T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BAŞLICA İNGUİNAL KANAL PATOLOJİLERİNİN MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE TAHMİN EDİLMESİ

Alihan YILMAZ

YÜKSEK LİSANS TEZİ
Matematik Anabilim Dalı
Matematik Programı

Danışman
Dr. Öğr. Üyesi Selmahan SELİM

Eş-Danışman
Doç. Dr. Emrah AYDIN

Şubat, 2024

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BAŞLICA İNGUİNAL KANAL PATOLOJİLERİNİN MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE TAHMİN EDİLMESİ

Alihan YILMAZ tarafından hazırlanan tez çalışması 05.02.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı Matematik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Dr. Öğr. Üyesi Selmahan SELİM
Yıldız Teknik Üniversitesi
Danışman

Doç. Dr. Emrah AYDIN
Tekirdağ Namık Kemal Üniversitesi
Eş-Danışman

Jüri Üyeleri

Dr. Öğr. Üyesi Selmahan SELİM, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Özgür YILDIRIM, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Şebnem ÖZDEMİR, Üye
İstinye Üniversitesi _____

Danışmanım Dr. Öğr. Üyesi Selmahan SELİM sorumluluğunda tarafımca hazırlanan Başlıca İnguinal Kanal Patolojilerinin Makine Öğrenmesi Algoritmaları ile Tahmin Edilmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Alihan YILMAZ

İmza

Aileme

ve

Sevgili eşime

TEŞEKKÜR

Bu tez çalışmasında bilgi ve tecrübesiyle bana her zaman yardımcı olan saygıdeğer danışman hocam Dr. Öğr. Üyesi Selmahan SELİM'e, veri setinin elde edilmesi, tıp alanındaki bilgi ve tecrübeleriyle bizleri yönlendiren ve desteklerini esirgemeyen saygıdeğer danışman hocam Doç. Dr. Emrah AYDIN'a, yüksek lisans eğitimim süresince ders aldığım tüm hocalarıma ve öğretim hayatım boyunca destekleri ile her zaman yanımda olan aileme ve sevgili eşime teşekkürlerimi sunarım.

Alihan YILMAZ

İÇİNDEKİLER

KISALTMA LİSTESİ vii

ŞEKİL LİSTESİ viii

TABLO LİSTESİ ix

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Literatür Özeti . . . 1
1.2 Tezin Hipotezi ve Amacı . . . 2
1.3 Orjinal Katkı . . . 2

2 MAKİNE ÖĞRENMESİ VE KAVRAMLARI 3

2.1 Veri . . . 3
2.1.1 Veri İşleme . . . 4
2.2 Makine Öğrenmesi . . . 5
2.3 Öğrenme Yöntemleri . . . 5
2.3.1 Denetimsiz Öğrenme . . . 5
2.3.2 Denetimli Öğrenme . . . 6
2.4 Öğrenme Yöntemine Karar Verme . . . 6
2.4.1 Sınıflandırma için Gereken Ölçütler . . . 6
2.4.2 Regresyon için Gereken Ölçütler . . . 7
2.5 Denetimli Öğrenme Algoritmaları . . . 8
2.5.1 Karar Ağaçları . . . 8
2.5.2 Naive Bayes Sınıflandırmaları . . . 10
2.5.3 Destek Vektör Makineleri (DVM) . . . 13

3 BAŞLICA İNGUİNAL (KASIK) KANAL PATOLOJİLERİ 18

3.1 İnmemiş Testis . . . 19
3.2 İnguinal Herni . . . 19
3.3 Hidrosel . . . 19

3.4 Tam Kan Sayımı Parametreleri 20

4 BAŞLICA İNGUİNAL KANAL PATOLOJİLERİNİN TAHMİN EDİLMESİ 22

4.1 Verilerin Düzenlenmesi 22
4.2 İstatistiksel Analiz 25
4.3 Hastalıkların Sınıflandırma Modellerine Göre Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesi 29
4.3.1 Kontrol ve Hasta Verilerini Sınıflandırma Modeli 31
4.3.2 Kontrol ve İnmemiş Testis Verilerini Sınıflandırma Modeli . 33
4.3.3 Kontrol ve İnguinal Herni Verilerini Sınıflandırma Modeli . 35
4.3.4 Kontrol ve Hidrosel Verilerini Sınıflandırma Modeli 37
4.3.5 İnmemiş Testis Olmayan ve İnmemiş Testis Verilerini Sınıflandırma Modeli 38
4.3.6 İnguinal Herni Olmayan ve İnguinal Herni Verilerini Sınıflandırma Modeli 40
4.3.7 Hidrosel Olmayan ve Hidrosel Verilerini Sınıflandırma Modeli 42
4.3.8 İnmemiş Testis ve İnguinal Herni Verilerini Sınıflandırma Modeli 44
4.3.9 İnmemiş Testis ve Hidrosel Verilerini Sınıflandırma Modeli 46
4.3.10 İnguinal Herni ve Hidrosel Verilerini Sınıflandırma Modeli . 48

5 SONUÇ VE ÖNERİLER 51

KAYNAKÇA 55

TEZDEN ÜRETİLMİŞ YAYINLAR 57

KISALTMA LİSTESİ

BDH	Bahçelievler Devlet Hastanesi
BA	Bazofil
DVM	Destek Vektör Makinesi
EO	Eozinofil
HCT	Hematokrit
HGB	Hemoglobin
LY	Lenfosit
MCH	Mean Corpuscular Hemoglobin (Ortalama Hücresel Hemoglobin)
MCHC	Mean Corpuscular Hemoglobin Concentration (Ortalama Hücresel Hemoglobin Konsantrasyonu)
MCV	Mean Corpuscular Volume (Ortalama Hücresel Hacim)
MPV	Mean Platelet Volume (Ortalama Trombosit Hacmi)
MO	Monosit
NE	Nötrofil
PLT	Platelet (Trombosit)
PDW	Platelet Distribution Width (Trombosit Dağılım Genişliği)
RBC	Red Blood Cell (Kırmızı Kan Hücresi)
RDW	Red Cell Distribution Width (Kırmızı Hücre Dağılım Genişliği)
SVM	Support Vector Machine (Destek Vektör Makinesi)
SMOTE	Synthetic Minority Oversampling Technique
TKNÜ	Tekirdağ Namık Kemal Üniversitesi
WBC	White Blood Cell (Beyaz Kan Hücresi)

ŞEKİL LİSTESİ

Şekil 2.1 SMOTE Algoritması Çalışma Sistemi [10] 4
Şekil 2.2 Karar Ağacı Yapısı [14] 9
Şekil 2.3 Destek Vektör Makineleri 13
Şekil 2.4 Çizilebilecek Hiper Düzlemler [17] 14
Şekil 2.5 En Uygun Hiper Düzlem [17] 14
Şekil 2.6 En Uygun Hiper Düzlemin Belirlenmesi [17] 15
Şekil 2.7 Doğrusal Ayrılamayan Veriler için Hiper Düzlemin Belirlenmesi [17] 16
Şekil 2.8 Doğrusal Ayrılamayan Veriler için Optimum Hiper Düzlemin Belirlenmesi [17] 16
Şekil 2.9 Kernel Fonksiyonu ile Verinin Boyutunun Yükseltilmesi [17] . . . 17
Şekil 3.1 Normal İnguinal Kanal 18
Şekil 3.2 İnguinal Herni 19
Şekil 3.3 Hidrosel 20
Şekil 5.1 Kontrol ve Hasta Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması 52
Şekil 5.2 Sırasıyla Kontrol ve İnmemiş Testis, Kontrol ve İnguinal Herni, Kontrol ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması 52
Şekil 5.3 Sırasıyla İnmemiş Testis Olmayan ve İnmemiş Testis, İnguinal Herni Olmayan ve İnguinal Herni, Hidrosel Olmayan ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması . 53
Şekil 5.4 Sırasıyla İnmemiş Testis ve İnguinal Herni, İnmemiş Testis ve Hidrosel, İnguinal Herni ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması 53

TABLO LİSTESİ

Tablo 2.1 Karşılaştırma Matrisi 6
Tablo 2.2 Hastaneye Başvuran Sayısı ve Hasta Oranı 10
Tablo 2.3 Sonuç Değerleri Tablosu 11
Tablo 4.1 TKNÜ Verilerinin Tanı ve Kişi Dağılımları 23
Tablo 4.2 Beyaz Kan Hücresi Parametreleri Arasındaki Pearson Korelasyon Analizi 24
Tablo 4.3 BDH Verilerinin Tanı ve Kişi Dağılımları 25
Tablo 4.4 Tanılara Göre Cinsiyet Sayıları 25
Tablo 4.5 Kategorilere Göre Değişkenlere Ait Ortalama ve Standart Sapma Değerleri 26
Tablo 4.6 Tanılara Ait Değişkenler için p Değerleri 27
Tablo 4.7 SMOTE Algoritması Sonrası Tanılara Göre Cinsiyet Sayıları . . 27
Tablo 4.8 SMOTE Algoritması Sonrası Kategorilere Göre Değişkenlere Ait Ortalama ve Standart Sapma Değerleri 28
Tablo 4.9 SMOTE Algoritması Sonrası Tanılara Ait Değişkenler İçin p Değerleri 29
Tablo 4.10 Makine Öğrenmesinde Kullanılacak Tanı ve Kişi Dağılımı 29
Tablo 4.11 SMOTE Algoritmasından Sonra Makine Öğrenmesinde Kullanılacak Tanı ve Kişi Dağılımı 31
Tablo 4.12 Kontrol ve Hasta Veri Grubunun Karşılaştırma Matrisi 31
Tablo 4.13 Kontrol ve Hasta Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 32
Tablo 4.14 Kontrol ve Hasta SMOTE Veri Grubunun Karşılaştırma Matrisi . 32
Tablo 4.15 Kontrol ve Hasta SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 33
Tablo 4.16 Kontrol ve İnmemiş Testis Veri Grubunun Karşılaştırma Matrisi . 33
Tablo 4.17 Kontrol ve İnmemiş Testis Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 34
Tablo 4.18 Kontrol ve İnmemiş Testis SMOTE Veri Grubunun Karşılaştırma Matrisi 34
Tablo 4.19 Kontrol ve İnmemiş Testis SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 35

Tablo 4.20 Kontrol ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisi . 35
Tablo 4.21 Kontrol ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 35
Tablo 4.22 Kontrol ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisi 36
Tablo 4.23 Kontrol ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 36
Tablo 4.24 Kontrol ve Hidrosel Veri Grubunun Karşılaştırma Matrisi 37
Tablo 4.25 Kontrol ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 37
Tablo 4.26 Kontrol ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi 38
Tablo 4.27 Kontrol ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 38
Tablo 4.28 İnmemiş Testis ve İnmemiş Testis Olmayan Veri Grubunun Karşılaştırma Matrisi 39
Tablo 4.29 İnmemiş Testis ve İnmemiş Testis Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 39
Tablo 4.30 İnmemiş Testis ve İnmemiş Testis Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi 40
Tablo 4.31 İnmemiş Testis ve İnmemiş Testis Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 40
Tablo 4.32 İnguinal Herni ve İnguinal Herni Olmayan Veri Grubunun Karşılaştırma Matrisi 41
Tablo 4.33 İnguinal Herni ve İnguinal Herni Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 41
Tablo 4.34 İnguinal Herni ve İnguinal Herni Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi 42
Tablo 4.35 İnguinal Herni ve İnguinal Herni Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 42
Tablo 4.36 Hidrosel ve Hidrosel Olmayan Veri Grubunun Karşılaştırma Matrisi 43
Tablo 4.37 Hidrosel ve Hidrosel Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 43
Tablo 4.38 Hidrosel ve Hidrosel Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi 44
Tablo 4.39 Hidrosel ve Hidrosel Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 44
Tablo 4.40 İnmemiş Testis ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisi 45

Tablo 4.41 İnmemiş Testis ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 45
Tablo 4.42 İnmemiş Testis ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisi 46
Tablo 4.43 İnmemiş Testis ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 46
Tablo 4.44 İnmemiş Testis ve Hidrosel Veri Grubunun Karşılaştırma Matrisi 47
Tablo 4.45 İnmemiş Testis ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 47
Tablo 4.46 İnmemiş Testis ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi 48
Tablo 4.47 İnmemiş Testis ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 48
Tablo 4.48 İnguinal Herni ve Hidrosel Veri Grubunun Karşılaştırma Matrisi . 48
Tablo 4.49 İnguinal Herni ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 49
Tablo 4.50 İnguinal Herni ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi 49
Tablo 4.51 İnguinal Herni ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar 49

ÖZET

Başlıca İnguinal Kanal Patolojilerinin Makine Öğrenmesi Algoritmaları ile Tahmin Edilmesi

Alihan YILMAZ

Matematik Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Selmahan SELİM
Eş-Danışman: Doç. Dr. Emrah AYDIN

Doğuma yakın dönemde ya da doğumu takiben testislerin skrotuma yerleşmesi gerekmektedir. Bazı hormon ve faktörlerin yetersiz düzeyde salgılanması gibi yaşanan bir takım sorunlar sebebiyle testislerin skrotuma inememesi inmemiş testis rahatsızlığına, kasık kanalını karın boşluğundan ayıran zarın tam kapanamaması ya da geç kapanması da inguinal herni veya hidrosel rahatsızlıklarına sebep olmaktadır. Bu çalışmanın amacı hastaneye başvuran çocuk hastalardan alınan tam kan sayımı (hemogram), yaş ve cinsiyet verilerinin ilişkisinden yola çıkarak, makine öğrenmesi algoritmalarıyla başvuranın inmemiş testis, inguinal herni ve hidrosel hastalıklarına sahip olup olmadığını öngörmektir. Gerçekleştirilen çalışmada öncelikle; 2014-2023 yılları arasında Tekirdağ Namık Kemal Üniversitesi Hastanesi Çocuk Cerrahisi ABD Kliniğine ve 2014-2016 yılları arasında Bahçelievler Devlet Hastanesi Çocuk Cerrahisi Kliniğine başvuran kişilerin verileri geriye dönük olarak toplanmıştır. Bu verilere temizleme, birleştirme ve dönüştürme işlemleri uygulandıktan sonra; 4164 sağlıklı, 555 inmemiş testis, 668 inguinal herni, 291 hidrosel veri gruplarının rastgele ayrılan %75'lik kısımları makine öğrenmesi algoritmaları modellerince 5 katlamalı çapraz doğrulama ile eğitilmiş olup, eğitilen modeller geriye kalan %25'lik kısmı ile test edilmiştir. Yapılan testler sonucunda hedeflenen sonuçların elde edilememesinin veri setinin dengesiz dağılımından kaynaklandığı düşünülerek, verilere SMOTE (Synthetic Minority Oversampling Technique) algoritması uygulanmıştır. SMOTE algoritması uygulandıktan sonra; 4164'er sağlıklı, inmemiş testis, inguinal herni ve hidrosel

veri gruplarının tekrardan rastgele ayrılan %70'lik kısımları makine öğrenmesi algoritmalarınca eğitilmiş olup, eğitilen modeller geriye kalan %30'luk kısmı ile test edilmiştir. Yapılan testler sonucunda inguinal kanal patolojilerinden toplumda en sık karşılaşılan inmemiş testis, inguinal herni ve hidrosel olgularının hemogram verileri ile ayrılabileceği doğrulanmıştır. Hasta verilerinin geriye dönük toplanmasından kaynaklı olarak inmemiş testis olgularına eşlik edebilen inguinal herni ya da hidrosel olgularının ayrımının yapılamaması, hemogram verilerini oluşturan hücrelerin yaşam sürelerinin günler ile sınırlı olmasından kaynaklı olarak çalışmada kullanılan yaş parametresinin gün olarak değil de yıl olarak hesaplanması gibi durumların giderilmesi halinde daha optimize sonuçlar elde edileceği düşünülmektedir.

