107). Renal hücreli karsinomun ilk dönemlerinde radyolojik ve ultrasonografik görüntüleme sistemleri ile gözlenemeyen tümör başlangıçları olabilmektedir. Bu sebeplerden dolayı renal hücreli karsinomun erken evrelerinde teşhisi için, makine öğrenme yöntemlerinin kullanımı ile yararlı sonuçlar sağlanmaktadır.

Makine öğrenme yöntemlerinin hastalıkların tahmininde kullanılması son yıllarda yaygınlaşan bir uygulama alanı olmuştur. Özellikle kanser hastalıklarının tahmininde makine öğrenme yöntemleri sıklıkla kullanılmaktadır (108-111). Makine öğrenme yöntemleri kanser hastalıklarının tahmin edilmesinde yüksek performans

sağlayabilmektedir. Yapay zeka ve makine öğrenmenin gelecekte hastalıkların tahmininde kullanımının artacağına, sağlık alanında birçok noktada yapay zeka yöntemlerinden yararlanılacağına dair öngörüler bulunmaktadır (112). Makine öğrenme yöntemlerinde hastalıkların sınıflandırılması ve tahmin edilmesinde başarılı tahminlerde bulunmasının farklı unsurları bulunmaktadır. Veri tabanlarında çok sayıda hastanın verilerinin bulunması, bu hastalıkların makine öğrenme yöntemleri ile sınıflandırılmasına olanak tanımaktadır. Araştırmalarda çoğunlukla toplumda sık görülen kanser türleri makine öğrenme yöntemleri ile tahmin edilmektedir. Göğüs kanseri, cilt kanseri, akciğer kanseri, prostat kanseri, mide kanseri, kolon kanseri ve servikal kanserler toplumda en fazla gözlenen kanser türlerindedir (113). Belirtilen bu kanser türlerinde veri kayıtlarının yüksek sayıda olması ve tahminde kullanılan değişkenlerin modeli iyi açıklaması, makine öğrenme yöntemlerinde sıklıkla bu kanser türleri üzerine çalışan araştırmacılar için tercih sebebi olmuştur. Göğüs kanseri, cilt kanseri ve akciğer kanseri üzerine makine öğrenme yöntemleri ile tahmin gerçekleştiren çok sayıda araştırma mevcuttur (106, 114-119). Toplumda prevalansı düşük olan kanser türlerine ilişkin makine öğrenme yöntemi kullanılarak yapılan sınıflandırma ve tahmin araştırmalarının az sayıda olduğu gözlemlenmiştir. Bu durumun temel nedenlerinden biride, prevalansı düşük olan kanser türlerinde yeterli hasta sayısının olmaması ve veriye ulaşamamaktan kaynaklanmaktadır. Makine öğrenme yöntemleri az sayıda örnekte, eğitim veri setindeki örüntüyü yeterince öğrenememesinden dolayı, düşük sınıflandırma performansı gösterebilmektedir.

Topluluk öğrenme yöntemleri, eğitim veri setini bir sınıflayıcı ile eğitmek yerine birden fazla sayıda sınıflayıcıyı modele alarak eğitmektedir. Birleştirme aşamasında, her bir sınıflayıcının tahminleri elde edildikten sonra sınıflayıcıların tahminleri birleştirilmektedir. Tahminlerdeki hata değeri düştüğü için, topluluk öğrenmede temel sınıflayıcılara göre daha yüksek performans elde edilebilmektedir (26). Toplumda görülme prevalansı düşük olan kanser türlerinin tahmin edilmesinde ve sınıflandırılmasında, topluluk öğrenme yöntemleri daha yüksek performans sağlayabilmektedir.

Bu çalışmada topluluk öğrenme yöntemleri kullanılarak renal hücreli karsinomun tahmin edilmesi amaçlanmıştır. Topluluk öğrenme modeline IB1, IBk, KStar, LWL, Random Forest, REPTree, SMO olmak üzere 7 sınıflayıcı dahil edilmiştir. Çalışmada topluluk öğrenme yöntemlerinin yanısıra temel sınıflayıcıların performansları da değerlendirilmiştir. Temel sınıflayıcılarda en yüksek performansı

REPTree sınıflayıcısı elde etmiştir. REPTree sınıflayıcısının doğruluk değeri 0.867 ve AUC değeri 0.932 olarak elde edilmiştir. Topluluk öğrenme yöntemleri içerisinde en yüksek performansı, stacking topluluk öğrenme yöntemlerinde IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcıların topluluğa dahil edildiği model 6 sağlamıştır. Modelde doğruluk değeri 0.906 ve AUC değeri 0.944 olarak elde edilmiştir. Renal hücreli karsinomların tahmin edilmesine yönelik yapılan benzer araştırmalardaki bulgular değerlendirilmiştir. Temel sınıflayıcılarda Jagga ve Gupta çalışmalarında renal hücreli karsinomların sınıflandırılmasında J48, Random Forest, Naive Bayes ve destek vektör makinesi sınıflayıcılarından yararlanmıştır. Sınıflama sonucuna göre en yüksek performansı Random Forest sınıflayıcısı 0.77 doğruluk değeri ile sağlamıştır. AUC değeri ise 0.80 olarak elde edilmiştir (120). Sing ve arkadaşları papiller renal hücreli karsinomların ileri dönem ve geç dönem ilerleme durumlarını tahmin etmek üzere Naive Bayes, Random Forest, k- en yakın komşu sınıflayıcısı ve destek vektör makinesini kullanmıştır. En yüksek sınıflandırma performansını, Random Forest sınıflayıcısı 0.885 değeri ile doğruluk ve 0.804 değeri ile AUC metriklerinde sağlamışlardır (121). Bektaş ve arkadaşları çalışmalarında bilgisayarlı tomografiden elde ettikleri görüntüleri berrak renal hücreli karsinomların sınıflandırılmasında k-en yakın komşu, Naive bayes, Random Forest, yapay sinir ağları sınıflayıcılarını kullanmışlardır. En yüksek sınıflandırma performansı doğruluk değeri 0.851 ve AUC değeri 0.860 olarak elde etmişlerdir (122). Koçak ve arkadaşları yapay sinir ağları ile renal hücreli karsinomlu hastaların tomografi görüntülerini sınıflandırmıştır. Sınıflandırmada doğruluk değeri 0.692 olarak elde etmişlerdir (123). Çalışmamızda elde edilen bulgulara göre temel sınıflayıcılarda doğruluk değeri ve AUC değerleri literatüre benzerlik göstermektedir. Literatürdeki araştırmalarda temel sınıflayıcılarda genellikle Random Forest algoritması en yüksek doğruluk değerini sunarken, çalışmamızda en yüksek doğruluk değerini REPTree sınıflayıcısı sağlamıştır. Çalışmamızda stacking topluluk öğrenme yönteminde doğruluk değeri 0.906 ve AUC değeri 0.944 olarak elde edilmiştir. Stacking topluluk öğrenme yöntemi, renal hücreli karsinomların tahmin edilmesine yönelik literatürdeki çalışmalardan daha yüksek sınıflandırma performansı sergilemiştir. Stacking topluluk öğrenme yöntemi temel sınıflayıcılara göre daha yüksek sınıflandırma performansı göstermiştir.