Anahtar Kelimeler: Sağlıkta Yapay Zeka, İnguinal Kanal Patolojileri, İnmemiş Testis, İnguinal Herni, Hidrosel

ABSTRACT

Prediction of Inguinal Canal Pathologies Using Machine Learning Algorithms

Alihan YILMAZ

Department of Mathematics
Master of Science Thesis

Supervisor: Asst. Prof. Selmahan SELİM
Co-supervisor: Assoc. Prof. Emrah AYDIN

In the period close to birth or following birth, the testicles must settle in the scrotum. Failure of the testicles to descend into the scrotum due to some problems such as insufficient secretion of some hormones and factors causes undescended testicle disorder, and incomplete closure or delayed closure of the membrane separating the inguinal canal from the abdominal cavity causes inguinal hernia or hydrocele. The aim of this study is to predict whether the applicant has undescended testicles, inguinal hernia and hydrocele diseases using machine learning algorithms, based on the relationship between complete blood count (hemogram), age and gender data taken from pediatric patients admitted to the hospital. In the study carried out, first of all; The data of people who applied to Tekirdağ Namık Kemal University Hospital Pediatric Surgery Clinic between 2014-2023 and Bahçelievler State Hospital Pediatric Surgery Clinic between 2014-2016 were collected retrospectively. After cleaning, combining and transforming this data; Randomly divided %75 parts of 4164 healthy, 555 undescended testicles, 668 inguinal hernia, 291 hydrocele data groups were trained by machine learning algorithm models with 5-fold cross validation, and the trained models were tested with the remaining %25 part. Considering that the failure to obtain the targeted results as a result of the tests was due to the unbalanced distribution of the data set, the SMOTE (Synthetic Minority Oversampling Technique) algorithm was applied to the data. After applying the SMOTE algorithm; %70 of the 4164 healthy, undescended testicles, inguinal hernia and hydrocele data groups were randomly

allocated again and were trained by machine learning algorithms, and the trained models were tested with the remaining %30. As a result of the tests, it was confirmed that undescended testicle, inguinal hernia and hydrocele cases, which are the most common inguinal canal pathologies in the society, can be distinguished with hemogram data. It would be more optimized if situations such as the inability to distinguish inguinal hernia or hydrocele cases that may accompany undescended testicle cases due to retrospective collection of patient data, and the calculation of the age parameter used in the study in years rather than days due to the life span of the cells that form the hemogram data being limited to days, are eliminated. It is thought that results will be achieved.

Keywords: Artificial Intelligence in Health, Inguinal Canal Pathologies, Undescended Testis, Inguinal Hernia, Hydrocele

1
GİRİŞ

1.1 Literatür Özeti

Testislerin, görevlerini yapabilmek için uygun sıcaklık koşullarında olması gerekmektedir. Testislerin oluştuğu karın içi sıcaklığı, sperm hücrelerinin sağlıklı gelişmesine elverişli olmadığı için doğuma yakın dönemde ya da doğumu takiben skrotuma yerleşmesi gerekmektedir. Ancak bazı hormon ve faktörlerin yetersiz düzeyde salgılanması gibi yaşanan bir takım sorunlar sebebiyle skrotuma yerleşememektedir. Bu şekilde testislerin inememesine inmemiş testis, açık kalan kanallardan; karın içi yapıların skrotuma ilerlermesine inguinal herni, karın içi sıvının testisin etrafını saran zarların arasına birikmesine hidrosel adı verilmektedir.

Yılmaz ve Özen'in yaptığı çalışmada [1]; doğumsal anomalilerde, inmemiş testis hastalığı sık görülen bir hastalık olarak tanımlanmaktadır. Doğum haftası ve doğum anındaki ağırlığı ile inmemiş testis hastalığının görülme sıklığı arasında bir bağlantı olduğu görülmektedir. 1,5 kg ve daha düşük ağırlıkta doğan erkek çocuklarda testislerin inmemesinin görülme sıklığı %60-70'leri bulmakta iken, prematür doğan bebeklerde bu oran %33'e düşmekte ve yenidoğan bebeklerde %3'e kadar inmektedir.

Alp ve arkadaşlarının yaptığı çalışmada [2]; çocuk cerrahisi kliniklerinin yaptığı ameliyatlardan %30-50'sini kasık fıtığı oluşturmaktadır. Bu oranların yaklaşık olarak %0,8-4,4'ü ise bebek ve çocuklardan oluşmaktadır. Fetal dönemde testis ilerlemesi ile kasık fıtığının oluşumu arasında paralellik olmasından dolayı, erkek çocukları kasık fıtığı sayısı ile kız çocukları kasık fıtığı sayısı oranı 6 olarak gözlemlenmektedir. Bu nedenle çalışmalardaki olguların çoğunluğunun erkek çocuklardan oluşması literatür ile uyumlu gözükmektedir. Fakat [2] çalışmasında bu oran 2,57 olarak tespit edilmiştir. Literatürde fıtıkların %60'ının sağda, %30'unun solda ve kalanlarının ise iki tarafta olduğu gözlemlenmektedir. Yapılan [2] çalışmasında ise fıtıkların %50'si sağ tarafta yer alırken %20'si sol tarafta ve kalanı ise iki tarafta olduğu saptanmıştır. Çocuk yaştaki gruplarda kasık fıtığının %100'e

yakını indirekt kasık fıtığı olarak saptanmakta ve [2] çalışmasında da tümü indirekt kasık fıtığı olarak saptanmıştır.

Yapılan diğer bir araştırmada [3]; 28 aylık dönemde çocuk cerrahisi polikliniğine başvuru yapan 3358 hastanın %6,9'u inguinal bölge patolojisine sahip olarak tanımlanmıştır. Hastalardan %82'si erkek, %18'i kız olarak saptanmıştır. Yaşları 1 gün ile 13 yıl arasında değişen kızların yaşları ortalamaları $3,67 \pm 4$ yıl olup, yaşları 1 gün ile 16 yıl arasında olan erkeklerin ortalama yaşı ise $3,32 \pm 9$ yıl olarak gözlemlenmiştir. Kızların ve erkeklerin yaşları karşılaştırıldığında istatistiksel anlamda farklılık bulunmamıştır. Gerçekleştirilen [3] araştırmasında, 233 hastadan %70,4'ünde inguinal herni görülüp hidrosel görülmezken, hidrosel görülen hastalardan %26,4'ünde ise sadece hidrosel görülmüştür. Geriye kalan hastalarda da kasık fıtığı ile hidroselin birlikte olduğu gözlemlenmiştir.

1.2 Tezin Hipotezi ve Amacı

Tezin hipotezi, hastaneye başvuran çocuk hastalardan alınan tam kan sayımı (hemogram) verileri ile o hastaların inmemiş testis, inguinal herni ve hidrosel hastalıklarından hangisine sahip olduğunun belirlenebileceğidir. Bu hipotezi gerçekleştirmek için çalışmanın ilk amacı, ön işleme ile geriye yönelik olarak toplanan hemogram verilerinin temizlenmesi, anonimleştirilmesi ve farklı dosyalarda bulunan verilerin tek dosyada birleştirilmesi işlemleri ile verilerin istatistiksel analize uygun hale getirilmesidir. Bu sayede mevcut veriler standardize hale getirilmiş olup analiz edilmeye uygun olacaklardır. Bu amacı gerçekleştirdikten sonra çalışmanın ikinci amacı, hemogram verileri ile inmemiş testis, inguinal herni ve hidrosel hastalıklarının arasındaki ilişkiyi yapay zeka algoritmaları ile analiz edebilmektir. Bu sayede tam kan sayımı parametreleri ile hastalıkların öngörülmesi hedeflenmektedir.

1.3 Orjinal Katkı

Yapılan literatür taraması sonucunda bildiğimiz kadarıyla insanlardan alınan tam kan sayımı parametreleri ile inmemiş testis, inguinal herni ve hidrosel hastalıklarına sahip olunması arasındaki ilişkiyi araştıran bir çalışmanın bulunmadığı tespit edilmiştir.

2
MAKİNE ÖĞRENMESİ VE KAVRAMLARI

2.1 Veri

Etrafımızdaki her şeyin bir veri kaynağına bağlı olduğu ve hayatımızdaki her şeyin dijital olarak kaydedildiği veri çağında yaşamaktayız. Örneğin; siber güvenlik verileri, iş verileri, akıllı telefon verileri, sosyal medya verileri, sağlık verileri vb. bir çok veri grubu çok çeşitli veri zenginliğine sahiptir [4].

Bilgisayarlar olmadan önce, insanlar veriler içerisinden belli ilişkiler çıkarmaya çalışmışlardır. Doğrusal denklemleri çözme yöntemleri (Newton Gauss), gradyan iniş yoluyla optimum öğrenme (Newton), polinom enterpolasyonu (Lagrange), en küçük kareler uydurma (Laplace) vb. matematiksel teknikler, bu ilişkileri ortaya çıkarmak için yardımcı tekniklerdir. XX.Yüzyılın ortalarına doğru ortaya çıkan bilgisayarlarla beraber, bu teknikler otomatikleştirilmiştir. Veriler, matematiksel tekniklere olan ihtiyacın artmasından dolayı, veri analizi için bilgisayarlara yüklenerek makinenin verileri öğrenmesi hedeflenmiştir. Böylece makinelerle farklı teknikler geliştirilmiş ve algoritmalar ortaya çıkarılmıştır [5].

İşlenmemiş ham bilgi veri anlamına gelmektedir. Verilerin kullanılabilirliği, genellikle bir öğrenme algoritması veya veriye dayalı gerçek dünya sistemleri oluşturmanın anahtarı olarak kabul edilmektedir [6]. Verilerin anlamlandırılabilir ve birleştirilebilir olması, aralarında bir ilişki olması durumunda mümkün olmaktadır. İlişkiden yola çıkarak anlamlandırılan bu veriler, belirlenen amaçlara yönelik işlenerek kullanılabilir hale getirilmektedir [7].

Çalışmanın bu kısmının devamında, belirlenen amaçlara yönelik verilerin işlenmesi, makine öğrenmesi, öğrenme yöntemleri, öğrenme yöntemine karar verme, denetimli öğrenme algoritmaları anlatılmaktadır.

2.1.1 Veri İşleme

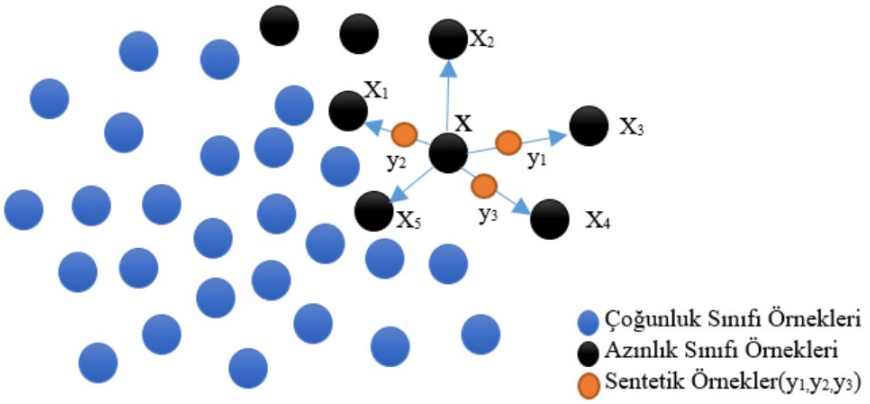
Veri Temizleme: Özniteliği eksik olan verilerin, veri grubundan silinmesi işlemidir [8].

Veri Birleştirme: Farklı dosyalarda bulunan verileri, ortak sütunlarından yararlanılarak bir dosyaya toplama işlemidir [8].

Veri Dönüştürme: Verilerin analiz edilebilmesi için uygun formlara uyarlanması işlemidir [8].

Veri İndirgeme: Yapılan işlemlerden elde edilecek sonucun veriyi olumsuz yönde etkilemeyeceği düşünüldüğünde veya etkisiz özniteliklerin var olması durumunda uygulanabilen verilerin azaltılması işlemidir [8].

Dengesiz Veri Setlerinde Aşırı Örnekleme: Algoritmaların, veri sayısının yetersizliği veya dengesizliğinden dolayı istenilen sonucu verememesi durumunda, orjinal veri grubundan yararlanılarak veri sayısını eşitleme işlemidir. Bunun için kullanılan en yaygın algoritma SMOTE algoritmasıdır [9].



Şekil 2.1 SMOTE Algoritması Çalışma Sistemi [10]

SMOTE algoritması, veri grubundaki azınlık veri sayısını sentetik bir şekilde artırarak dengeli bir veri seti oluşturmaktadır [10]. Algoritmanın çalışma adımları;

1. Azınlık veri grubundaki herhangi bir veri (y_k) seçilir.
2. Seçilen verinin öznitelikleri (x_i) ile o özniteliklere en yakın olan öznitelikler (x_j) arasındaki fark alınır ve 0 ile 1 arasında seçilen rastgele bir sayı (α) ile çarpılarak Denklem 2.1 elde edilir.

$$\alpha \cdot (x_i - x_j) \tag{2.1}$$

3. Denklem 2.1 ile seçilen verinin özniteliği (x_i) toplanarak Denklem 2.2 elde edilir.

$$x_{yeni} = x_i + \alpha \cdot (x_i - x_j) \tag{2.2}$$

4. Bulunan yeni öznitelik değerleri ile yeni veri (y_{yeni}) oluşturulur.

5. Veri sayısı az olan veri grubu ile çok olan veri grubu sayıca eşitlenene kadar 1-5. adımlar tekrarlanır.

2.2 Makine Öğrenmesi

Yapay zekanın alt bir dalı olan makine öğrenmesi, istatistiksel ve matematiksel işlemlerle beraber verilerin içerisinden çıkarımlar yaparak tahminler yapan sistemin bilgisayarlar ile modellenmesi anlamına gelmektedir. Veri setlerinin birbiri arasındaki örüntülerden yararlanarak çeşitli algoritmalarla bir mantık oluşturmayı amaçlayan makine öğrenmesi, performans bakımından en yüksek sonucu vermek istemesi sebebiyle, birçok makine öğrenmesi algoritmasının geliştirilmesine sebep olmuş ve bu algoritmaların bazılarıyla kümeleme, bazılarıyla ise sınıflandırma ve tahminleme yapılabilmektedir [11].

2.3 Öğrenme Yöntemleri

Makine öğrenmesi temel olarak denetimli ve denetimsiz öğrenme şeklinde iki grupta incelenebilmektedir. Denetimsiz öğrenme bağımsız değişkenler arasındaki ilişkiden yola çıkarak öğrenirken, denetimli öğrenme bağımsız değişkenler ile bağımsız değişkenlere bağlı olan değişkenler arasındaki ilişkiden yola çıkarak öğrenme işlemini gerçekleştirmektedir.

2.3.1 Denetimsiz Öğrenme

Bağımlı değişkenler olmadan sadece bağımsız değişkenler arasındaki ilişkiyi ortaya koymayı amaçlamaktadır. Algoritma, bu değerler arasındaki en yakın olanları kendi içerisinde gruplandırarak kümelemeyi oluşturmaktadır. Yeni bağımsız değişkenlerin değerleri bu kümelerden hangisiyle daha yakın değerlere sahipse o kümeye ait sayılmaktadır. Bu kümeleme işlemini en doğru şekilde yapan algoritmanın, yeni bağımsız değişken değerlerinin kümelemesini de en doğru şekilde verebilmesi beklenmektedir.

2.3.2 Denetimli Öğrenme

Bağımsız değişkenlerin değerleri ile bağımlı değişkenlerin değerlerinin beraber verilmesiyle aralarındaki ilişkiyi ortaya koymayı amaçlamaktadır. Bağımlı değişkenlerin değerleri; belirli bir değer kümesine göre ayrılabiliyorsa sınıflandırma yöntemleri, bir aralıkta ki herhangi bir sayısal değeri alabiliyorsa regresyon yöntemleri kullanılmaktadır. Bağımlı ve bağımsız değişkenlerin arasındaki ilişkiyi en doğru şekilde öğrenen modelin, yeni verilerin bağımsız değişkenlerine bakarak çıktı değerlerini de en doğru şekilde verebilmesi beklenmektedir.

2.4 Öğrenme Yöntemine Karar Verme

Bu bölümde, öğrenme yöntemine karar vermek için gereken ölçütlerden bahsedilmektedir.

2.4.1 Sınıflandırma için Gereken Ölçütler

2.4.1.1 Karşılaştırma Matrisi (Confusion Matrix)

İkili karşılaştırmalarda gerçekleştirilen öğrenmede, Tablo 2.1'deki değerlerden yararlanarak; sınıflandırmanın doğruluğu, kesinliği, özgüllüğü ve duyarlılığı hakkında bize bilgi vermektedir.

Tablo 2.1 Karşılaştırma Matrisi

Gerçek Sınıfı	**Tahmin Sınıfı**	
	Y_1	Y_2
Y_1	D_{Y_1}(Doğru, Y_1)	Y_{Y_2}(Yanlış, Y_2)
Y_2	Y_{Y_1}(Yanlış, Y_1)	D_{Y_2}(Doğru, Y_2)

2.4.1.2 Doğruluk(Accuracy)

Veri setiyle eğitilen algoritma modelinin, iki değişkene ait doğru olarak tahmin ettiği veri sayısının toplam veri sayısına oranı anlamına gelmektedir.

$$\text{Doğruluk} = \frac{D_{Y_1} + D_{Y_2}}{D_{Y_1} + Y_{Y_1} + D_{Y_2} + Y_{Y_2}} \tag{2.3}$$

2.4.1.3 Kesinlik(Specificite)

Eğitilen algoritma modelinin, değişkenlerden birine ait doğru tahmin sayısının, o değişkene ait doğru ve yanlış tahmin sayısına oranı anlamına gelmektedir.

$$\text{Kesinlik} = \frac{D_{Y_1}}{D_{Y_1} + Y_{Y_1}} \tag{2.4}$$

2.4.1.4 Duyarlılık(Sensitivite)

Algoritma modelinin, bir değişkene ait doğru tahmin sayısının, o değişkene ait veri sayısına oranı anlamına gelmektedir.

$$\text{Duyarlılık} = \frac{D_{Y_1}}{D_{Y_1} + Y_{Y_2}} \tag{2.5}$$

2.4.1.5 Özgüllük(Spesifite)

Eğitilen algoritma modelinin, diğer değişkene ait doğru tahmin sayısının, o değişkene ait veri sayısına oranı anlamına gelmektedir.

$$\text{Özgüllük} = \frac{D_{Y_2}}{Y_{Y_1} + D_{Y_2}} \tag{2.6}$$

2.4.1.6 F_1-Skoru

Kesinlik değeri ile duyarlılık değerin harmonik ortalamasıyla bulunmaktadır.

$$\text{F}_1 = 2 \cdot \frac{\text{Kesinlik} \cdot \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \tag{2.7}$$

2.4.2 Regresyon için Gereken Ölçütler

2.4.2.1 Açıklanan Varyans Puanı

Örnek veya veri kümesindeki verilerin yayılışının istatistiksel ölçüsü demektir. Algoritma başarısı sonuçla doğru orantılı olup, en yüksek sonuç 1.0 dır.

2.4.2.2 Ortalama Mutlak Hata (Mean Square Error)

Regresyon eğrisinin, bir dizideki noktalara yakınlığı hakkında bilgi vermektedir. Gerçek değerler ile tahmin değerleri arasındaki hataların karelerinin ortalamaları hesaplanmaktadır.

$$\frac{1}{n}\sum_{t=1}^{n}(e_t)^2 \tag{2.8}$$

2.4.2.3 Kare Ortalamalarının Karekökü (RMSE)

Ortalama mutlak hatanın karekökü anlamına gelmektedir.

$$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2} \tag{2.9}$$

2.4.2.4 R^2-Score

Doğrusal regresyon için uygunluk ölçüsüne denilmektedir. Bu değerin yüksek olması, veri setinin başarılı olduğunu gösteriyor olabileceği gibi aşırı öğrenme yapmış olabileceğini ve sonuçların yanıltıcı olabileceğini de gösteriyor olabilmektedir. Aşırı öğrenmelerde sonuç yüksek olsa bile, öğrenme başarısı yüksek olamayabilmektedir.

2.5 Denetimli Öğrenme Algoritmaları

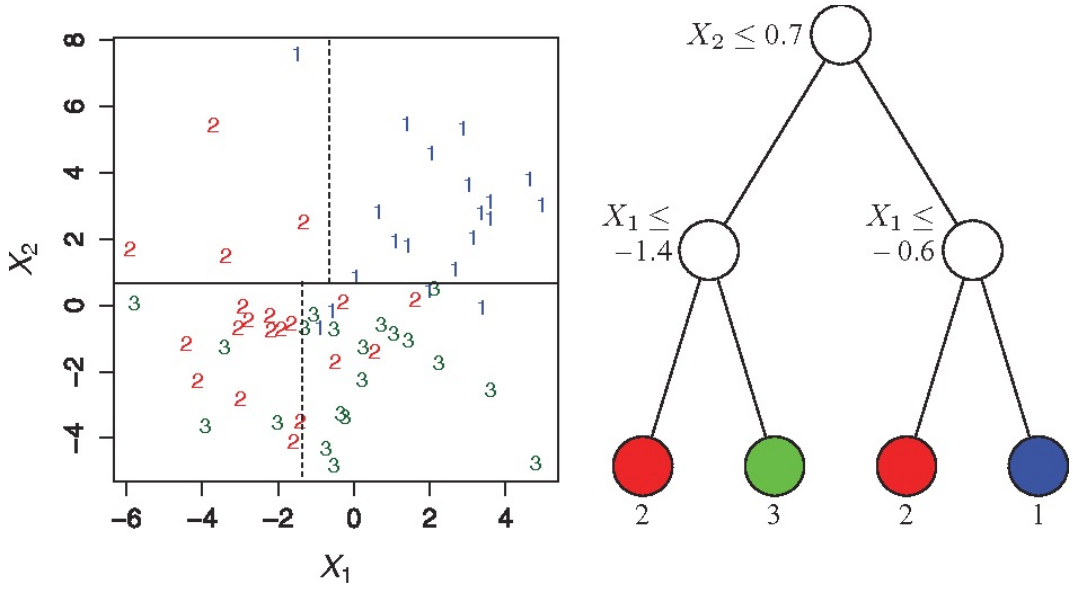
Bu bölümde, en yaygın olarak kullanılan denetimli makine öğrenmesi algoritmalarından bahsedilmektedir.

2.5.1 Karar Ağaçları

Ağaç yapılarının oluşturulmasında kullanılan kuralların anlaşılabilir ve sade olması sebebiyle, son yıllarda literatürde yaygın kullanımı olan bir sınıflandırma ve örüntü tanımlama algoritmasıdır [12].

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları; kolay yorumlanabilirliği, veri tabanlarına entegre edilmelerinin kolay olması, daha güvenilirlikli olması sebepleri ile sınıflandırma yöntemlerinde kullanımı en yaygın tekniktir [13].

Bu yöntem, ismi gibi daha çok karar verme mekanizmalarında, örneğin; bir hastalık tanısı koymak, alışverişlerde müşterinin profiline bağlı olarak satın alma durumunu tahmin etmek vb. durumlarda kullanılabilmektedir.



Şekil 2.2 Karar Ağacı Yapısı [14]

Şekil 2.2'deki gibi bir grafik olsun. Değerler bir ağaç yapısına göre sınıflandırılmak istenirse eğer, sağ taraftaki gibi bir karar ağacı ortaya çıkmaktadır. Artık bir veri sınıflandırılmak istendiği zaman bu ağaç yapısına bakılarak hangi sınıfa ait olduğu bulunabilmektedir.

Örneğin; $x_1 = 2$ ve $x_2 = 4$ verileriyle giriş yapıldığında, bu karar ağacına göre bu değerlerden çıktının mavi renk sınıfına ait olduğu söylenebilmektedir. Veri sayısı arttıkça ağacın yapısı değişerek çok daha fazla dallı hale gelebilmektedir.

Bu algoritma, çizilen sınırlar içerisine gelen değerlerin; regresyon problemlerinde sınır sayısı sayılamaz olduğundan kendisini değilde ortalamasını, sınıflandırma problemlerinde ise sınır sayısı sayılabilir olduğundan kendisini vermektedir. Dolayısıyla, regresyon problemlerinden daha çok sınıflandırma problemlerinin çözümünde kullanılmaktadır.

Sınırların çizilmesi işleminde, ilk başta entropi hesapları yapılmaktadır. Entropi işlemi, verilerin belirsizliğinin ölçüsü anlamına gelmektedir.

$$H = -\sum p(x) \cdot log_2[p(x)] \tag{2.10}$$

Denklem 2.10'da yer alan $p(x)$ belirli bir sınıfa ait veri grubunun yüzdesini, H değeri ise entropiyi belirtmektedir. Daha sonra, karar ağacının entropi sonucunu en aza indiren bölünmeler yapması istenmektedir. Bu bölünme işlemi için de Denklem 2.11'deki gibi *Information Gain (Bilgi Kazancı)* denklemi kullanılmaktadır.

$$Gain(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} \cdot H(V) \tag{2.11}$$

Burada; S orjinal veri kümesi, D kümenin bölünmüş bir parçası, V ise birleşimleri S'yi oluşturan ayrık alt kümeleri anlamına gelmektedir [14].

2.5.2 Naive Bayes Sınıflandırmaları

Bir değişkenin hangi olasılıkla hangi sınıfa ait olduğunu hesaplamaktadır. Burada iki önemli kabul vardır [15];

1. Tüm değişkenlerin önem derecesi aynıdır.
2. Değişkenler birbirinden bağımsızdır.

2.5.2.1 Bayes Teoremi

Bir sonucun sebebini bulurken koşullu olasılıktan faydalanarak sonucun hangi olasılıkla hangi sebepten kaynaklandığını bulma işleminde kullanılmaktadır.

$$P(A \backslash B) = \frac{P(B \backslash A) \cdot P(A)}{P(B)} \tag{2.12}$$

Örneğin; Bir hastaneye başvuran bir grup insan olsun. Hastaneye başvuranlar kadın ve erkek olmak üzere ayrıldığı zaman, Tablo 2.2'deki gibi veri havuzu oluşmaktadır;

Tablo 2.2 Hastaneye Başvuran Sayısı ve Hasta Oranı

	Kadın	**Erkek**
Hastaneye Başvuran Sayısı	1000	3000
Toplam Hasta Oranı	%5	

Tablo 2.2'den yararlanarak bulunması gereken sonuç; bir gün içinde hastaneye başvuran ve hasta tanımlananlardan kadın olma ihtimali nedir?

Bir günde hastaneye başvuran toplam 4000 kişi içerisinden, 1000 tanesi kadın 3000 tanesi erkek olsun. 4000 başvuran içerisinden %5 oran ile 200 hasta teşhisinin konulmuş olduğu ve hastaların cinsiyete göre dağılımının da eşit olduğu kabul

edilirse, 100 kişi kadın ve 100 kişi erkek olmuş olur. O halde hastaların kadın olma olasılığı;

$$\frac{100}{1000} = \%10 \tag{2.13}$$

Denklem 2.13'deki gibi hesaplanmış olur. Örnek tablolaştırılırsa, Tablo 2.3 elde edilmiş olunur.

Tablo 2.3 Sonuç Değerleri Tablosu

	İfade	**Değeri**
Hastaneye Başvuranlarda Kadın Oranı	P(K)	0,25
Kadınlarda Hasta Oranı	P(K|H)	0,5
Hastaneye Başvuranın Hasta Olma İhtimali	P(H)	0,05
Hastanın Kadın Olma İhtimali	P(H|K)	0,1

Tablo 2.3'de gösterilen ifadeler, Denklem 2.12'de yerine konulursa;

$$P(H|K) = \frac{P(K|H) \cdot P(H)}{P(K)} \tag{2.14}$$

Denklem 2.14 elde edilmiş olunur ve değerler Denklem 2.14'de yerine konulduğunda;

$$\frac{0,5 \cdot 0,05}{0,25} = \%10 \tag{2.15}$$

Denklem 2.15 sonucu elde edilir.

2.5.2.2 Bayes Sınıflandırıcı

Naive Bayes Sınıflandırıcı bir veri setinin bağımsız değişkenlerine bakarak, veriyi o değerlere yakın olan değerlere göre sınıflandırmaktadır. Bu işlemi, sınıflandırmak istediği kategoriye göre yapmaktadır. Öncelikle bir kategori belirlenmekte, verinin o kategoriye ait olma ihtimali ile olmama ihtimalinin toplamı bire eşit olacağı için, verinin tüm kategorilere göre ihtimali hesaplanmakta ve hangi ihtimal yüksekse o kabul edilmektedir.

İlk olarak, belirlenen kategoriye ait veri sayısının toplam veri sayısına olan

oranı bulunmaktadır. Eğitim veri setinde, sınıflandırılacak verinin bağımsız değişkenlerine yakın değerler içeren bir daire çizilerek o dairede bulunan veri sayısının toplam veri sayısına oranı bulunmaktadır. Daire içerisindeki belirlenen kategoriye ait veri sayısının, tüm verideki o kategoriye ait veri sayısına oranı bulunmaktadır.

Kategoriye ait veri sayısının toplam veri sayısına oranı ile daire içerisindeki kategori veri sayısının tüm verideki kategori veri sayısına oranı çarpılarak, daire içerisindeki veri sayısının tüm veri sayısına olan oranına bölünmekte ve olasılık hesaplanmaktadır. Aynı işlemler diğer kategoriler için de teker teker hesaplanarak ihtimaller içerisinden yüksek olan sınıfa dahil edilmektedir.

Örneğin; Hataneye başvuran bir grup insanın kan değerleri (bağımsız değişkenleri) ile hasta yada sağlıklı olduğu bilgileri (bağımlı değişkenleri) elimizde olsun. Hasta olma ihtimali ile sağlıklı olma ihtimalinin toplamı bire eşit olacağından, yeni bir insanın kan değerleri bilgileri ile Naive Bayes Sınıflandırıcısı kullanılarak hasta yada sağlıklı olduğu şu şekilde tahmin edilir;

Kişinin önce hasta olma olasılığı, sonra da sağlıklı olma olasılığı hesaplanarak, büyük olan oran hangi oran ise kişi o sınıfa dahil edilmektedir.

1. Burada kan değerleri bilgisi olmadığından, hasta olanların veri sayısının toplam veri sayısına oranı bulunur.

$$\text{P(H)} = \frac{\text{Hasta Sayısı}}{\text{Toplam Kişi}}$$

2. Hastaneye başvuran yeni bir kişinin tanısını koymaya karar verdiğimizde, o kan değerlerine yakın (o değerlere benzeyen) değerleri içeren bir daire çizilerek o dairede bulunan veri sayısının toplam veri sayısına olan oranı bulunur.

$$\text{P(D)} = \frac{\text{Dairedeki Veri Sayısı}}{\text{Toplam Veri Sayısı}}$$

3. Daire içerisindeki hasta verileri sayısının toplam hasta verileri sayısına oranı bulunur.

$$\text{P(D|H)} = \frac{\text{Dairedeki Hasta Veri Sayısı}}{\text{Toplam Hasta Veri Sayısı}}$$

4. 1.maddedeki oran (P(H)) ve 3. maddedeki oran (P(D|H)) çarpılarak 2.maddedeki orana (P(D)) bölünür ve hasta olma olasılığı bulunur.

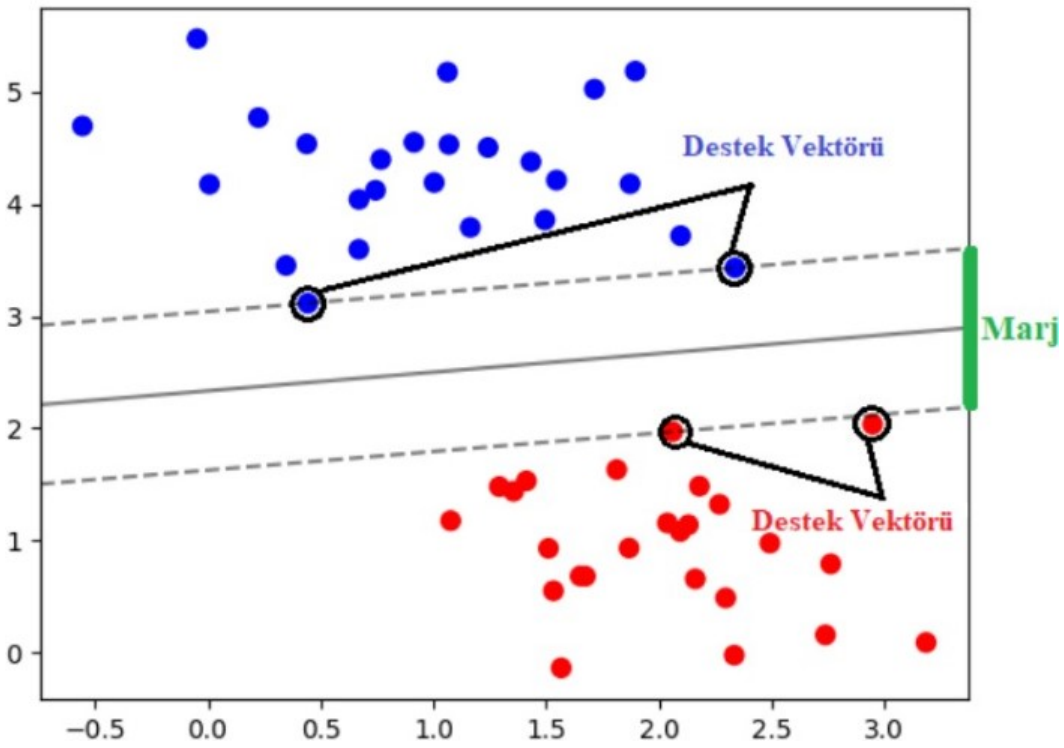
$$\text{P(Hasta)} = \frac{\text{P(H)} \cdot \text{P(D|H)}}{\text{P(D)}}$$

5. Daha sonra ilk dört maddedeki işlemlerin aynısı sağlıklı olanlar için tekrarlanarak, aynı kişinin sağlıklı olma olasılığı hesaplanır.