Topluluk öğrenme yöntemleri, doğru sınıflayıcılar ve birleştirme teknikleri kullanıldıkları takdirde oldukça yüksek sınıflandırma performansı sunmaktadır. Topluluk öğrenme yöntemleri kanser hastalıklarının tahmininde oldukça başarılı

sonular elde etmektedir. Mohebian ve arkadaşları alıřmalarında gėős kanserinin tahmin edilmesinde optimize edilmiř topluluk ėrenme yntemini nermiřlerdir (124). nerdikleri bu modelde destek vektr makinesi, karar aėacı ve ok katmanlı algılayıcı sınıflayıcı ile topluluk ėrenme model oluřturulmuřtur. Modelin sınıflandırma performansı 0.892 doėruluk deėerine sahiptir. Hsieh ve arkadaşları alıřmalarında gėős kanserinin tahmini iin topluluk ėrenme modelinden yararlanmıřlardır. Modelde neural fuzzy, k-en yakın komřu algoritması ve kuadratik sınıflayıcılar dahil edilmiřtir. modeldeki topluluk ėrenme yntemlerinin doėruluk deėeri 0.679 olarak elde edilmiřtir (28). Cai ve arkadaşları alıřmalarında akciėer kanserinin sınıflandırılmasında topluluk yntemlerinden yararlanılmıřtır. alıřmalarında doėruluk deėerini 0.865 olarak elde etmiřlerdir (125). alıřmamızda stacking topluluk ėrenme yntemi ile doėruluk deėeri 0.906 ve AUC deėeri 0.944 olarak belirlenmiřtir. Bulgular literatürdeki bulguları desteklemek ile birlikte literatürdeki doėruluk deėerlerinden daha yüksek bir doėruluk deėeri iermektedir. Modelimiz oldukça bařarılı tahminler gerekleřtirmiřtir.

Topluluk ėrenme yntemleri, tek sınıflayıcının tahmin performansından ziyade birden fazla sınıflayıcının tahmin performanslarını birleřtirerek daha yüksek performansa sahip sınıflandırma gerekleřtirebilecek model oluřturmayı amalamaktadır. Veri setine uygun sınıflayıcı ve doėru birleřtirme teknikleri seildiėinde, sınıflandırma hatasını dřürmek ve doėruluk performansını artırmak mmkündür. alıřmamızdaki bulgular bu bilgileri doėrular niteliktedir. Temel sınıflayıcılar ile elde edilen en yüksek doėruluk deėeri 0.867 iken, topluluk ėrenme yntemi ile elde edilen en yüksek doėruluk deėer 0.906 olarak elde edilmiřtir. alıřmanın bulgularında temel sınıflayıcılarda en yüksek tahmin performansı REPTree sınıflayıcısında elde edilmiřtir. REPTree sınıflayıcısında doėruluk deėeri, 0.867, duyarlılık deėeri, 0.867, kesinlik deėeri 0.867, zgllük deėeri 0.861 ve AUC deėeri 0.932 olarak elde edilmiřtir. Bagging topluluk ėrenme ynteminde en yüksek sınıflandırma performansı yine REPTree sınıflayıcısı ile elde edilmiřtir. Boosting topluluk ėrenme ynteminde en yüksek sınıflandırma performansını REPTree sınıflayıcısı ile saėlanmıřtır. Voting topluluk ėrenme ynteminde en yüksek sınıflandırma performansı Model 7 de elde edilmiřtir. Model 7'nin doėruluk deėeri 0.878, AUC deėeri 0.926 olarak elde edilmiřtir. Son olarak stacking topluluk ėrenme ynteminde en yüksek sınıflandırma performansı model 6 ile gerekleřtirilmiřtir. Model 6'nın doėruluk deėeri 0.906, AUC deėeri 0.944 olarak elde edilmiřtir. Bu bilgiler eřliėinde en yüksek sınıflandırma performansını stacking topluluk ėrenme yntemi

sağlamıştır. Stacking topluluk öğrenme yöntemi temel sınıflayıcılara ve diğer topluluk öğrenme yöntemlerine göre daha yüksek performans sergilemiştir.

Çalışmanın temel amacı olan topluluk öğrenme yöntemi ile renal hücreli karsinom hastalığı başarılı bir şekilde tahmin edilmiştir. Modelin sınıflandırmadaki doğruluk değeri 0.906, duyarlılık değeri 0.906, kesinlik değeri 0.906, özgüllük değeri 0.910 ve AUC değeri 0.944 olarak elde edilmiştir.