6. İki olasılık karşılaştırıldığında, hangi olasılık daha yüksek ise kişi o sınıfa (sağlıklı yada hasta) yerleştirilir.

2.5.3 Destek Vektör Makineleri (DVM)

Genelleme performansı bakımından yüksek makine öğrenimi algoritmalarından olup, sınıflandırma problemlerini kareli optimizasyon problemine dönüştürerek çözmekte ve bu durum öğrenmedeki işlem sayısını azaltarak algoritmayı hızlandırmaktadır [16]. Bu algoritmanın hedefi, N özellikli bir veri kümesinde, verileri belirgin olacak şekilde gruplandırabilen bir hiperdüzlem belirlemektir. Bunun için, Şekil 2.3'deki gibi grupların birbirlerine uzaklıklarının en az olduğu noktalar destek noktaları olarak, grupları birbirlerinden ayırmak için hiper düzlemlerin en optimumunu bulmaya yarayan destek konumları karar noktaları olarak, bu noktaların meydana getirdiği boşluktaki vektörler ise destek vektörleri olarak tanımlanmaktadır.

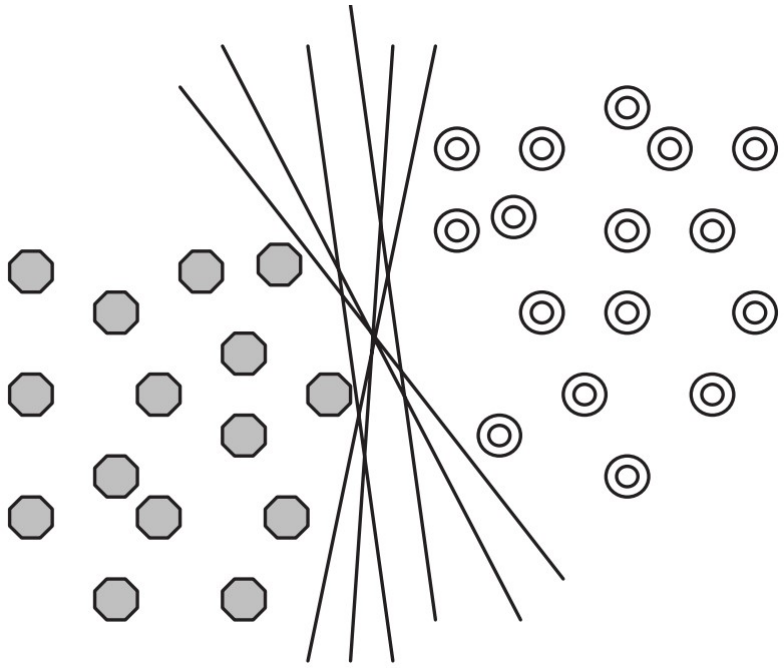


Şekil 2.3 Destek Vektör Makineleri

2.5.3.1 Doğrusal Ayrılabilen Veriler İçin DVM

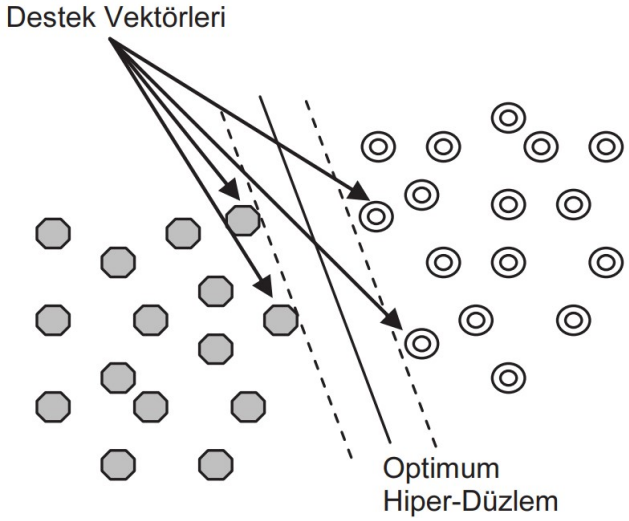
Eğitim verileri ile bir karar fonksiyonunu belirledikten sonra, {+1,-1} sınıf etiketleri yardımıyla iki gruba ait verileri belirlenen karar fonksiyonu ile Şekil 2.4'deki gibi

birbirinden ayırmaktadır.



Şekil 2.4 Çizilebilecek Hiper Düzlemler [17]

Bu ayırma işlemine en uygun olan hiper-düzlemin belirlenebilmesi için gereken noktalar destek vektörleridir ve bu destek vektörlerinden yararlanarak bu düzleme paralel iki hiper düzlem Şekil 2.5'deki gibi belirlenmektedir.



Şekil 2.5 En Uygun Hiper Düzlem [17]

N boyutlu vektör uzayından bir eleman ($x \in R^N$), ağırlık vektörü (w) ve eğilim değeri (b) kullanılarak $y \in \{-1, +1\}$ değeri Denklem 2.16'daki gibi elde edilmektedir.

$$w.x + b = y \tag{2.16}$$

w değerinin minimum olması, hiper düzlem sınırını maksimum yapmakta ve bu durum da Denklem 2.17'de verilen kısıtlı optimizasyon problemi ile çözülmektedir.

$$min \left[\frac{1}{2} \parallel w \parallel^2 \right] \tag{2.17}$$

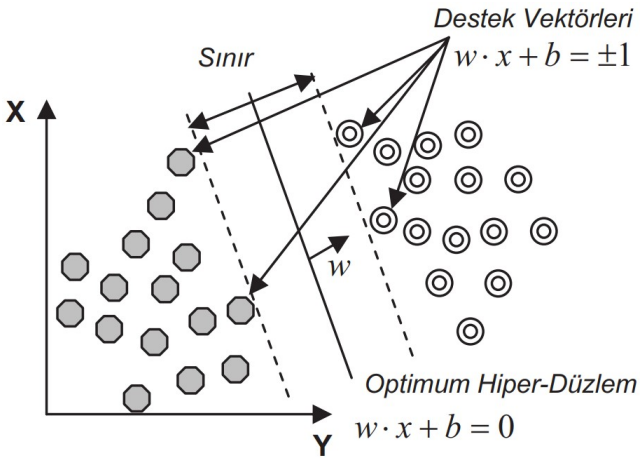
Kısıtlar, Lagrange Denklemi 2.18'deki gibi çözülerek karar fonksiyonu oluşturulmakta,

$$y_i(w.x_i + b) - 1 \geq 0 \text{ ve } y_i \in \{1, -1\} \tag{2.18}$$

Oluşturulan karar fonksiyonu Denklem 2.19 yardımı ile çözülmektedir.

$$f(x) = sign \sum_{i=1}^{k} \lambda_i [y_i(w.x_i + b)] \tag{2.19}$$

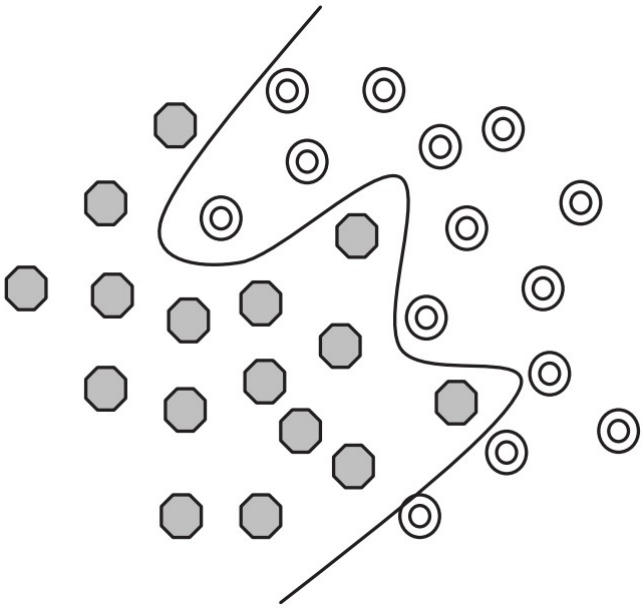
Denklem 2.19'dan yararlanarak Şekil 2.6'daki gibi en uygun Hiper-Düzlem bulunmaktadır.



Şekil 2.6 En Uygun Hiper Düzlemin Belirlenmesi [17]

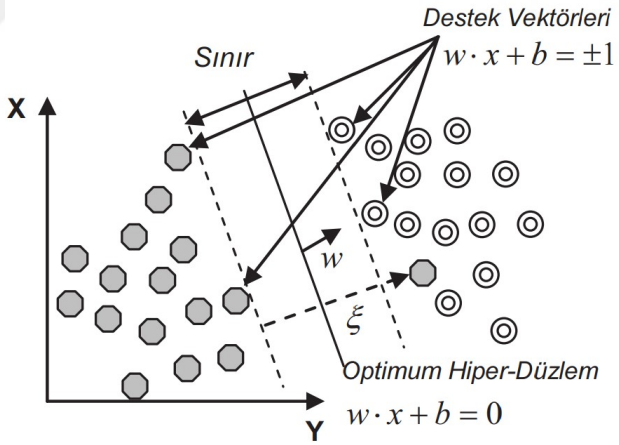
2.5.3.2 Doğrusal Ayrılamayan Veriler İçin DVM

Eğitim verileri içerisinden bazı verilerin Şekil 2.7'deki gibi hiper düzlemlerin en optimum olanıyla ayrışamadığı durumlarda bir problem oluşmaktadır.



Şekil 2.7 Doğrusal Ayrılamayan Veriler için Hiper Düzlemin Belirlenmesi [17]

Bu problemin çözümü için Şekil 2.8'deki gibi pozitif yapay bir (ξ_i) değişkeni tanımlanmaktadır.



Şekil 2.8 Doğrusal Ayrılamayan Veriler için Optimum Hiper Düzlemin Belirlenmesi [17]

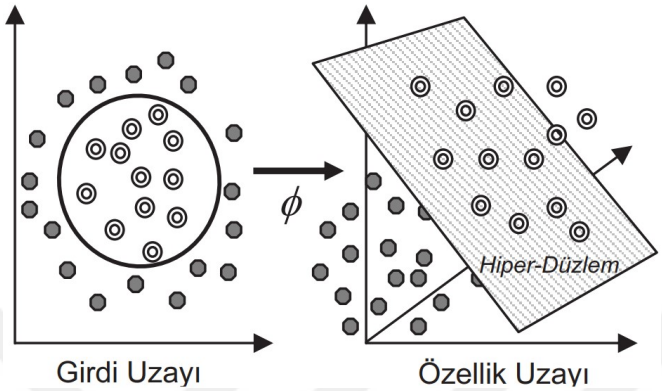
Bu durumda Denklem 2.20'deki gibi bir kernel fonksiyonuna ihtiyaç duyulmaktadır.

$$K(x_i, x_j) = \varphi(x).\varphi(x_j) \tag{2.20}$$

Kernel fonksiyonundan yararlanarak Denklem 2.21'deki gibi karar fonksiyonu elde edilmektedir.

$$f(x) = sign\left(\sum_i \alpha_i y_i \varphi(x).\varphi(x_i) + b\right) \tag{2.21}$$

Karar fonksiyonundan yararlanarak Şekil 2.9'daki gibi verilerin yüksek boyutta doğrusal ayrılabilmesi sağlanmaktadır.



Şekil 2.9 Kernel Fonksiyonu ile Verinin Boyutunun Yükseltilmesi [17]

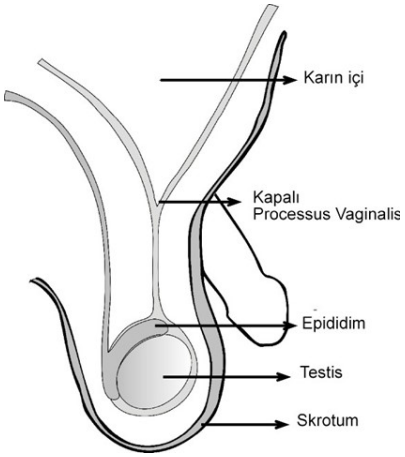
3
BAŞLICA İNGUİNAL (KASIK) KANAL PATOLOJİLERİ

Kasık bölgesi, kalçanın öndeki kemikli çıkıntısı olan spina iliaka anterior süperior seviyesinin alt bölümündeki karın ön kısmını tanımlamaktadır [18]. Çocuk cerrahisi tarafından en sık karşılaşılan patolojiler bu bölgede yer almakta olup hidroselden testis tümörüne kadar çeşitlilik göstermektedir [19].

İnguinal Kanal: Karın ön duvarındaki kaslar arasında oluşan karın boşluğu ile skrotum arasında uzanan bir kanaldır [20].

Testis: Embriyolojik dönemde karın içerisinde, henüz gelişmekte olan böbreklerin hemen altında oluşmaya başlayan ve doğuma yakın dönemde ya da doğumu takiben skrotuma yerleşen oval şeklinde sperm üretme organıdır [21].

Skrotum: Doğuma yakın dönemde ya da doğumu takiben testislerin yerleştiği ve sperm üretebilmek için sıcaklığın uygun olduğu yumurtalık torbasıdır [22].



Şekil 3.1 Normal İnguinal Kanal

Testislerin görevlerini normal bir şekilde yapabilmeleri için, vücut ısısından $3,5 \pm 0,5$ derece daha düşük bir ortamda bulunmaları gerekmektedir. Ancak karın

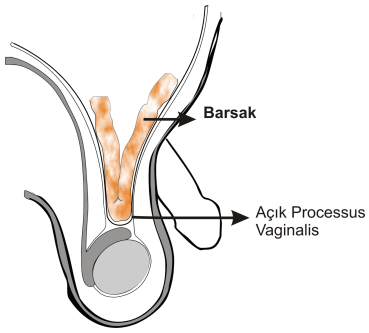
içerisindeki $37°C$ sıcaklığın, testislerin içerisindeki sperm hücrelerinin sağlıklı gelişmeleri için çok yüksek olmasından dolayı, fetal dönemde böbrekler yukarı doğru (olmaları gereken yere) çıkarken testisler de kasık kanalının girişine doğru göç etmektedirler [1]. Hamileliğin son 2 ayında ise salgılanan bazı hormon ve faktörlerin etkisiyle yeniden harekete geçen testisler, her iki kasıkta ayrı ayrı oluşan iki kanaldan geçip, sıcaklığı daha uygun olan skrotuma ($33°C$) doğuma yakın dönemde ya da doğumu takiben yerleşmiş olur.

3.1 İnmemiş Testis

Testislerin skrotuma inerken geçtiği kanallar, normalde doğuma birkaç gün kala vazifeleri bittiği için kapanmakta ve böylece karın içi ile dışı arasında bağlantı kalmamaktadır. Ancak bazen skrotuma yerleşmesi gereken testisler bazı hormon ve faktörlerin yetersiz düzeyde salgılanması sebebiyle inişlerini tamamlayamamakta ve inişini tamamlayamayan testislerden dolayı bu kanallar kapanamamaktadır. Karın içinde ya da bu kanallarda kalan ve keseye inemeyen testislere 'İnmemiş Testis' denmektedir.

3.2 İnguinal Herni

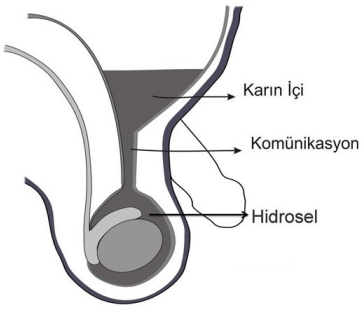
Kasık kanalını karın boşluğundan ayıran zarın kapanmamasından dolayı karın boşluğu ile skrotum arasındaki ilişkinin devam etmesi ve bu açıklıktan karın içi yapıların skrotuma ilerlemesi 'İnguinal Herni (Kasık Fıtığı)' dir.



Şekil 3.2 İnguinal Herni

3.3 Hidrosel

Kasık bölgesinde karın boşluğu ile skrotumu ayıran zarın tam kapanamaması yada geç kapanmasından dolayı karın içindeki sıvının testisin etrafını saran zarların arasında birikmesi 'Hidrosel (Su Fıtığı)' dir.



Şekil 3.3 Hidrosel

3.4 Tam Kan Sayımı Parametreleri

Bir insanın kanında, vücudun oksijeni taşıma, yenilenme, pıhtılaşma, bağışıklık vb. gibi çeşitli görevlerinden sorumlu, birbirinden ayrı hücresel bileşenler bulunmaktadır. Kanın dört ana bileşenleri; plazma, kırmızı kan hücreleri (eritrosit), trombositler ve beyaz kan hücreleridir (lökosit) [23].

Tam kan sayımı testi ise bu bileşenlerin sayı ve özellikleri ile insanın sağlık durumu hakkında bilgi sağlayan bir testtir.

- **Beyaz Kan Hücresi (White Blood Cell) Sayımı:** Lökosit olarak da adlandırılan beyaz kan hücreleri, vücudun savunma sistemi ve vücudun bağışıklık sistemi için gereken hücre sayısını göstermektedir.
  - **Lenfositler (LY):** Bakteri ve virüs gibi zararlılara karşı vücudun direncini artırmak için antikor üretmektedirler. Testteki değer ise beyaz kan hücrelerindeki lenfosit sayısını göstermektedir.
  - **Nötrofiller (NE):** Sayı olarak diğer beyaz kan hücrelerinden daha fazla olan nötrofiller bakterileri öldürerek sindirmekte ve vücudun enfeksiyonlara karşı savaşında görev yapmaktadır. Testteki değeri ise beyaz kan hücrelerindeki nötrofil sayısını göstermektedir.
  - **Monositler (MO):** Ömrü diğer beyaz kan hücrelerinden daha uzun olan monositler bakterilerin ortadan kaldırılmasında görev yapmaktadırlar. Testteki değer ise beyaz kan hücrelerindeki monosit sayısını göstermektedir.
  - **Eozinofiller (EO):** Parazitleri öldürerek alerjiye bağlı olarak gelişen iltihabı önlemektedirler. Ayrıca kanser hücrelerinin öldürülmesinde de kullanılmaktadırlar. Testteki değeri ise beyaz kan hücrelerindeki eozinofil sayısını göstermektedir.

  - **Bazofiller (BA):** Kanda enfeksiyon gördüklerinde bağışıklık sistemine haber vermektedirler. Testteki değeri ise beyaz kan hücrelerindeki bazofil sayısını göstermektedir.

- **Kırmızı Kan Hücresi (Red Blood Cell) Sayımı:** Eritrosit olarak da adlandırılan kırmızı kan hücreleri, vücudun oksijen taşıyan ve karbondioksiti uzaklaştıran hücre sayısını göstermektedir.

  - **Hemoglobin (HGB):** Kırmızı kan hücrelerindeki oksijen taşıyan protein sayısını göstermektedir.
  - **Hematokrit (HCT):** Kırmızı kan hücrelerinin kan içindeki oranını göstermektedir.
  - **Ortalama Korpuskular Hacim (Mean Corpuscular Volume):** Kırmızı kan hücrelerinin boyutlarını göstermektedir.
  - **Ortalama Korpuskular Hemoglobin (Mean Corpuscular Hemoglobin):** Kırmızı kan hücrelerinin her birindeki hemoglobin miktarını göstermektedir.
  - **Ortalama Korpuskular Hemoglobin Konsantrasyonu (Mean Corpuscular Hemoglobin Concentration):** Kırmızı kan hücrelerinin her birindeki hemoglobin miktarının kırmızı kan hücresinin kütlesine oranını göstermektedir.
  - **Eritrosit dağılım genişliği (RDW):** Kırmızı kan hücrelerinin boyutlarının ne kadar farklı olduğunu göstermektedir.

- **Trombosit (Platelet) Sayımı:** Kanın pıhtılaşmasını sağlayan hücre sayısını göstermektedir.

  - **Ortalama Trombosit Hacmi (Mean Platelet Volume):** Trombositlerin ortalama büyüklüğünü göstermektedir.
  - **Trombosit Dağılım Genişliği (Platelet Distribution Width):** Aynı anda kanda bulunan trombositler arasındaki boyut farklılığını ölçmektedir.

4
BAŞLICA İNGUİNAL KANAL PATOLOJİLERİNİN TAHMİN EDİLMESİ

2021.58.02.21 numaralı etik kurulu onayıyla, 2014-2023 yılları arasında Tekirdağ Namık Kemal Üniversitesi (TKNÜ) Hastanesi Çocuk Cerrahisi ABD Kliniğine ve 2014-2016 yılları arasında Bahçelievler Devlet Hastanesi (BDH) Çocuk Cerrahisi Kliniğine başvuran kişilerin verileri geriye dönük olarak toplanmıştır. Hastane sistemlerinden temin edilen veriler, her yıl için ayrı ayrı 9 poliklinik verisi ve 10 servis verisi olmak üzere 19 adet excel dosyası, her yılın laboratuvar sonuçları ve her yılın radyoloji sonuçları olmak üzere 2 adet excel dosyası ve her yılın BDH verisi olmak üzere toplam 22 adet Excel dosyalarından oluşmaktadır. Bu excel dosyaları; Hasta Adı Soyadı, Cinsiyet, Yaş, Barkod Numarası, Dosya No, Başvuru No, Doğum Tarihi, Başvuru Tarihi, Alan Adı, Açıklama, Tanı, Testte Ölçülen Değerler, Test Kabul ve Sonuç Tarihleri sütunlarından oluşmaktadır.

Çalışmamızın amacı; hastaneye başvuran kişilerin tam kan sayımı (hemogram) verileri, yaş ve cinsiyet verileri ile kişinin inmemiş testis, inguinal herni ve hidrosel hastalıklarına sahip olup olmayacağı arasındaki ilişkiyi incelemektir. İnceleme için hastaların hemogram verileri ile yaş ve cinsiyet bilgileri arasındaki ilişkiden yola çıkarak makine öğrenmesi algoritmaları kullanılarak sınıflandırıldı. Bu kapsamda öncelikle, veriler için yapılması gereken veri temizleme işlemleri gerçekleştirildi ve verilere ortak sütunlar üzerinden birleştirme işlemleri uygulandı.

Çalışmamızın devamında sırasıyla verilerin düzenlenmesi, istatistiksel analiz ve makine öğrenmesi algoritmalarıyla hastalıkların tahmin edilmesi aşamaları anlatılmaktadır.

4.1 Verilerin Düzenlenmesi

1. Alınan veriler kişiye göre değil değişkene göre satırlaştırılmış olup, okunabilirliğin ve işlenebilirliğin kolay olması açısından aşağıdaki işlemler

ile bir kişinin verileri bir satırda olacak şekilde birleştirildi.

(a) Hemogram değerlerinin her bir değeri için dokümanlarda aynı sütunda satırlaştırılmış şekilde yer alan değerler aynı satırda olacak şekilde sütunlaştırılarak düzenlendi.

(b) Bu işlem; poliklinik verilerinin her bir yılı için, servis verilerinin her bir yılı için, radyoloji verileri için ve laboratuvar verileri için ayrı ayrı uygulandı. BDH verileri alındığında bu şekilde düzenlenmiş olduğundan herhangi bir düzenleme işlemi yapılmayarak temizleme işlemlerinden sonra verilere eklendi.

(c) Uygulanan bu işlemlerden sonra, dokümanlarda kişiye ait tekil değer içeren ortak sütunlardan yararlanılarak excel dosyaları birleştirildi.

2. Birleştirilen excel dosyasında, ölçüm yapılacak hastalığı içeren ve içermeyen kişiler tanı sütunundan yararlanılarak belirlendi.

3. Belirleme işleminden sonra TKNÜ verilerinin tanılara göre kişi sayısı dağılımı Tablo 4.1'de gösterildiği gibidir;

Tablo 4.1 TKNÜ Verilerinin Tanı ve Kişi Dağılımları

Tanılar	**Kişi Sayısı**
İnmemiş Testis	447
İnguinal Herni	465
Hidrosel	142
İnmemiş Testis ve İnguinal Herni	10
İnmemiş Testis ve Hidrosel	18
İnguinal Herni ve Hidrosel	2
Üç hastalığa da sahip olan	1
Kontrol (Bu üç hastalığa da sahip olmayan)	1461
Toplam	2546

4. Bu kategorilerden, birden fazla hastalığa sahip olan 31 kişinin verileri ayırt ediciliğin artması açısından veri setinden çıkartılarak, inmemiş testis, inguinal herni, hidrosel ve kontrol (bu üç hastalığa da sahip olmayan) gruplarındaki 2515 kişinin verisi olmak üzere 4 kategoriye indirgendi.

5. Bu işlemden sonra; kontrol kategorisinden 9 kişi, inguinal herni kategorisinden 1 kişi, inmemiş testis kategorisinden 1 kişi olmak üzere toplamda 11 kişinin verisi kan değerleri eksik olması sebebiyle veri setinden temizlendi ve 2504 kişinin verileri ile devam edildi.

6. Lenfosit (LY), Nötrofil (NE), Eozinofil (EO), Monosit (MO) ve Bazofil (BA) değerleri ile birlikte yüzde değerleri de bulunmaktadır. Birbirine bağımlı değişkenler olduğu için, yüzdelik değerler kalacak şekilde, diğer değerleri içeren sütunlar veri setinden çıkartıldı.

7. Veri setimizde 997 (%39,81) kişinin BA değeri olmadığından, kişileri veri setinden çıkarmak yerine LY+NO+EO+MO değerleri ile BA değerleri arasındaki korelasyon ilişkisinin %100 olmasından (Tablo 4.2) yola çıkarak BA sütunu veri setinden çıkartıldı.

Tablo 4.2 Beyaz Kan Hücresi Parametreleri Arasındaki Pearson Korelasyon Analizi

	LY+NO+EO+MO	BA
LY+NO+EO+MO	1	-1,000
BA	-1,000	1

8. 5 Kişinin kan değerlerinde makinenin ölçüm ayarlarından kaynaklı değerler 10 katı olarak gösterilmiş olup, bu durum aykırı değere yol açacağından dolayı veriler diğer verilerle aynı değerlere uyarlandı.

9. Veri seti, ölçüm yaparken işe yaramayacak olan verilerden ve kişisel verilerden ayıklanarak hemogram verileri, yaş ve cinsiyet sütunları kalacak şekilde düzenlendi.

10. BDH dosyasında 3185 kişinin verilerinden yüzdesel olmayan veri sütunları çıkartıldı.

11. BDH verilerinde; tanıları belirtilmeyen 6 kişi verisi, hemogram verileri olmayan kontrol kategorisinden 4 kişi ve inmemiş testis kategorisinden 1 kişi olmak üzere toplam 11 kişi veri setinden çıkartılarak 3174 kişinin verileri ile devam edildi.

12. BDH veri seti, ölçüm yaparken işe yaramayacak olan verilerden ve kişisel verilerden ayıklanarak, hemogram verileri, yaş ve cinsiyet sütunları kalacak şekilde indirgenerek düzenlenen TKNÜ verileriyle birleştirilebilir olması için sütunların sıralaması değiştirildi.

13. Son olarak BDH verilerinin tanılara göre kişi sayısı dağılımı Tablo 4.3'de gösterildiği gibidir;

Tablo 4.3 BDH Verilerinin Tanı ve Kişi Dağılımları

Tanılar	Kişi Sayısı
Kontrol	2712
İnmemiş Testis	109
İnguinal Herni	204
Hidrosel	149
Toplam	3174

14. Birleştirilen excel dosyasında, ortalama ve standart sapma değerleri hesaplaması için, cinsiyet sütununda olan erkek ve kadın değerleri sırasıyla 1 ve 0 olmak üzere değiştirilerek, sayısal hale getirildi.

4.2 İstatistiksel Analiz

Verilerin düzenlenmesi işlemleri tamamlandıktan sonra, 5678 kişinin, belirlenen kategorilere (inmemiş testis, inguinal herni, hidrosel ve kontrol) ve cinsiyete göre sayıca dağılımı Tablo 4.4'de gösterildiği gibidir;

Tablo 4.4 Tanılara Göre Cinsiyet Sayıları

	Kontrol	İnmemiş Testis	İnguinal Herni	Hidrosel	Toplam
Kadın	299	0	206	0	505
Erkek	3865	555	462	291	5173
Toplam	4164	555	668	291	5678

Tablo 4.4'e bakıldığında, kadın verileri toplam veri grubunun %8,89'unu oluştururken, bunlardan %59,2'si kontrol veri grubunda, %40,8'i ise inguinal herni veri grubundadır. İnmemiş testis ve hidrosel veri gruplarında kadın bulunmadığı için bu hastalıkların %100'ü erkeklerden oluşmaktadır. Erkekler ise %74,71'i kontrol, %10,73'ü inmemiş testis, %8,93'ü inguinal herni, %5,63'ü hidrosel olarak dağılmıştır. Veri setindeki değişkenlerin, kontrol ve hastalıklara ait alt sınıflara göre ortalama ve standart sapma değerleri Tablo 4.5'de gösterildiği gibidir.

Tablo 4.5 Kategorilere Göre Değişkenlere Ait Ortalama ve Standart Sapma Değerleri

Değişkenler	Kontrol	İnmemiş Testis	İnguinal Herni	Hidrosel
Yaş	4,57± 3,77	2,77±2,64	3,50±4,47	3,96±3,64
WBC	9,76±17,08	9,01±2,93	8,87±2,98	9,32±4,36
LY%	44,71±16,24	49,70±13,39	49,24±16,01	45,18±15,65
NE%	41,96±16,93	37,72±13,34	37,33±16,86	41,63±15,92
EO%	3,24±2,78	3,52±2,60	3,47±2,49	3,13±2,63
MO%	8,25±2,97	8,01±2,45	8,54±3,05	8,37±3,35
PLT	328,65±99,19	336,73±90,89	354,10±123,33	326,32±114,91
MPV	7,51±1,58	8,07±1,27	8,13±1,50	7,55±1,44
PDW	15,92±4,29	13,55±3,88	14,51±3,88	15,20±4,43
RBC	4,69±0,62	4,56±0,47	4,25±0,79	4,66±0,53
HCT	36,54±4,70	35,44±3,58	35,28±5,26	36,13±3,99
HGB	12,09±1,68	11,77±1,29	11,72±1,84	12,01±1,43
MCV	77,78±9,38	77,94±7,44	83,56±11,20	77,51±7,53
RDW-CV	13,34±2,17	14,53±2,31	13,62±2,14	13,79±2,30
MCH	25,76±3,51	25,89±2,72	27,75±3,77	25,80±2,90
MCHC	32,86±3,07	33,14±1,88	33,02±2,89	33,13±2,41

Hastaneye başvuranlardan sağlıklı, inmemiş testis, inguinal herni ve hidrosel tanılarına sahip olanlara ait değişkenler için Kruskal-Wallis H testi uygulanmıştır. Hesaplanan p değerleri 0,05'den küçük olan değişkenler istatistiki olarak anlamlı kabul edilmiştir. Tablo 4.6'daki değerlere bakıldığında sağlıklı, inmemiş testis, inguinal herni ve hidrosel tanılarına sahip olanların parametrelerinin ortalamaları arasında anlamlı istatistiksel farklar bulunmuştur.

Tablo 4.6 Tanılara Ait Değişkenler için p Değerleri

Değişkenler	p Değeri (Kruskal-Wallis H Testi)
Yaş	0,000
WBC	0,009
LY%	0,000
NE%	0,000
EO%	0,000
MO%	0,027
PLT	0,000
MPV	0,000
PDW	0,000
RBC	0,000
HCT	0,000
HGB	0,000
MCV	0,000
RDW-CV	0,000
MCH	0,000
MCHC	0,013

Verilere SMOTE algoritması uygulandıktan sonra 16656 verinin, belirlenen kategorilere (inmemiş testis, inguinal herni, hidrosel ve kontrol) ve cinsiyete göre sayıca dağılımı Tablo 4.7'de gösterildiği gibidir;

Tablo 4.7 SMOTE Algoritması Sonrası Tanılara Göre Cinsiyet Sayıları

	Kontrol	**İnmemiş Testis**	**İnguinal Herni**	**Hidrosel**	**Toplam**
Kadın	299	0	2047	0	2346
Erkek	3865	4164	2117	4164	14310
Toplam	4164	4164	4164	4164	16656

Tablo 4.7'ye bakıldığında, kadın verileri toplam veri grubunun %14,08'ini oluştururken, bunlardan %12,74'ü kontrol veri grubunda, %87,25'i ise inguinal herni veri grubundadır. İnmemiş testis ve hidrosel veri gruplarında kadın bulunmadığı için bu hastalıkların %100'ü erkeklerden oluşmaktadır. Erkekler ise %27'si kontrol, %29,09'u inmemiş testis, %14,79'u inguinal herni, %29,09'u hidrosel olarak dağılmıştır. Veri setindeki değişkenlerin, kontrol ve hastalıklara ait alt sınıflara göre ortalama ve standart sapma değerleri Tablo 4.8'de gösterildiği gibidir.