Renal hücreli karsinomların tahmin edilmesinde güçlü bir topluluk öğrenme modeli oluşturulmuştur. Oluşturulan topluluk öğrenme modeli, stacking topluluk öğrenme yönteminde IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcıların birleştirilmesinden elde edilmiştir.

Renal hücreli karsinom hastalığının toplumda görülme prevalansının düşük olması nedeniyle araştırmacılar bu alanda makine öğrenme yöntemleri ile az sayıda çalışma gerçekleştirmiştir. Yapılan az sayıda çalışmada ise çoğunlukla renal hücreli karsinomlarda gen ekspresyonu ve radyolojik görüntülerin işlenmesi üzerine uygulanmıştır. Az sayıda çalışma gerçekleştirilmesinin sebeplerinden biri de yeterli örnek büyüklüğüne ulaşamama, laboratuvar ve hemogram değerleri ile yüksek performansa sahip sınıflandırma gerçekleştirilememesidir.

Bu çalışmada laboratuvar ve hemogram bulguları ile renal hücreli karsinom hastalığı tahmin edilmiştir. Sınıflandırmada yüksek performans sağlanmıştır. Renal hücreli karsinomların tahmin edilmesinde stacking topluluk öğrenme yöntemini uygulayan literatür bilgisine rastlanmamıştır. Çalışma, bu anlamda bilime büyük katkı ve yenilik katacaktır. IB1, IBk, KStar, LWL, Random Forest ve REPTree sınıflayıcılarının stacking topluluk öğrenme yöntemi ile oluşturulan modelin yüksek sınıflama performansına sahip olduğu gözlemlenmiştir. Bu modelin farklı veri setlerinde uygulanması yüksek sınıflama performansı sağlanmasında önemli katkı sağlayacaktır.

6.SONUÇ VE ÖNERİLER

Çalışmada topluluk öğrenme yöntemleri kullanılarak renal hücreli karsinom hastalığı tahmin edilmiştir. Modelde bagging topluluk öğrenme yöntemi, boosting topluluk öğrenme yöntemi, voting topluluk öğrenme yöntemi ve stacking topluluk öğrenme yönteminin yanısıra temel sınıflayıcıların performansları da değerlendirilmiştir. Topluluk yöntemi ve sınıflayıcıların performansları doğruluk, duyarlılık, kesinlik, özgüllük ve AUC metriklerinde değerlendirilmiştir.

En yüksek sınıflandırma performansını stacking topluluk öğrenme yönteminde IB1, IBk, Kstar, LWL, Random Forest ve REPTree sınıflayıcılarından oluşan model sağlamıştır. Stacking topluluk öğrenme yöntemi renal hücreli karsinomları başarılı bir şekilde tahmin etmiştir. Stacking topluluk öğrenme yöntemi ile sınıflandırmada % 90.6 doğruluk, % 94.4 oranında da AUC değerleri elde edilmiştir.

Topluluk öğrenme yöntemleri renal hücreli karsinomları tahmin etmede temel sınıflayıcılara göre daha yüksek performans elde etmektedir. Stacking topluluk öğrenme yöntemleri, boosting topluluk öğrenme yöntemi ve bagging topluluk öğrenme yöntemlerinden daha yüksek sınıflandırma performansı sağlamıştır. Bagging topluluk öğrenme yöntemi ve boosting topluluk öğrenme yöntemleri birbirlerine yakın sınıflandırma performansı sağlamıştır.

Stacking topluluk öğrenme yöntemi, laboratuvar ve hemogram değerlerini içeren değişkenlerde başarılı tahminler gerçekleştirebilmektedir. Stacking topluluk öğrenme yöntemi ile yapılacak çalışmalarda sınıflandırma performansının artırılması adına modele çok sayıda tahminci değişken alınması önerilmektedir. Stacking topluluk öğrenme yönteminde meta sınıflayıcının seçimi, sınıflandırma performansını önemli derecede etkilemektedir. Veri setine ve modele uygun meta sınıflayıcının seçilmesi önerilmektedir. Topluluk öğrenme yöntemlerinde veri setindeki örnek sayısının yeterli sayıda olması, sınıflandırma performansı için önemlidir. Topluluk öğrenme yöntemleri ile yapılacak araştırmalarda veri setinde yeterli sayıda örneğin modele dahil edilmesi önerilmektedir.

Topluluk öğrenme yöntemlerinin geliştirilmesi ve daha yüksek performans ile sınıflandırma gerçekleştirebilecek yöntemler üzerine yapılacak araştırmalarda, farklı sınıflayıcıları modele dahil eden, birleştirme tekniğinde sınıflayıcıların entropilerini dikkate alan ve ağırlıklı birleştirme tekniğini esas alan, modelde karar ağacı tabanlı sınıflayıcılara yer veren yeni yöntemler üzerinde çalışılması önerilmektedir.

KAYNAKLAR

1. Güçer H. Renal Hücreli (Berrak Hücreli Tip) Karsinomlarda Cox-2 Ekspresyonu Ve Mikrodamar Yoğunluğunun Klinikopatolojik Parametrelerle Karşılaştırılması. Taksim Eğitim ve Araştırma Hastanesi, Patoloji Kliniği. Tıpta Uzmanlık Tezi, İstanbul: Taksim Eğitim ve Araştırma Hastanesi, 2006.
2. Demirkıran ED. Renal Hücreli Karsinomlarda Tümör Hacminin, Böbrek Hacmine Oranının Prognostik Faktörler ile İlişkisi. Tıp Fakültesi, Üroloji Anabilimdalı. Tıpta Uzmanlık Tezi. Zonguldak: Zonguldak Bülent Ecevit Üniversitesi, 2019.
3. Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, Gore JL, Sun M, Wood C, Russo P. Epidemiology of renal cell carcinoma. *Eur Urol* 2019, 75: 74-84.
4. Eble JN, Togashi K, Pisani P. Renal Cell Carcinoma. In : Eble JN, Sauter G, Epstein JI, Sesterhenn IA (eds). *World Health Organization Classification of Tumours. Pathology and genetics of Tumours of the urinary system and male genital organs* IARC Press, 2004: 9-87.
5. Pesch B, Haerting J, Ranft U, Klimpel A, Oelschlägel B, Schill W. Occupational risk factors for renal cell carcinoma: agent-specific results from a case-control study in Germany. *Int J Epidemiol* 2000, 29: 1014-24.
6. Campbell S, Lane B. Malignant Renal tumors. In: Wein A, Kavoussi L, Partin A, Novick A, Peters C (eds). *Campbell-Walsh Urology*, 10th ed. Philadelphia, Saunders, 2012:1421.
7. Brierley JD, Gospodarowicz MK, Wittekind C, (eds). *TNM classification of malignant tumours*. John Wiley & Sons, 2017.
8. Yaşar S. Renal Hücreli Karsinom Olgularında Evre, Nükleer Grade Ve Histolojik Subtipte Radiomics Bulguları. Kocaeli Derince Sağlık Uygulama ve Araştırma Merkezi, Radyoloji Anabilimdalı. Tıpta Uzmanlık Tezi. Kocaeli: Sağlık Bilimleri Üniversitesi, 2019.
9. Ritchie AW, Chisholm GD. The natural history of renal carcinoma. *Semin Oncol* 1983, 10: 390-400.

10. Kim HL, Belldegrun AS, Freitas DG, Bui MH, Han KR, Dorey FJ, Figlin RA. Paraneoplastic signs and symptoms of renal cell carcinoma: implications for prognosis. *J Urol* 2003, 170: 1742-6.
11. Ficarra V, Prayer-Galetti T, Novella G, Bratti E, Maffei N, Dal Bianco M, Artibani W, Pagano F. Incidental detection beyond pathological factors as prognostic predictor of renal cell carcinoma. *Eur Urol* 2003, 43: 663-9.
12. Jayson M, Sanders H. Increased incidence of serendipitously discovered renal cell carcinoma. *Urology*. 1998, 51: 203-5.
13. Taştekin E. Renal Hücreli Karsinomlarda Anjiyogenetik Ve Prognostik Faktörlerin Karşılaştırılması. Tıp Fakültesi, Tıbbi Patoloji Anabilimdalı. Tıpta Uzmanlık Tezi. Edirne: Trakya Üniversitesi, 2019.
14. Sufrin G, Chasan S, Golio A, Murphy GP. Paraneoplastic and serologic syndromes of renal adenocarcinoma. *Semin Urol* 1989, 7: 158-71.
15. Roubidoux MA, Dunnick NR, Sostman HD, Leder RA. Renal carcinoma: detection of venous extension with gradient-echo MR imaging. *Radiology* 1992, 182: 269-72.
16. Dutcher JP, de Souza P, McDermott D, Figlin RA, Berkenblit A, Thiele A, Krygowski M, Strahs A, Feingold J, Hudes G. Effect of temsirolimus versus interferon- α on outcome of patients with advanced renal cell carcinoma of different tumor histologies. *Med Oncol* 2009, 26: 202-9.
17. Cho DS, Joo HJ, Oh DK, Kang JH, Kim YS, Lee KB, Kim SJ. Cyclooxygenase-2 and p53 expression as prognostic indicators in conventional renal cell carcinoma. *Yonsei Med J* 2005, 46: 133-40.
18. Witten IH, Frank E. *Data Mining: Practical machine learning tools and techniques*. San Francisco, CA, USA, Morgan Kaufmann Publishers Inc, 2005:1-37.
19. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI magazine* 1996, 17: 37.
20. Balaban ME, Kartal E. *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları*, İstanbul, Çağlayan Kitabevi, 2015.
21. Güldoğan E. Çeşitli Çekirdek Fonksiyonları İle Oluşturulan Destek Vektör Makinesi Modellerinin Performanslarının İncelenmesi: Bir Klinik Uygulama. Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilimdalı. Doktora Tezi, Malatya: İnönü Üniversitesi, 2017.

22. Akman M, Genç Y, Ankarali H. Random forests yöntemi ve sağlık alanında bir uygulama. *Turkiye Klinikleri J Biostat* 2011, 3: 36-48.
23. Dietterich T. Overfitting and undercomputing in machine learning. *ACM Comput Surv* 1995, 27: 326-7.
24. Jabbar H, Khan DR. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, 2015.
25. Lee H, Kim J, Kim S. Gaussian-Based SMOTE Algorithm for Solving Skewed Class Distributions. *International Journal of Fuzzy Logic and Intelligent Systems* 2017, 17: 229-34.
26. Zhang C, Ma Y, (Eds). *Ensemble machine learning: methods and applications*. Springer Science, Business Media, 2012.
27. Sun S. Ensemble learning methods for classifying EEG signals. In: *International Workshop on Multiple Classifier Systems*, Berlin, Heidelberg, Springer, 2007: 113-20.
28. Hsieh SL, Hsieh SH, Cheng PH, Chen CH, Hsu KP, Lee IS, Wang Z, Lai F. Design ensemble machine learning model for breast cancer diagnosis. *J Med Syst* 2012, 36: 2841-7.
29. Polikar R. Ensemble learning. In: Zhang C, Ma Y (eds). *Ensemble machine learning*. Springer Science, Business Media, 2012: 1-34.
30. Tukey JW. *Exploratory data Analysis*. Addison Wesley, Reading, MA, 1977.
31. Rokach L. *Pattern classification using ensemble methods*. World Scientific, 2010.
32. Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 1990, 10: 993-1001.
33. Schapire RE. The strength of weak learnability. *Mach Learn* 1990, 5: 197-227.
34. Freund Y, Schapire RE. Schapire R: Experiments with a new boosting algorithm. In: *Thirteenth International Conference on ML*, 1996.
35. Zhou ZH. *Ensemble methods: foundations and algorithms*. Chapman and Hall/CRC, 2012.
36. Domingos P. Using partitioning to speed up specific-to-general rule induction. In: *Proceedings of the AAAI-96 Workshop on Integrating Multiple Learned Models*, AAAI Press, 1996: 29-34.