Tablo 4.8 SMOTE Algoritması Sonrası Kategorilere Göre Değişkenlere Ait Ortalama ve Standart Sapma Değerleri

Değişkenler	Kontrol	İnmemiş Testis	İnguinal Herni	Hidrosel
Yaş	4,57± 3,77	2,52±2,41	3,26±4,13	3,41±3,05
WBC	9,76±17,08	8,81±2,36	8,81±2,62	9,12±3,53
LY%	44,71±16,24	49,88±12,52	49,58±15,48	46±14,28
NE%	41,96±16,93	37,78±12,59	37,25±16,36	41,27±14,41
EO%	3,24±2,78	3,43±2,13	3,44±2,13	3,06±2,05
MO%	8,25±2,97	8±2,07	8,42±2,57	8,29±2,64
PLT	328,65±99,19	335,40±89,34	356,15±119,66	329,28±104,72
MPV	7,51±1,58	8,14±1,02	8,14±1,31	7,55±1,17
PDW	15,92±4,29	13,43±3,22	14,46±3,34	15,03±3,78
RBC	4,69±0,62	4,55±0,37	4,25±0,72	4,66±0,42
HCT	36,54±4,70	35,49±2,83	35,38±4,72	36,19±3,19
HGB	12,09±1,68	11,80±1,04	11,76±1,64	12,03±1,15
MCV	77,78±9,38	78,24±6,01	83,72±10,24	77,70±6,01
RDW-CV	13,34±2,17	14,59±1,92	13,63±1,85	13,86±1,88
MCH	25,76±3,51	26,02±2,23	27,81±3,41	25,84±2,32
MCHC	32,86±3,07	33,22±1,30	33,05±2,61	33,15±1,79

SMOTE algoritması uygulandıktan sonra; sağlıklı, inmemiş testis, inguinal herni ve hidrosel tanılarına ait değişkenler için Kruskal-Wallis H testi uygulanmıştır. Hesaplanan p değerleri 0,05'den küçük olan değişkenler istatistiki olarak anlamlı kabul edilmiştir. Tablo 4.9'daki değerlere bakıldığında sağlıklı, inmemiş testis, inguinal herni ve hidrosel tanılarına sahip olanların parametrelerinin ortalamaları arasında anlamlı istatistiksel farklar bulunmuştur.

Tablo 4.9 SMOTE Algoritması Sonrası Tanılara Ait Değişkenler İçin p Değerleri

Değişkenler	p Değeri (Kruskal-Wallis H Testi)
Yaş	0,000
WBC	0,000
LY%	0,000
NE%	0,000
EO%	0,000
MO%	0,000
PLT	0,000
MPV	0,000
PDW	0,000
RBC	0,000
HCT	0,000
HGB	0,000
MCV	0,000
RDW-CV	0,000
MCH	0,000
MCHC	0,000

4.3 Hastalıkların Sınıflandırma Modellerine Göre Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesi

Makine öğrenmesinde kullanılan veri setinin tanı ve kişi dağılımı Tablo 4.10'da gösterildiği gibidir;

Tablo 4.10 Makine Öğrenmesinde Kullanılacak Tanı ve Kişi Dağılımı

Tanılar	Kişi Sayısı	Eğitim Veri Sayısı	Test Veri Sayısı
Kontrol	4164	3123	1041
İnmemiş Testis	555	416	139
İnguinal Herni	668	501	167
Hidrosel	291	218	73
Toplam	5678	4258	1420

Veri seti, MatLab programının sınıflandırma öğreticisi (Classification Learner) aracı kullanılarak;

- Karar Ağaçları Algoritması
  - Fine Tree Modeli
  - Medium Tree Modeli
  - Coarse Tree Modeli
- Naive Bayes Sınıflandırmaları Algoritması
  - Gaussian Naive Bayes Modeli
  - Kernel Naive Bayes Modeli
- Destek Vektör Makineleri Algoritması
  - Quadratic Destek Vektör Makineleri Modeli
  - Cubic Destek Vektör Makineleri Modeli
  - Fine Gaussian Destek Vektör Makineleri Modeli
  - Medium Gaussian Destek Vektör Makineleri Modeli
  - Coarse Gaussian Destek Vektör Makineleri Modeli

algoritmaları için MatLab varsayılan parametre değerleri kullanılarak tahminlemeler yapıldı. Bu teknikleri uygularken, veriler rastgele olarak %75 eğitim için, %25 test için olacak şekilde ayrıldı. Eğitim verisine 5 katlamalı çapraz doğrulama uygulanarak modeller eğitildi. Eğitilen modeller ile yapılan testler sonucunda istenilen sonuçların, veri setinin dengesiz dağılımından kaynaklandığı düşünülerek, verilere *Dengesiz Veri Setlerinde Aşırı Örnekleme* işlemindeki SMOTE algoritması uygulandı.

SMOTE algoritması uygulandıktan sonra elde edilen yeni veri seti rastgele olarak %70 eğitim için, %30 test için olacak şekilde ayrıldı. Bu işlemlerden sonra veri setinin tanı ve kişi dağılımı Tablo 4.11'de gösterildiği gibidir;

Tablo 4.11 SMOTE Algoritmasından Sonra Makine Öğrenmesinde Kullanılacak Tanı ve Kişi Dağılımı

Tanılar	Kişi Sayısı	Eğitim Veri Sayısı	Test Veri Sayısı
Kontrol	4164	2915	1249
İnmemiş Testis	4164	2915	1249
İnguinal Herni	4164	2915	1249
Hidrosel	4164	2915	1249
Toplam	16656	11660	4996

Çalışmanın devamında asıl veriler ile SMOTE algoritması uygulandıktan sonra elde edilen verilerin; inmemiş testis, inguinal herni, hidrosel, hasta (bu üç tanısı olanları hasta olarak etiketleyerek), kontrol (bu üç hastalıktan herhangi birine sahip olmayan) kategorilerinin sınıflandırma kombinasyonlarına göre algoritmaların modellerinden; doğruluk (tüm verilerin doğru tahmin edilme) oranının en yüksek olduğu, duyarlılık (iki değişkenden birinin doğru tahmin edilme) oranının en yüksek olduğu ve özgüllük (iki değişkenden diğerinin doğru tahmin edilme) oranının en yüksek olduğu modeller için elde edilen sonuçlar verildi.

4.3.1 Kontrol ve Hasta Verilerini Sınıflandırma Modeli

4.3.1.1 Asıl Veriler

Kontrol ve hasta verilerinin sınıflandırılmasında, kontrol verilerinin ne kadarının kontrol ne kadarının hasta olarak, hasta verilerinin ne kadarının hasta ne kadarının kontrol olarak tahmin edildiği Tablo 4.12'de gösterildiği gibidir.

Tablo 4.12 Kontrol ve Hasta Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**Hasta**
Medium Gaussian SVM	**Kontrol**	1000	41
	Hasta	295	84
Lineer SVM	**Kontrol**	1041	0
	Hasta	379	0
Gaussian Naive Bayes	**Kontrol**	635	406
	Hasta	129	250

Karşılaştırma matrisinden elde edilen sonuçlar ile doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.13'de gösterildiği gibi elde edildi.

Tablo 4.13 Kontrol ve Hasta Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Gaussian SVM	%76,3	%96,1	%22,2
Lineer SVM	%73,3	%100	%0
Gaussian Naive Bayes	%62,3	%61	%66

Tablo 4.13'ye bakıldığında Medium Gaussian SVM modelinin; kontrol verilerinin %96,1'ini, hasta verilerinin ise %22,2'sini doğru tahmin etmesinden dolayı doğruluk oranı %76,3 olarak çıksa da, hasta veri grubunu doğru tahmin etme oranının düşük olmasından dolayı başarılı bir ayrım yapamadığı görülmektedir.

Lineer SVM modelinin tüm verileri kontrol veri grubunda tahmin ederek iki grup arasında ayrım yapamadığı görülmektedir.

Gaussian Naive Bayes modelinin doğruluk oranı %62,3 olarak diğer modellerden düşük çıksa da, kontrol veri grubunu %61 oranında ayırt edebilirken hasta veri grubunu %66 oranında doğru tahmin edebildiği ve daha dengeli bir ayrım yapabildiği görülmektedir.

4.3.1.2 SMOTE Verileri

Sentetik olarak üretilen verilerden kontrol ve hasta verilerinin sınıflandırılmasında, kontrol verilerinin ne kadarının kontrol ne kadarının hasta olarak, hasta verilerinin ne kadarının hasta ne kadarının kontrol olarak tahmin edildiği Tablo 4.14'de gösterildiği gibidir.

Tablo 4.14 Kontrol ve Hasta SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**Hasta**
Fine Gaussian SVM	**Kontrol**	1108	141
	Hasta	292	956
Medium Tree	**Kontrol**	1010	239
	Hasta	448	800
Cubic SVM	**Kontrol**	937	312
	Hasta	288	960

Karşılaştırma matrisinden elde edilen sonuçlar ile doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.15'de gösterildiği gibi elde edildi.

Tablo 4.15 Kontrol ve Hasta SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%82,7	%88,7	%76,6
Medium Tree	%72,5	%80,9	%64,1
Cubic SVM	%76	%75	%76,9

Tablo 4.15'e bakıldığında, kontrol veri grubunu hasta veri grubundan ayırma işleminde %82,7 oranla Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğunu, hasta veri grubunu kontrol veri grubundan ayırma işleminde ise %76,9 oranında Cubic SVM modelinin diğer modellere göre daha başarılı olduğu ancak kontrol veri grubunu hasta veri grubundan ayırma işleminde Fine Gaussian SVM modeli kadar başarılı olamadığı görülmektedir.

4.3.2 Kontrol ve İnmemiş Testis Verilerini Sınıflandırma Modeli

4.3.2.1 Asıl Veriler

Kontrol ve inmemiş testis verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.16'de gösterildiği gibidir.

Tablo 4.16 Kontrol ve İnmemiş Testis Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**İnmemiş Testis**
Fine Gaussian SVM	**Kontrol**	1040	1
	İnmemiş Testis	136	3
Lineer SVM	**Kontrol**	1041	0
	İnmemiş Testis	139	0
Kernel Naive Bayes	**Kontrol**	906	135
	İnmemiş Testis	70	69

Karşılaştırma matrisinden elde edilen sonuçlar ile doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.17'da gösterildiği gibi elde edildi.

Tablo 4.17 Kontrol ve İnmemiş Testis Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%88,4	%99,9	%2,2
Lineer SVM	%88,2	%100	%0
Kernel Naive Bayes	%82,6	%87	%49,6

Tablo 4.17'ya bakıldığında Fine Gaussian SVM modeli doğruluk oranı %88,4 olarak en başarılı gibi gözükse de, özgüllük oranı yani inmemiş testis verilerini kontrol verilerinden ayırma başarısı %2,2 olduğu için başarılı bir ayrım yapamadığı görülmektedir.

Kernel Naive Bayes modelinin kontrol verilerini inmemiş testis den %87 oranında ve inmemiş testis verilerini kontrol verilerinden %49,6 oranında ayırarak daha dengeli bir ayrım yapabildiği görülmektedir.

4.3.2.2 SMOTE Verileri

Sentetik olarak üretilen verilerden kontrol ve inmemiş testis verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.18'de gösterildiği gibidir.

Tablo 4.18 Kontrol ve İnmemiş Testis SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**İnmemiş Testis**
Fine Gaussian SVM	**Kontrol**	1174	75
	İnmemiş Testis	80	1169
Medium Gaussian SVM	**Kontrol**	972	277
	İnmemiş Testis	138	1111
Cubic SVM	**Kontrol**	5	1244
	İnmemiş Testis	0	1249

Karşılaştırma matrisinden elde edilen sonuçlar ile doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.19'de gösterildiği gibi elde edildi.

Tablo 4.19 Kontrol ve İnmemiş Testis SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%93,8	%94	%93,6
Medium Gaussian SVM	%83,4	%77,8	%89
Cubic SVM	%50,2	%0,4	%100

Tablo 4.19'e bakıldığında, hem kontrol veri grubunu inmemiş testis veri grubundan hem de inmemiş testis veri grubunu kontrol veri grubundan ayırma işleminde Fine Gaussian SVM modelinin diğer modellere göre daha iyi başarı yakaladığı görülmektedir.

4.3.3 Kontrol ve İnguinal Herni Verilerini Sınıflandırma Modeli

4.3.3.1 Asıl Veriler

Kontrol ve inguinal herni verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.20'da gösterildiği gibidir.

Tablo 4.20 Kontrol ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**İnguinal Herni**
Medium Gaussian SVM	**Kontrol**	1029	12
	İnguinal Herni	120	47
Fine Gaussian SVM	**Kontrol**	1041	0
	İnguinal Herni	163	4
Gaussian Naive Bayes	**Kontrol**	887	154
	İnguinal Herni	83	84

Karşılaştırma matrisinden elde edilen sonuçlara bakarak doğruluk, duyarlılık ve özgüllük değerleri Tablo 4.21'de gösterildiği gibi elde edildi.

Tablo 4.21 Kontrol ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Gaussian SVM	%89,1	%98,8	%28,1
Fine Gaussian SVM	%86,5	%100	%2,4
Gaussian Naive Bayes	%80,4	%85,2	%50,3

Tablo 4.21'ye bakıldığında Medium Gaussian SVM modelinin, verilerin %89,1'ini doğru tahmin ettiğini ancak inguinal herni verilerini kontrol verilerinden ayırma başarısını yakalayamadığı görülmektedir.

Fine Gaussian SVM modelinin, verilerin çoğunluğunu kontrol olarak tahmin etmesi ve kontrol verilerinin fazla olmasından kaynaklı %86,5 gibi yüksek bir doğruluk oranına sahip olması başarılı bir ayrım yapabildiği değil, aksine inmemiş testisi istenilen derecede ayıramadığı için başarılı bir ayrım yapamadığı görülmektedir.

Gaussian Naive Bayes modelinin doğruluk oranı %80,4 ile diğer modellere göre daha düşük gibi gözükse de, kontrol veri grubu ile inmemiş testis veri grubunu birbirinden daha dengeli ayırabildiği görülmektedir.

4.3.3.2 SMOTE Verileri

Sentetik olarak üretilen verilerden kontrol ve inguinal herni verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.22'de gösterildiği gibidir.

Tablo 4.22 Kontrol ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**İnguinal Herni**
Fine Gaussian SVM	**Kontrol**	1199	50
	İnguinal Herni	128	1121
Coarse Gaussian SVM	**Kontrol**	1130	119
	İnguinal Herni	411	838
Cubic SVM	**Kontrol**	1065	184
	İnguinal Herni	172	1077

Karşılaştırma matrisinden elde edilen sonuçlara bakarak doğruluk, duyarlılık ve özgüllük değerleri Tablo 4.23'de gösterildiği gibi elde edildi.

Tablo 4.23 Kontrol ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%92,9	%96	%89,8
Coarse Gaussian SVM	%78,8	%90,5	%67,1
Cubic SVM	%85,7	%85,3	%86,2

Tablo 4.23'ye bakıldığında, hem kontrol veri grubunu inguinal herni veri grubundan hem de inguinal herni veri grubunu kontrol veri grubundan ayırma işleminde Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğu görülmektedir.

4.3.4 Kontrol ve Hidrosel Verilerini Sınıflandırma Modeli

4.3.4.1 Asıl Veriler

Kontrol ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.24'de gösterildiği gibidir.

Tablo 4.24 Kontrol ve Hidrosel Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**Hidrosel**
Coarse Tree	**Kontrol**	1041	0
	Hidrosel	73	0
Coarse Gaussian SVM	**Kontrol**	1041	0
	Hidrosel	73	0
Fine Tree	**Kontrol**	999	42
	Hidrosel	64	9

Karşılaştırma matrisinden elde edilen sonuçlara bakarak doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.25'de gösterildiği gibi elde edildi.