37. Bauer E, Kohavi R. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Mach Learn* 1999, 36: 105-39.
38. Quinlan JR. Bagging, boosting, and C4.5. In: *AAAI/IAAI*, 1996, 1: 725-30.
39. Bartlett P, Shawe-Taylor J. Generalization performance of support vector machines and other pattern classifiers. *Advances in Kernel methods—support vector learning* 1999, 43-54.
40. Zhang Y, Liu B, Cai J, Zhang S. Ensemble weighted extreme learning machine for imbalanced data classification based on differential evolution. *Neural Comput Appl* 2017, 28: 259-67.
41. Turhan S. Kolektif Öğrenmede Sınıf Dengesizliği Problemi: Hastalık Tanısı Sınıflandırma. Fen Bilimleri Enstitüsü, İstatistik Anabilimdalı. Yüksek Lisans Tezi, Muğla: Muğla Sıtkı Koçman Üniversitesi, 2019.
42. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 1998, 20: 832-44.
43. Breiman L. Random forests. *Mach Learn* 2001, 45: 5-32.
44. Wolpert DH. Stacked generalization. *Neural Netw* 1992, 5: 241-59.
45. Aha DW, Kibler D, Albert MK. Instance-based learning algorithms. *Mach Learn* 1991, 6: 37-66.
46. Weka Sourceforge. Class IB1.
<https://weka.sourceforge.io/doc.stable/weka/classifiers/lazy/IB1.html>. 21 Kasım 2019.
47. Bal M, Amasyali MF, Sever H, Kose G, Demirhan A. Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system. *The Scientific World Journal* 2014, 2014.
48. Weka Sourceforge. Class IBk.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/lazy/IBk.html>. 21 Kasım 2019.
49. Weinberger KQ, Blitzer J, Saul LK. Distance metric learning for large margin nearest neighbor classification. In: *Advances in neural information processing systems* 2006: 1473-80.
50. Atkeson CG, Moore AW, Schaal S. Locally weighted learning. In: *Lazy learning*, Springer Dordrecht, 1997: 11-73.

51. Weka Sourceforge. Class LWL.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/lazy/LWL.html>. 22 Kasım 2019.
52. Cleary JG, Trigg LE. K*: An instance-based learner using an entropic distance measure. In: *Machine Learning Proceedings*, Morgan Kaufmann, 1995: 108-14.
53. Weka Sourceforge. Class KStar.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/lazy/KStar.html>. 22 Kasım 2019.
54. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. *Comput Stat Data Anal* 2008, 52: 2249-60.
55. Akar Ö, Güngör O. Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Jeodezi ve Jeoinformasyon Dergisi* 2012: 139-46.
56. Weka Sourceforge. Class RandomForest.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/RandomForest.html>. 22 Kasım 2019.
57. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2016.
58. Quinlan JR. Simplifying decision trees. *International journal of man-machine studies* 1987, 27: 221-34.
59. Weka Sourceforge. Class REPTree.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/trees/REPTree.html>. 23 Kasım 2019.
60. Doğan Y, Ataş M. Prediction of adaptive exposure time in hyperspectral bands for industrial cameras. In: *2015 23rd Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2015: 657-60.
61. Platt JC. Fast training of support vector machines using sequential minimal optimization, *Advances in kernel methods support vector learning* 1999, 185-208.
62. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KR. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput* 2001, 13: 637-49.
63. Tan B. Belediyelerde sosyal yardım hizmetleri dağıtımına yönelik karar destek sistemi geliştirilmesi. Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilimdalı. Yüksek Lisans Tezi. Elazığ: Fırat Üniversitesi, 2018.

64. Weka Sourceforge. Class SMO.
<https://weka.sourceforge.io/doc.dev/weka/classifiers/functions/SMO.html> 24 Kasım 2019.
65. Vapnik V. *The nature of statistical learning theory*. Springer science & business media, 2013.
66. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995, 20: 273-97.
67. Nitze I, Schulthess U, Asche H. Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. In: *Proceedings of the 4th GEOBIA*, Rio de Janeiro, Brazil. 2012, 79: 3540.
68. Soman KP, Loganathan R, Ajay V. *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd, 2009.
69. Ayhan S, Erdoğan Ş. Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi* 2014, 9: 175-201.
70. Arslan A, Şen B. Detection of non-coding RNA's with optimized support vector machines. In: *2015 23rd Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2015: 1668-71.
71. Chen S, Yu P. Support vector networks puning on fllood forecasting and data mining. *J Hydrol* 2007, 347: 67-78.
72. Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Med Inform Decis Mak* 2010, 10: 16.
73. Won Y, Song HJ, Kang TW, Kim JJ, Han BD, Lee SW. Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons. *Proteomics* 2003, 3: 2310-6.
74. Liu D, Shi T, DiDonato JA, Carpten JD, Zhu J, Duan ZH. Application of genetic algorithm/k-nearest neighbor method to the classification of renal cell carcinoma. In: *Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference*, IEEE, 2004: 558-9.
75. Fuchs TJ, Wild PJ, Moch H, Buhmann JM. Computational pathology analysis of tissue microarrays predicts survival of renal clear cell carcinoma patients. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Berlin, Heidelberg, 2008: 1-8.

76. Gönen M, Ulaş A, Schüffler P, Castellani U, Murino V. Combining data sources nonlinearly for cell nucleus classification of renal cell carcinoma. In: *International Workshop on Similarity-Based Pattern Recognition*, Springer, Berlin, Heidelberg, 2011:250-60.
77. Lee HS, Hong H, Jung DC, Park S, Kim J. Differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma in contrast-enhanced MDCT images using quantitative feature classification. *Med Phys* 2017, 44: 3604-14.
78. Lin F, Cui EM, Lei Y, Luo LP. CT-based machine learning model to predict the Fuhrman nuclear grade of clear cell renal cell carcinoma. *Abdom Radiol* 2019: 1-7.
79. Tabibu S, Vinod PK, Jawahar CV. Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning. *Sci Rep* 2019, 9: 1-9.
80. Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification. *Appl Bioinformatics* 2003, 2: S75–S83.
81. Luo ST, Cheng BW. Diagnosing breast masses in digital mammography using feature selection and ensemble methods. *J Med Syst* 2012, 36: 569-77.
82. Nagi S, Bhattacharyya DK. Classification of microarray cancer data using ensemble approach. *Netw Model Anal Health Inform Bioinform* 2013, 2: 159-73.
83. Onan A. On the performance of ensemble learning for automated diagnosis of breast cancer. In: *Artificial Intelligence Perspectives and Applications*, Springer, Cham, 2015: 119-29.
84. Wang Y, Wang D, Geng N, Wang Y, Yin Y, Jin Y. Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection. *Appl Soft Comput* 2019, 77: 188-204.
85. Alpar R. *Spor, sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik*, Ankara, Detay Yayıncılık, 2016.
86. Sümbüloğlu V, Sümbüloğlu K. *Klinik saha araştırmalarında örnekleme yöntemleri ve örneklem büyüklüğü*, Ankara, Hatiboğlu Yayınevi, 2005.
87. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: identifying density-based local outliers. *ACM sigmod record* 2000, 29: 93-104.

88. Lee J, Kang B, Kang SH. Integrating independent component analysis and local outlier factor for plant-wide process monitoring. *J Process Contr* 2011, 21: 1011-21.
89. Weka Sourceforge. Class GridSearch.
<https://weka.sourceforge.io/doc.stable/weka/classifiers/meta/GridSearch.html>.
24 Kasım 2019.
90. Multisearch weka package. <https://github.com/fracpete/multisearch-weka-package>. 24 Kasım 2019.
91. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*. 2009, 11:10-8.
92. Team RC. R: A language and environment for statistical computing. *www.R-project.org*. Accessed: September 2019.
93. Witten IH, Frank E, Hall MA, Pal CJ. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann; 2016.
94. TIBCO Software Inc. Statistica (data analysis software system), version 13 Academic. 2017.
95. IBM Corp. *IBM SPSS statistics for windows, version 22.0*, Armonk, NY:IBM Corp. 2013.
96. Dirim A, Özkardeş H. Böbrek kanserlerinde insidans, etiyoloji ve sınıflandırma. *Türkiye Klinikleri J Urology-Special Topics* 2011, 4: 10-7.
97. Patard JJ, Rodriguez A, Rioux-Leclercq N, Guille F, Lobel B. Prognostic significance of the mode of detection in renal tumours. *BJU international* 2002, 90: 358-63.
98. Kato M, Suzuki T, Suzuki Y, Terasawa Y, Sasano H, Arai Y. Natural history of small renal cell carcinoma: evaluation of growth rate, histological grade, cell proliferation and apoptosis. *J Urol* 2004, 172: 863-6.
99. Chow WH, Linehan WM, Devesa SS. Re: Rising incidence of small renal masses: a need to reassess treatment effect. *Journal of the National Cancer Institute* 2007, 99: 569-70.
100. Gül Ü. Böbrek Hücreli Karsinomda Risk Faktörleri ve Etyoloji. *Bulletin of Urooncology* 2014, 13: 135-8.
101. Richards MA. The national awareness and early diagnosis initiative in England: assembling the evidence. *Br J Cancer* 2009, 101: S1.

102. Wardle J, Robb K, Vernon S, Waller J. Screening for prevention and early diagnosis of cancer. *Am Psychol* 2015, 70: 119.
103. Ng EK, Fok SC, Peh YC, Ng FC, Sim LS. Computerized detection of breast cancer with artificial intelligence and thermograms. *J Med Eng Technol* 2002, 26: 152-7.
104. Zubi ZS, Saad RA. Using some data mining techniques for early diagnosis of lung cancer. In: *Proceedings of the 10th WSEAS international conference on Artificial intelligence Knowledge Engineering and Data Bases* 2011: 32-7.
105. Giger ML. Computerized analysis of images in the detection and diagnosis of breast cancer. *Seminars in Ultrasound, CT and MRI* 2004, 25: 411-8. WB Saunders.
106. Abdel-Zaher AM, Eldeib AM. Breast cancer classification using deep belief networks. *Expert Syst Appl* 2016, 46: 139-44.
107. Cairns P. Renal cell carcinoma. *Cancer Biomark* 2011, 9: 461-73.
108. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015, 13: 8-17.
109. Asri H, Mousannif H, Al Moatassime H, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Comput Sci* 2016, 83: 1064-9.
110. Liu HX, Zhang RS, Luan F, Yao XJ, Liu MC, Hu ZD, Fan BT. Diagnosing breast cancer based on support vector machines. *J Chem Inf Comput Sci* 2003, 43: 900-7.
111. Maglogiannis I, Zafiropoulos E, Anagnostopoulos I. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Appl Intell* 2009, 30: 24-36.
112. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019, 25: 44-56.
113. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA: a cancer journal for clinicians* 2011, 61: 69-90.
114. Wolberg WH, Street WN, Mangasarian OL. Image analysis and machine learning applied to breast cancer diagnosis and prognosis. *Analytical and Quantitative cytology and histology*. 1995, 17: 77-87.