Tablo 4.25 Kontrol ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Coarse Tree	%93,4	%100	%0
Coarse Gaussian SVM	%93,4	%100	%0
Fine Tree	%90,5	%96	%12,3

Tablo 4.25'e bakıldığında modellerin doğruluk oranları yüksek çıksa da, hidrosel verilerini kontrol verilerinden ayırma oranlarının düşük çıkmasından dolayı başarılı bir ayrım yapamadıkları görülmektedir.

4.3.4.2 SMOTE Verileri

Sentetik olarak üretilen verilerden kontrol ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.26'de gösterildiği gibidir.

Tablo 4.26 Kontrol ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Kontrol	**Hidrosel**
Fine Gaussian SVM	**Kontrol**	1210	39
	Hidrosel	74	1175
Coarse Tree	**Kontrol**	1068	181
	Hidrosel	668	581
Cubic SVM	**Kontrol**	971	278
	Hidrosel	118	1131

Karşılaştırma matrisinden elde edilen sonuçlara bakarak doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.27'da gösterildiği gibi elde edildi.

Tablo 4.27 Kontrol ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%95,5	%96,9	%94,1
Coarse Tree	%66	%85,5	%46,5
Cubic SVM	%84,1	%77,7	%90,6

Tablo 4.27'ya bakıldığında, hem kontrol veri grubunu hidrosel veri grubundan hem de hidrosel veri grubunu kontrol veri grubundan ayırma işleminde Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğu görülmektedir.

4.3.5 İnmemiş Testis Olmayan ve İnmemiş Testis Verilerini Sınıflandırma Modeli

4.3.5.1 Asıl Veriler

İnmemiş testis ve inmemiş testis olmayan patolojik verilerin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.28'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.29'de gösterildiği gibidir.

Tablo 4.28 İnmemiş Testis ve İnmemiş Testis Olmayan Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis Olmayan	**İnmemiş Testis**
Medium Gaussian SVM	**İnmemiş Testis Olmayan**	158	82
	İnmemiş Testis	32	107
Coarse Tree	**İnmemiş Testis Olmayan**	240	0
	İnmemiş Testis	139	0
Medium Tree	**İnmemiş Testis Olmayan**	136	105
	İnmemiş Testis	26	113

Tablo 4.29 İnmemiş Testis ve İnmemiş Testis Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Gaussian SVM	%69,9	%65,8	%77
Coarse Tree	%63,3	%100	%0
Medium Tree	%65,5	%56,4	%81,3

Tablo 4.28 ve Tablo 4.29'e bakıldığında; Medium Gaussian SVM modelinin tüm verileri birbirinden dengeli bir şekilde ayırabildiği, Coarse Tree modelinin ise inmemiş testisi diğer patolojik hastalardan ayırma oranı %0 iken diğer patolojik hastalıkları inmemiş testisten ayırma oranının %100 olduğu yani ayrımı yapamadığı ve verilerin hepsini diğer patolojik hastalık olarak sınıflandırdığı görülmektedir.

4.3.5.2 SMOTE Verileri

Sentetik olarak üretilen verilerden inmemiş testis ve inmemiş testis olmayan patolojik verilerin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.30'da; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.31'da gösterildiği gibidir.

Tablo 4.30 İnmemiş Testis ve İnmemiş Testis Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis Olmayan	**İnmemiş Testis**
Fine Gaussian SVM	**İnmemiş Testis Olmayan**	1228	19
	İnmemiş Testis	73	1176
Coarse Tree	**İnmemiş Testis Olmayan**	1032	215
	İnmemiş Testis	649	600
Medium Gaussian SVM	**İnmemiş Testis Olmayan**	907	340
	İnmemiş Testis	85	1164

Tablo 4.31 İnmemiş Testis ve İnmemiş Testis Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%96,3	%98,5	%94,2
Coarse Tree	%65,4	%82,8	%48
Medium Gaussian SVM	%83	%72,7	%93,2

Tablo 4.31'a bakıldığında, inmemiş testis veri grubunu inmemiş testis olmayan veri grubundan ayırma işleminde %96,3 oranla Fine Gaussian SVM modelinin diğer modellere göre daha iyi bir başarı yakaladığı, ancak Coarse Tree modelinin inmemiş testis olmayan veri grubunu inmemiş testis veri grubundan ayırmada ki başarısını inmemiş testis veri grubunu inmemiş testis olmayan veri grubundan ayırma işleminde yakalayamadığı görülmektedir.

4.3.6 İnguinal Herni Olmayan ve İnguinal Herni Verilerini Sınıflandırma Modeli

4.3.6.1 Asıl Veriler

İnguinal herni ve inguinal herni olmayan patolojik verilerin sınıflandırılmasından elde edilen karşılaştırma matrisi Tablo 4.32'de; karşılaştırma matrisinden elde

edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.33'de gösterildiği gibidir.

Tablo 4.32 İnguinal Herni ve İnguinal Herni Olmayan Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnguinal Herni Olmayan	**İnguinal Herni**
Medium Tree	**İnguinal Herni Olmayan**	201	11
	İnguinal Herni	73	94
Coarse Tree	**İnguinal Herni Olmayan**	210	2
	İnguinal Herni	89	78
Fine Tree	**İnguinal Herni Olmayan**	170	42
	İnguinal Herni	62	105

Tablo 4.33 İnguinal Herni ve İnguinal Herni Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Tree	%77,8	%94,8	%56,3
Coarse Tree	%76	%99,1	%46,7
Fine Tree	%72,6	%80,2	%62,9

Tablo 4.32 ve Tablo 4.33'ye bakıldığında; Fine Tree modelinin diğer patolojileri inguinal herniden ayırma başarısında diğer modellere göre düşük kaldığı ve Coarse Tree modelinin ise %99,1 oranla daha başarılı olduğu görülmektedir.

4.3.6.2 SMOTE Verileri

Sentetik olarak üretilen verilerden inguinal herni ve inguinal herni olmayan patolojik verilerin sınıflandırılması halinde halinde elde edilen karşılaştırma matrisi Tablo 4.34'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.35'de gösterildiği gibidir.

Tablo 4.34 İnguinal Herni ve İnguinal Herni Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnguinal Herni Olmayan	**İnguinal Herni**
Fine Gaussian SVM	**İnguinal Herni Olmayan**	1188	60
	İnguinal Herni	23	1226
Medium Gaussian SVM	**İnguinal Herni Olmayan**	1224	24
	İnguinal Herni	290	959
Cubic SVM	**İnguinal Herni Olmayan**	1180	68
	İnguinal Herni	163	1086

Tablo 4.35 İnguinal Herni ve İnguinal Herni Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%96,7	%95,2	%98,2
Medium Gaussian SVM	%87,4	%98,1	%76,8
Cubic SVM	%90,7	%94,6	%86,9

Tablo 4.35'e bakıldığında, inguinal herni veri grubunu inguinal herni olmayan veri grubundan ayırma işleminde Fine Gaussian SVM modelinin %98,2 oranla en başarılı model olduğu, inguinal herni olmayan veri grubunun inguinal herni veri grubundan ayırma işleminde ise Medium Gaussian SVM modelinin %98,1 oranla diğer modellere göre daha başarılı olduğu görülmektedir.

4.3.7 Hidrosel Olmayan ve Hidrosel Verilerini Sınıflandırma Modeli

4.3.7.1 Asıl Veriler

Hidrosel ve hidrosel olmayan patolojik verilerin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.36'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.37'da gösterildiği gibidir.

Tablo 4.36 Hidrosel ve Hidrosel Olmayan Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Hidrosel Olmayan	**Hidrosel**
Quadratic SVM	**Hidrosel Olmayan**	295	11
	Hidrosel	61	12
Lineer SVM	**Hidrosel Olmayan**	306	0
	Hidrosel	73	0
Cubic SVM	**Hidrosel Olmayan**	259	47
	Hidrosel	45	28

Tablo 4.37 Hidrosel ve Hidrosel Olmayan Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Quadratic SVM	%81	%96,4	%16,4
Lineer SVM	%80,7	%100	%0
Cubic SVM	%75,7	%84,6	%38,4

Tablo 4.36 ve Tablo 4.37'ya bakıldığında; Lineer SVM modelinin hidrosel rahatsızlığının hemogram verilerini diğer patalojik rahatsızlıkların hemogram verilerinden ayırmada başarısız kaldığı, Cubic SVM modelinin Lineer SVM modeline göre daha başarılı olduğu gözükse de, Quadratic SVM modelinin diğer patolojik rahatsızlıkların hemogram verilerini hidrosel rahatsızlığının hemogram verilerinden ayırma oranında ve tüm verilerde doğru tahmin etme oranında en başarılı olduğu görülmektedir.

4.3.7.2 SMOTE Verileri

Sentetik olarak üretilen verilerden hidrosel ve hidrosel olmayan patolojik verilerin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.38'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.39'de gösterildiği gibidir.

Tablo 4.38 Hidrosel ve Hidrosel Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		Hidrosel Olmayan	**Hidrosel**
Fine Gaussian SVM	**Hidrosel Olmayan**	1228	20
	Hidrosel	75	1174
Cubic SVM	**Hidrosel Olmayan**	1019	229
	Hidrosel	75	1174
Medium Gaussian SVM	**Hidrosel Olmayan**	912	336
	Hidrosel	113	1136

Tablo 4.39 Hidrosel ve Hidrosel Olmayan SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%96,2	%98,4	%94
Cubic SVM	%87,8	%81,7	%94
Medium Gaussian SVM	%82	%73,1	%91

Tablo 4.39'e bakıldığında, hem hidrosel veri grubunu hidrosel olmayan veri grubundan ayırma işleminde hem de hidrosel olmayan veri grubunu hidrosel veri grubundan ayırma işleminde en başarılı modelin Fine Gaussian SVM modeli olduğu görülmektedir.

4.3.8 İnmemiş Testis ve İnguinal Herni Verilerini Sınıflandırma Modeli

4.3.8.1 Asıl Veriler

İnmemiş testis ve inguinal herni verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.40'da; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.41'da gösterildiği gibidir.

Tablo 4.40 İnmemiş Testis ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis	**İnguinal Herni**
Medium Gaussian SVM	**İnmemiş Testis**	132	7
	İnguinal Herni	62	105
Coarse Tree	**İnmemiş Testis**	135	4
	İnguinal Herni	83	84
Fine Gaussian SVM	**İnmemiş Testis**	49	90
	İnguinal Herni	15	152

Tablo 4.41 İnmemiş Testis ve İnguinal Herni Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Gaussian SVM	%77,5	%95	%62,9
Coarse Tree	%71,6	%97,1	%50,3
Fine Gaussian SVM	%65,7	%35,3	%91

Tablo 4.41'a bakıldığında; inmemiş testis rahatsızlığının hemogram verilerini inguinal herni hemogram verilerinden %97,1 oranında ayırabilmesiyle Coarse Tree modelinin diğer modellere göre daha başarılı olduğu ve Fine Gaussian SVM modelinin ise bu ayrımı başarılı şekilde yapamadığı görülmektedir.

4.3.8.2 SMOTE Verileri

Sentetik olarak üretilen verilerden inmemiş testis ve inguinal herni verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.42'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.43'de gösterildiği gibidir.

Tablo 4.42 İnmemiş Testis ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis	**İnguinal Herni**
Fine Gausssian SVM	**İnmemiş Testis**	1204	45
	İnguinal Herni	13	1236
Coarse Gaussian SVM	**İnmemiş Testis**	1235	14
	İnguinal Herni	354	895
Cubic SVM	**İnmemiş Testis**	1205	44
	İnguinal Herni	105	1144

Tablo 4.43 İnmemiş Testis ve İnguinal Herni SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gausssian SVM	%97,7	%96,4	%99
Coarse Gaussian SVM	%85,3	%98,9	%71,7
Cubic SVM	%94	%96,5	%91,6

Tablo 4.43'ye bakıldığında, inmemiş testis veri grubunu inguinal herni veri grubundan ayırma işleminde %98,9 oranla en iyi başarıyı Coarse Gaussian SVM modelinin, inguinal herni veri grubunu inmemiş testis veri grubundan ayırma işleminde ise Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğu görülmektedir.

4.3.9 İnmemiş Testis ve Hidrosel Verilerini Sınıflandırma Modeli

4.3.9.1 Asıl Veriler

İnmemiş testis ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.44'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.45'de gösterildiği gibidir.

Tablo 4.44 İnmemiş Testis ve Hidrosel Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis	**Hidrosel**
Quadratic SVM	**İnmemiş Testis**	124	15
	Hidrosel	42	31
Fine Gaussian SVM	**İnmemiş Testis**	139	0
	Hidrosel	72	1
Medium Tree	**İnmemiş Testis**	92	47
	Hidrosel	24	49

Tablo 4.45 İnmemiş Testis ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Quadratic SVM	%73,1	%89,2	%42,5
Fine Gaussian SVM	%66	%100	%1,4
Medium Tree	%66,5	%66,2	%67,1

Tablo 4.45'e bakıldığında; Fine Gaussian SVM modelinin inmemiş testis rahatsızlığını hidrosel rahatsızlığından ayırma işleminde %100 başarı yakaladığı gözükse de, hidrosel rahatsızlığını inmemiş testis rahatsızlığından ayırma işleminde başarı yakalayamadığı ve ayrımı dengeli bir şekilde yapamadığı görülmektedir.

4.3.9.2 SMOTE Verileri

Sentetik olarak üretilen verilerden inmemiş testis ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.46'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.47'da gösterildiği gibidir.

Tablo 4.46 İnmemiş Testis ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnmemiş Testis	**Hidrosel**
Fine Gaussian SVM	**İnmemiş Testis**	1237	12
	Hidrosel	54	1195
Cubic SVM	**İnmemiş Testis**	1088	161
	Hidrosel	99	1150
Medium Gaussian SVM	**İnmemiş Testis**	1049	200
	Hidrosel	179	1070

Tablo 4.47 İnmemiş Testis ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%97,4	%99	%95,7
Cubic SVM	%89,6	%87,1	%92,1
Medium Gaussian SVM	%84,8	%84	%85,7

Tablo 4.47'ya bakıldığında, inmemiş testis veri grubu ile hidrosel veri grubunu birbirinden ayırma işleminde en başarılı modelin Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğu görülmektedir.

4.3.10 İnguinal Herni ve Hidrosel Verilerini Sınıflandırma Modeli

4.3.10.1 Asıl Veriler

İnguinal herni ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.48'de; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.49'de gösterildiği gibidir.

Tablo 4.48 İnguinal Herni ve Hidrosel Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnguinal Herni	**Hidrosel**
Medium Tree	**İnguinal Herni**	141	26
	Hidrosel	28	45
Fine Gaussian SVM	**İnguinal Herni**	166	1
	Hidrosel	69	4
Kernel Naive Bayes	**İnguinal Herni**	111	56
	Hidrosel	16	57

Tablo 4.49 İnguinal Herni ve Hidrosel Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Medium Tree	%77,5	%84,4	%61,6
Fine Gaussian SVM	%70,8	%99,4	%5,5
Kernel Naive Bayes	%70	%66,5	%78,1

Fine Gaussian SVM modelinin inguinal herni verilerini hidrosel verilerinden başarılı şekilde ayırabildiği ancak hidrosel verilerini inguinal herni verilerinden ayırmada başarılı olamadığı görülmektedir. Kernel Naive Bayes modelinin özgüllük oranında başarıyı yakaladığı, ancak doğruluk ve duyarlılık oranlarında Medium Tree modelinin daha yüksek sonuçlar elde ederek daha başarılı olduğu görülmektedir.

4.3.10.2 SMOTE Verileri

Sentetik olarak üretilen verilerden inguinal herni ve hidrosel verilerinin sınıflandırılması halinde elde edilen karşılaştırma matrisi Tablo 4.50'da; karşılaştırma matrisinden elde edilen doğruluk, duyarlılık ve özgüllük değerleri ise Tablo 4.51'de gösterildiği gibidir.

Tablo 4.50 İnguinal Herni ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisi

	Gerçek Sınıfı	**Tahmin Sınıfı**	
		İnguinal Herni	**Hidrosel**
Fine Gaussian SVM	**İnguinal Herni**	1242	7
	Hidrosel	58	1191
Cubic SVM	**İnguinal Herni**	1125	124
	Hidrosel	26	1223
Medium Gaussian SVM	**İnguinal Herni**	1028	221
	Hidrosel	18	1231

Tablo 4.51 İnguinal Herni ve Hidrosel SMOTE Veri Grubunun Karşılaştırma Matrisinden Elde Edilen Sonuçlar

	Doğruluk	**Duyarlılık**	**Özgüllük**
Fine Gaussian SVM	%97,4	%99,4	%95,4
Cubic SVM	%94	%90,1	%97,9
Medium Gaussian SVM	%90,4	%82,3	%98,6

Tablo 4.51'ye bakıldığında, inguinal herni veri grubunu hidrosel veri grubundan ayırma işleminde Fine Gaussian SVM modelinin %99,4 oranla en başarılı model olduğu, hidrosel veri grubunu inguinal herni veri grubundan ayırma işleminde ise Medium Gaussian SVM modelinin %98,6 oranla Fine Gaussian SVM modelinden daha başarılı olduğu görülmektedir.