115. Ramos-Pollán R, Guevara-López MA, Suárez-Ortega C, Díaz-Herrero G, Franco-Valiente JM, Rubio-Del-Solar M, González-De-Posada N, Vaz MA, Loureiro J, Ramos I. Discovering mammography-based machine learning classifiers for breast cancer diagnosis. *J Med Syst* 2012, 36: 2259-69.
116. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017, 542: 115.
117. Elgamal M. Automatic skin cancer images classification. (IJACSA) *International Journal of Advanced Computer Science and Applications* 2013, 4: 287-94.
118. Huo D, Kiehn M, Scherzinger A. Investigation of Low-Dose CT Lung Cancer Screening Scan “Over-Range” Issue Using Machine Learning Methods. *J Digit Imaging* 2019: 1-8.
119. Singh GA, Gupta PK. Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Comput Appl* 2019, 31: 6863-77.
120. Jagga Z, Gupta D. Classification models for clear cell renal carcinoma stage progression, based on tumor RNAseq expression trained supervised machine learning algorithms. *BMC proceedings* 2014, 8: S2.
121. Singh NP, Bapi RS, Vinod PK. Machine learning models to predict the progression from early to late stages of papillary renal cell carcinoma. *Comput biol med* 2018, 100: 92-9.
122. Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, Erdim C, Kilickesmez O. Clear cell renal cell carcinoma: machine learning-based quantitative computed tomography texture analysis for prediction of Fuhrman nuclear grade. *Eur radiol* 2019, 29: 1153-63.
123. Kocak B, Yardimci AH, Bektas CT, Turkcanoglu MH, Erdim C, Yucetas U, Koca SB, Kilickesmez O. Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with independent external validation. *Eur J Radiol* 2018, 107: 149-57.
124. Mohebian MR, Marateb HR, Mansourian M, Mañanas MA, Mokarian F. A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. *Comput Struct Biotechnol J* 2017, 15: 75-85.

- 125.** Cai Z, Xu D, Zhang Q, Zhang J, Ngai SM, Shao J. Classification of lung cancer using ensemble-based feature selection and machine learning methods. *Mol Biosyst* 2015, 11: 791-800.



EK-1. ÖZGEÇMİŞ

A. Kişisel Bilgiler

Adı Soyadı	Adem DOĞANER
Doğum Tarihi	1985
Doğum Yeri	Adana
Yabancı Dil Bilgisi	İngilizce-Almanca
Görev Yeri	Kahramanmaraş Sütçü İmam Üniversitesi Tıp Fakültesi
İletişim Bilgileri	adoganer@ksu.edu.tr

B. Eğitim Bilgileri

Derece	Alan	Üniversite	Yıl
Lisans	İstatistik	Muğla Üniversitesi	2004-2008
Yüksek Lisans	İstatistik	Fırat Üniversitesi	2009-2011
Doktora	İstatistik	Fırat Üniversitesi	2011-2015
II. Doktora	Biyostatistik	İnönü Üniversitesi	2016-2020

C. İş Tecrübesine Ait Bilgiler

Yıllar	Çalıştığı Kurum
2009-2015	Fırat Üniversitesi
2015-Devam Ediyor	Kahramanmaraş Sütçü İmam Üniversitesi

EK-2. KLİNİK ARAŞTIRMALAR ETİK KURULU ONAY FORMU

KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ TIP FAKÜLTESİ KLİNİK ARAŞTIRMALAR ETİK KURULU KARAR FORMU

ARAŞTIRMANIN AÇIK ADI	Topluluk Öğrenme Yöntemleri İle Renal Hücreli Karsinomun Tahmin Edilmesi
VARSA ARAŞTIRMANIN PROTOKOL KODU	162

ETİK KURUL BİLGİLERİ	ETİK KURULUN ADI	KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ TIP FAKÜLTESİ KLİNİK ARAŞTIRMALAR ETİK KURULU
	AÇIK ADRESİ:	KSÜ Tıp Fakültesi Dekanlığı Adres: Kayseri/Kahramanmaraş Yolu Üzeri Avşar Yerleşkesi 46000/ K.MARAŞ
	TELEFON	(0344)3003424
	FAKS	(0344)3003409
	E-POSTA	tipkaek@ksu.edu.tr

BAŞVURU BİLGİLERİ	KOORDİNATÖR/SORUMLU ARAŞTIRMACI UNVANI/ADI/SOYADI	Dr. Öğr. Üyesi Adem DOĞANER		
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ UZMANLIK ALANI	Biyostatistik ve Tıbbi Bilişim AD		
	KOORDİNATÖR/SORUMLU ARAŞTIRMACININ BULUNDUĞU MERKEZ	KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ TIP FAKÜLTESİ		
	VARSA İDARİ SORUMLU UNVANI/ADI/SOYADI			
	DESTEKLEYİCİ	yok		
	PROJE YÜRÜTÜCÜSÜ UNVANI/ADI/SOYADI (TÜBİTAK vb. gibi kaynaklardan destek alanlar için)			
	DESTEKLEYİCİNİN YASAL TEMSİLCİSİ			
	ARAŞTIRMANIN FAZİ VE TÜRÜ	FAZ 1	<input type="checkbox"/>	
		FAZ 2	<input type="checkbox"/>	
		FAZ 3	<input type="checkbox"/>	
FAZ 4		<input type="checkbox"/>		
Gözlemsel ilaç çalışması		<input type="checkbox"/>		
Tıbbi cihaz klinik araştırması		<input type="checkbox"/>		
İn vitro tıbbi tanı cihazları ile yapılan performans değerlendirme çalışmaları		<input type="checkbox"/>		
İlaç dışı klinik araştırma	<input checked="" type="checkbox"/>			
	- Dosya kullanılarak yapılan arşiv taraması - Ses ve görüntü (Radyolojik / dijital / analog) kayıtları kullanılarak yapılan arşiv taraması			
ARAŞTIRMAYA KATILAN MERKEZLER	TEK MERKEZ <input checked="" type="checkbox"/>	ÇOK MERKEZLİ <input type="checkbox"/>	ULUSAL <input type="checkbox"/>	ULUSLARARASI <input type="checkbox"/>

Etik Kurul Başkanı
Unvanı/Adı/Soyadı: Doç.Dr. Can ACIPAYAM
İmza:

Not: Etik kurul başkanı, imzasının yer almadığı her sayfaya imza atmalıdır.

**KAHRAMANMARAŞ SÜTÇÜ İMAM ÜNİVERSİTESİ TIP FAKÜLTESİ
KLİNİK ARAŞTIRMALAR ETİK KURULU KARAR FORMU**

DEĞERLENDİRİLEN BELGELER	Belge Adı	Tarihi	Versiyon Numarası	Dili			
	ARAŞTIRMA PROTOKOLÜ	yok		Türkçe <input checked="" type="checkbox"/>	İngilizce <input type="checkbox"/>	Diğer <input type="checkbox"/>	
	BİLGİLENDİRİLMİŞ GÖNÜLLÜ OLUR FORMU	yok		Türkçe <input checked="" type="checkbox"/>	İngilizce <input type="checkbox"/>	Diğer <input type="checkbox"/>	
	OLGU RAPOR FORMU	yok		Türkçe <input checked="" type="checkbox"/>	İngilizce <input type="checkbox"/>	Diğer <input type="checkbox"/>	
	ARAŞTIRMA BROŞÜRÜ	yok		Türkçe <input checked="" type="checkbox"/>	İngilizce <input type="checkbox"/>	Diğer <input type="checkbox"/>	
DEĞERLENDİRİLEN DİĞER BELGELER	Belge Adı	Açıklama					
	SİGORTA	<input type="checkbox"/>					
	ARAŞTIRMA BÜTÇESİ	<input type="checkbox"/>					
	BIYOLOJİK MATERYEL TRANSFER FORMU	<input type="checkbox"/>					
	İLAN	<input type="checkbox"/>					
	YILLIK BİLDİRİM	<input type="checkbox"/>					
	SONUÇ RAPORU	<input type="checkbox"/>					
	GÜVENLİLİK BİLDİRİMLERİ	<input type="checkbox"/>					
DIĞER:	<input checked="" type="checkbox"/>	Başvuru Dilekçesi, Başvuru Formu, Özgeçmişler, Arşiv izni					
KARAR BİLGİLERİ	Karar No: 21	Tarih: 04.04.2018	Oturum: 2018/07				
	Yukarıda bilgileri verilen başvuru dosyası ile ilgili belgeler araştırmacının/çalışmanın gerekece, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş ve uygun bulunmuş olup araştırmacının/çalışmanın başvuru dosyasında belirtilen merkezlerde gerçekleştirilmesinde etik ve bilimsel sakınca bulunmadığına toplantıya katılan etik kurul üye tam sayısının salt çoğunluğu ile karar verilmiştir. Kök Hücre, doku nakli, organ nakli ve yeni bir cerrahi yöntem ile ilgili çalışmalar ve geleneksel tıp uygulamaları ve tıbbi ürünler ile ilgili çalışmalar için ayrıca Sağlık Hizmetleri Genel Müdürlüğünden izin alınması gerekmektedir. İlaç ve Biyolojik Ürünlerin Klinik Araştırmaları Hakkında Yönetmelik kapsamında yer alan araştırmalar/çalışmalar için Türkiye İlaç ve Tıbbi Cihaz Kurumu'ndan izin alınması gerekmektedir.						
KLİNİK ARAŞTIRMALAR ETİK KURULU							
ETİK KURULUN ÇALIŞMA ESASI		İlaç ve Biyolojik Ürünlerin Klinik Araştırmaları Hakkında Yönetmelik, İyi Klinik Uygulamaları Kılavuzu					
BAŞKAN UNVANI / ADI / SOYADI:		Doç. Dr. Can ACIPAYAM					
Unvanı/Adı/Soyadı	Uzmanlık Alanı	Kurumu	Cinsiyet	Araştırma ile İlişki	Katılım *	İmza	
BASKAN Doç. Dr. Can ACIPAYAM	Çocuk Sağlığı ve Hastalıkları	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Doç. Ahmet Çağrı AYKAN Başkan Yardımcısı Üye	Kardiyoloji	KSU Tıp Fakültesi	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Doç. Dr. Sezgin KOÇARSLAN Üye	Tıbbi Patoloji	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Doç. Dr. Mete GÜLER Üye	Göz Hastalıkları	KSU Tıp Fakültesi	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Dr.Öğr. Üyesi Günen ÖKSÜZ Üye	Anesteziyoloji ve Reanimasyon	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Dr.Öğr. Üyesi Aysegül ERDOĞAN Üye	Halk Sağlığı	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Dr.Öğr. Üyesi Selma YAMAN Üye	Biyofizik	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Dr.Öğr. Üyesi Nadire ESER Üye	Farmakoloji	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Dr.Öğr. Üyesi Adem DOĞANER Üye	Biyoistatistik	KSU Tıp Fakültesi	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>	ARAŞTIRMACI	
Dr.Öğr. Üyesi Nağihan BİLAL Üye	Kulak, Burun, Boğaz Hastalıkları	KSU Tıp Fakültesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Üm.Ecz. Dilara Algül DOKUMACI Üye	Eczacı	Dünya Eczanesi	E <input type="checkbox"/> K <input checked="" type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Öğt.Gör. Ahmet KARATUT Üye	Hukukçu	KSU Pazarlık MYO	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Hakan SERBETÇİOĞLU Üye	Mühendis	Mavi-Yeşil Yazılım	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
Hacı Ömer DOKUMACI Üye	Mühendis	Serbest	E <input checked="" type="checkbox"/> K <input type="checkbox"/>	E <input type="checkbox"/> H <input checked="" type="checkbox"/>	E <input checked="" type="checkbox"/> H <input type="checkbox"/>		
SERH(VARSA)							

*:Toplantıda Bulunma

Etik Kurul Başkanı
Unvanı/Adı/Soyadı: Doç. Dr. Can ACIPAYAM
İmza:

Not: Etik kurul başkanı, imzasının yer almadığı her sayfaya imza atmalıdır.