5
SONUÇ VE ÖNERİLER

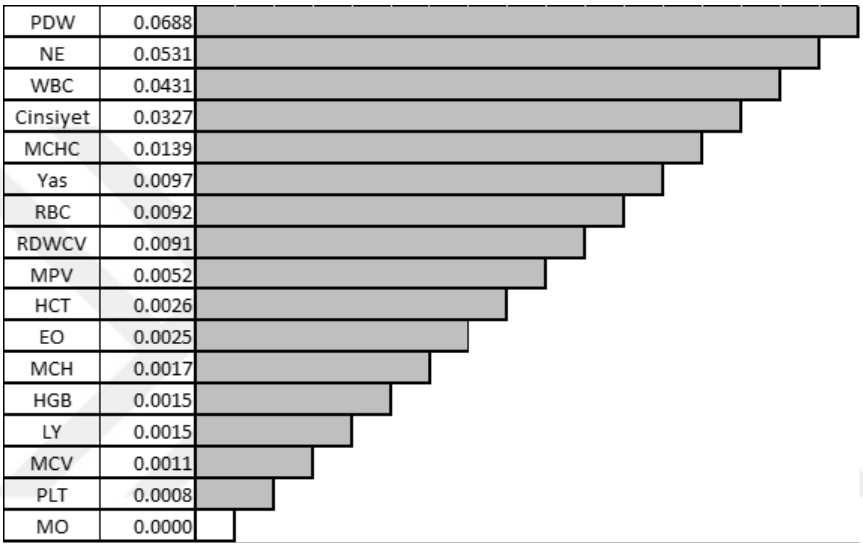
Bu çalışmada Tekirdağ Namık Kemal Üniversitesi Hastanesi ve Bahçelievler Devlet Hastanesine inguinal kanal bölgesine yönelik patolojiler ile başvuranlardan alınan hemogram, cinsiyet ve yaşlardan oluşan veri setinden yola çıkarak, başvuranların inguinal herni, inmemiş testis ve hidrosel rahatsızlıklarına sahip olup olmama durumları incelendi ve makine öğrenmesi algoritmalarıyla yapılan sınıflandırmalar ele alındı. Bu kapsamda, öncelikle 2014-2023 yılları arasındaki Tekirdağ Namık Kemal Üniversitesi Hastanesi Çocuk Cerrahisi ABD Kliniği verileri ve 2014-2016 yılları arasındaki Bahçelievler Devlet Hastanesi Çocuk Cerrahisi Kliniği verileri toplanmıştır. Toplanan veriler, veri düzenleme ve birleştirme gibi ön aşamalardan geçirilerek verilerin hemogram, yaş ve cinsiyet değerleri arasındaki ilişki istatiksel olarak analiz edildi.

Veri setinde inmemiş testis, inguinal herni ve hidrosel hastalıklarına sahip olmayan veri grupları **kontrol** olarak, inmemiş testis hastalığına sahip olanlar **inmemiş testis** olarak, inguinal herni hastalığına sahip olanlar **inguinal herni** olarak, hidrosel hastalığına sahip olanlar **hidrosel** olarak ve bu üç hastalığın verileri birleştirilerek inmemiş testis, inguinal herni ve hidrosel rahatsızlıklarının hepsi **hasta** olarak etiketlendi.

Hastaneye başvuran insanların hemogram verilerinden patolojik rahatsızlıklara sahip olup olmadığının tahmin edilebilmesi için veri seti; MatLab Sınıflandırma Öğreticisi uygulamasında bulunan Karar Ağaçları, Naive Bayes Sınıflandırmaları ve Destek Vektör Makinesi algoritmalarıyla eğitildi ve test edildi. Asıl verilerle yapılan sınıflandırma modellerinde, veri gruplarını birbirinden ayırma işlemlerinde veri sayısı fazla olan veri grubuna ait sınıflandırma başarısının yüksek çıktığı görülürken veri sayısı az olan veri grubuna ait sınıflandırma başarısının düşük çıktığı görüldü. Bunun nedeni veri setinin dengesiz dağılımından kaynaklı olabilir diye düşünülerek veri gruplarının veri sayıları, SMOTE algoritmasından yararlanılarak veri sayısı en fazla olan kontrol veri grubuyla eşitlendi.

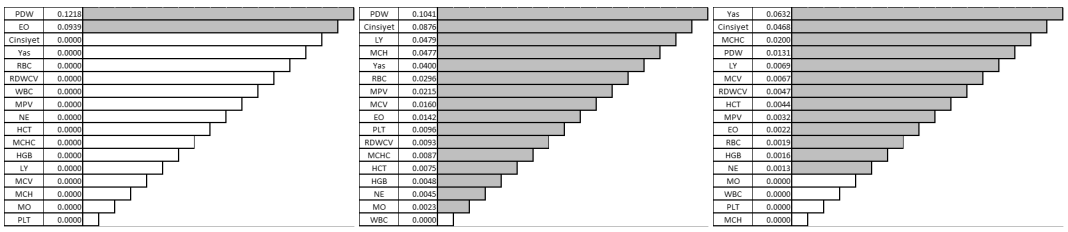
Sentetik olarak üretilen verilerden elde edilen yeni veri gruplarının sınıflandırılmalarında, Fine Gaussian SVM modelinin diğer modellere göre daha başarılı olduğu görülmektedir. Fine Gaussian SVM modelinden sonra ise Cubic SVM, Coarse Gaussian SVM, Medium Gaussian SVM modellerinin diğer modellere göre daha iyi başarı yakaladığı görülmektedir. Algoritmaların bu başarıları elde etmesini sağlayan özniteliklerin önem katsayıları şu şekilde elde edildi;

Kontrol ve hasta verilerinin sınıflandırılması halinde en çok etkili olan özniteliklerin sıralaması Şekil 5.1'de gösterildiği gibidir.



Şekil 5.1 Kontrol ve Hasta Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması

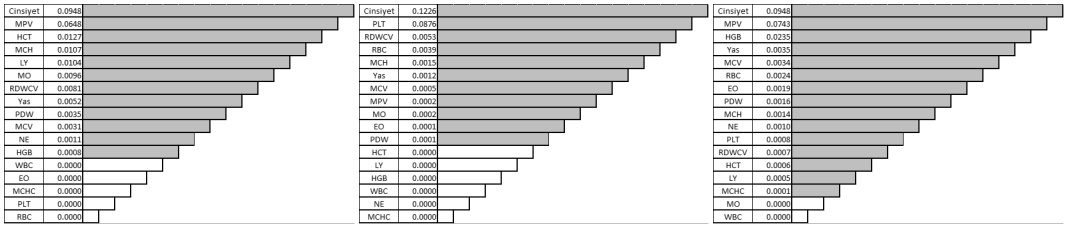
Kontrol ve sırasıyla inmemiş testis, inguinal herni, hidrosel verilerinin sınıflandırılması halinde en çok etkili olan özniteliklerin sıralaması Şekil 5.2'de gösterildiği gibidir.



Şekil 5.2 Sırasıyla Kontrol ve İnmemiş Testis, Kontrol ve İnguinal Herni, Kontrol ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması

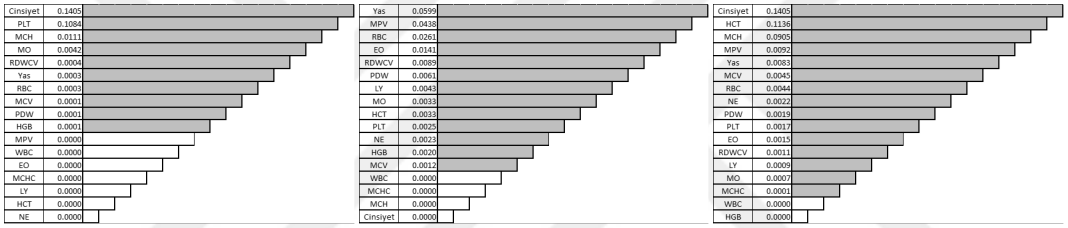
Sırasıyla; inmemiş testis ve inmemiş testis olmayan patolojik verilerinin, inguinal herni ve inguinal herni olmayan patolojik verilerinin, hidrosel ve hidrosel olmayan

patolojik verilerinin sınıflandırılması halinde en çok etkili olan özniteliklerin sıralaması Şekil 5.3'de gösterildiği gibidir.



Şekil 5.3 Sırasıyla İnmemiş Testis Olmayan ve İnmemiş Testis, İnguinal Herni Olmayan ve İnguinal Herni, Hidrosel Olmayan ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması

Hastalıkların her birine ait verilerin birbiriyle sınıflandırılması halinde en çok etkili olan özniteliklerin sıralaması sırasıyla Şekil 5.4'de gösterildiği gibidir.



Şekil 5.4 Sırasıyla İnmemiş Testis ve İnguinal Herni, İnmemiş Testis ve Hidrosel, İnguinal Herni ve Hidrosel Verilerinin Sınıflandırılmasında Öznitelik Katsayıları Sıralaması

Sonuç olarak, inguinal kanal patolojilerinden toplumda en sık karşılaşılan inmemiş testis, inguinal herni ve hidrosel olgularının hemogram verileri ile ayrılabileceği hipotezi doğrulanmış olmaktadır. Bu üç hastalık grubunun birbiri ile iç içe olması tanıyı zorlaştıran etkenlerden biri olduğu gibi çalışmamızda sonuçları etkileyen önemli bir parametredir. İnmemiş testis olgularına inguinal herni eşlik edebildiği gibi inguinal herni hidrosel ile birlikte de olabilmektedir. Hasta verileri geriye dönük olarak toplandığı için mevcut çalışmada bunun ayrımı yapılamamaktadır ve bu durumun algoritmaların başarısını düşüren önemli bir etken olduğu düşünülmektedir. Ayrıca, çalışmada kullanılan parametrelerden biri olan yaş parametresi yıl olarak hesaplanmıştır. Fakat hemogram verilerini oluşturan hücrelerin yaşam süreleri günler ile sınırlı olduğundan, yıl yerine gün kullanma durumunda daha optimize sonuçlar elde edileceği düşünülmektedir. Bu çalışmanın sonucu göstermektedir ki, yeterli miktarda temiz veri olması durumunda inmemiş testis, inguinal herni ve hidrosel ayrımı hemogram verileri ile yapılma potansiyeline sahiptir. Özellikle SMOTE ile artırılmış veri sayısı sonrası

örneklem boyutlarının eşitlenmesi ile önerilen algoritmaların başarılarında önemli ölçüde gelişme izlenmiştir. Mevcut klinik uygulamada bu üç hastalığın tanısı fiziki muayene ile yapılmaktadır. Fiziki muayene hemogram verilerine kıyasla non invaziv bir tanı yöntemidir. Ancak gelişen teknoloji ile birlikte tam kan sayımı ölçümlerinin yapılma tekniğinin gelişmesi ve değişmesi durumunda makine öğrenmesi algoritmaları kitlesel taramalar amacıyla kullanılma potansiyeline sahiptir. Ayrıca bu çalışmadan çıkarılacak bir diğer sonuç ise bu hastalıkların her birinin patofizyolojik olarak farklı olduğu, vücut içinde inflamatuar bir süreç başlattığı ve bireylerin farklı hastalıkları olmamasına rağmen kan parametrelerini değiştirdiğidir. Temel bilimler alanında yapılmakta olan çalışmalar için yol gösteren bir bulgu olduğu düşünülmektedir.

KAYNAKÇA

[1] Y. Yılmaz, İ. O. Özen, "Inmemiş testis kliniğinde güncel yaklaşımlar," *Sted*, vol. 13, no. 6, pp. 211–214, 2004.

[2] A. Tosun Alp, E. Yeniaras, M. Sahtiyancı, G. Daş, "Çocuklarda kasık fıtığının erken tanısı," *İstanbul Gelişim Üniversitesi Sağlık Bilimleri*, no. 7, pp. 697–701, 2019.

[3] S. Ulusoy Tangül, "Çocuk cerrahisine başvuran olgularda inguinal bölge patolojilerinin değerlendirilmesi: Tek hekimin deneyimleri ve yaşadığı zorluklar," vol. 10, no. 1, pp. 83–87, 2020.

[4] L. Cao, "Data science: A comprehensive overview," *ACM Computing Surveys (CSUR)*, vol. 50, no. 3, pp. 1–42, 2017.

[5] J. Biamonte, P. Wittek, N. Pancotti, et al., "Cybersecurity data science: An overview from machine learning perspective," *Nature*, vol. 549, pp. 195–202, 2017.

[6] I. H. Sarker, A. Kayes, S. Badsha, H. Alqahtani, P. Watters, A. Ng, "Cybersecurity data science: An overview from machine learning perspective," *Journal of Big data*, vol. 7, pp. 1–29, 2020.

[7] S. Gesoğlu, *Yeni doganların yogun bakım ihtiyacinin makine ögrenmesi teknikleri ile dogum oncesinde tahmin edilmesi*, 2022.

[8] A. Oğuzlar, "Veri ön işleme," *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, no. 21, 2003.

[9] N. Ankara, H. Şahintürk, "Dengesiz kredi skorlama veri setlerinde kolektif öğrenme algoritmalarının performans değerlendirmesi," *PressAcademia Procedia*, vol. 9, no. 1, pp. 180–185, 2019.

[10] M. Yavaş, A. Güran, M. Uysal, "Covid-19 veri kümesinin smote tabanlı örnekleme yöntemi uygulanarak sınıflandırılması," *Avrupa Bilim ve Teknoloji Dergisi*, pp. 258–264, 2020.

[11] S. Ray, "A quick review of machine learning algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 35–39. doi: 10.1109/COMITCon.2019.8862451.

[12] T. Kavzoğlu, İ. Çölkesen, "Karar ağaçları ile uydu görüntülerinin sınıflandırılması," *Harita Teknolojileri Elektronik Dergisi*, vol. 2, pp. 36–45, 2010.

[13] A. S. Albayrak, Ş. K. Yılmaz, "Veri mdenciliği: Karar ağacı algoritmaları ve imkb verileri üzerine bir uygulama," *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, vol. 14, pp. 31–52, 2009.

[14] *Karar ağaçları (decision tree)*. [Online]. Available: https://yavuz.github.io/karar-agaclari/ (visited on 01/20/2024).

[15] R. Solmaz, M. Günay, A. Alkan, "Fonksiyonel tiroit hastalığı tanısında naive bayes sınıflandırıcının kullanılması," *Akademik Bilişim Konferansı, Mersin*, pp. 891–897, 2014.

[16] S. Ayhan, Ş. Erdoğmuş, "Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi," *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, vol. 15, no. 1, pp. 175–201, 2014.

[17] T. Kavzaoğlu, İ. Çölkesen, "Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi," *Harita Dergisi*, vol. 76, pp. 73–82, 2010.

[18] H. Emir, M. Eliçevik, *İnguinal Bölge Patolojileri*.

[19] M. Rowe, "The newborn as a surgical patient," *Pediatric Surgery. 5th ed. Philadelphia: Mosby year book Inc*, pp. 43–55, 1998.

[20] A. Mesci, *Linchtenstein herni onarımının testiküler kan akımı üzerinde olan etkisi*, 2005.

[21] G. Coşkun, "Testis inişinin moleküler mekanizması," *Arşiv Kaynak Tarama Dergisi*, vol. 30, no. 1, pp. 22–30, 2021. doi: 10.17827/aktd.699932.

[22] Y. K. Haspolat, F. Aktar, S. Ege, S. Bilici, "Cocuk ve ergenlerde gonad hastalıkları," *Orient*, 2019.

[23] A. Girdhar, H. Kapur, V. Chahar, "Classification of white blood cell using convolution neural network," *Biomedical Signal Processing and Control*, 2022.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. A.Yılmaz, S.Selim, E.Aydın, "Başlıca inguinal kanal patolojilerinin makine öğrenmesi algoritmaları ile tahmin edilmesi", *7.Uluslararası Araştırmacılar, İstatistikçiler ve Genç İstatistikçiler Kongresi*, Kasım 2023, İstanbul TÜRKİYE