

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HİPOSPADİAS OLGULARINDA İNFLAMATUAR SÜRECİN VERİ
MADENCİLİĞİ YÖNTEMLERİ İLE TESPİTİ

Hülya ACAR

YÜKSEK LİSANS TEZİ

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Danışman

Doç. Dr. Birol ASLANYÜREK

Eş Danışman

Doç. Dr. Emrah AYDIN

Eylül, 2022

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HİPOSPADİAS OLGULARINDA İNFLAMATUAR SÜRECİN VERİ
MADENCİLİĞİ YÖNTEMLERİ İLE TESPİTİ

Hülya ACAR tarafından hazırlanan tez çalışması 12.09.2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Mühendisliği Anabilim Dalı, Matematik Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Birol ASLANYÜREK
Yıldız Teknik Üniversitesi
Danışman

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

Jüri Üyeleri

Doç. Dr. Birol ASLANYÜREK, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Hülya ŞAHİNTÜRK, Üye
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Oğuzhan KIVRAK, Üye
Bandırma Onyedli Eylül Üniversitesi

Danışmanım Doç. Dr. Birol ASLANYÜREK ve eş danışmanım Doç. Dr. Emrah AYDIN sorumluluğunda tarafımca hazırlanan Hipospadias Olgularında İnflamatuvar Sürecin Veri Madenciliği Yöntemleri ile Tespiti 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.

Hülya ACAR

İmza



Aileme

TEŞEKKÜR

Bu çalışmanın yürütülmesi sırasında desteğini esirgemeyen, beni yüreklendiren ve çalışma sürecinde sabırla her zaman yanımda olan danışmanım Doç. Dr. Birol ASLANYÜREK'e ve eş danışmanım Doç. Dr. Emrah AYDIN'a, lisans ve yüksek lisans eğitimim boyunca beni destekleyen aileme teşekkürlerimi sunarım.

Hülya ACAR



İÇİNDEKİLER

KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ	ix
ÖZET	x
ABSTRACT	xii
1 GİRİŞ	1
1.1 Literatür Özeti.....	1
1.2 Tezin Amacı	4
1.3 Orjinal Katkı.....	5
2 HİPOSPADİAS	6
2.1 Etiyoloji.....	6
2.2 Genetik Faktörler	6
2.3 Hormonal ve Çevresel Etkiler	7
2.4 Sendromlar ve Birlikte Gözlenen Anomaliler.....	7
2.5 Tedavi.....	7
3 VERİ MADENCİLİĞİ	9
3.1 Veri Madenciliği Tarihi.....	10
3.2 Veri Madenciliği Süreci	10
3.3 Veri Madenciliği Uygulama Alanları.....	11
3.4 Veri Madenciliği Metodları.....	13
3.4.1 Sınıflandırma ve Regresyon.....	13
3.4.2 Destek Vektör Makinesi.....	16
3.4.3 Kümeleme	16
3.4.4 Birliktelik Kuralları.....	18
4 HİPOSPADİAS OLGULARINDA İNFLAMATUAR SÜRECİN TESPİTİ	19
4.1 İşi Anlama	20
4.1.1 İşin Hedeflerinin Belirlenmesi	20
4.1.2 Durum Değerlendirmesi	20
4.1.3 Veri Madenciliğinin Amacının Belirlenmesi	20
4.1.4 Proje Planı Üretme	21
4.2 Veriyi Anlama	21

4.2.1 İlk Verilerin Toplanması	21
4.3 Veri Ön İşleme	22
4.3.1 Veri Seçme	22
4.3.2 Veri Temizleme	24
4.4 Modelleme	27
4.4.1 Özniteliklerin İstatistiksel Analizi	27
4.4.2 Modelin İnşası	29
4.4.3 Modellerden Çıkan Sonuçların Değerlendirilmesi.....	30
5 SONUÇ VE ÖNERİLER	37
KAYNAKÇA	39
A ÖZNİTELİKLERİN KUTU DİYAGRAMLARI	44
TEZDEN ÜRETİLMİŞ YAYINLAR	54

AUC	Area Under Curve
CRIPS-DM	Cross Industry Standard Process for Data Mining
ICD	International Classification Diseases and Related Health
KNN	K-Nearest Neighbors
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine



ŞEKİL LİSTESİ

Şekil 2.1 Hipospadias cerrahi operasyonu.....	8
Şekil 3.1 Veri madenciliği tarihi	10
Şekil 3.2 Veri tabanında bilgi keşif süreci	11
Şekil 3.3 Veri madenciliğinin sağlıkta kullanım alanları.....	13
Şekil 3.4 Karar ağacı yapısı.....	15
Şekil 3.5 Yapay sinir ağları	16
Şekil 3.6 Kümeleme örneği.....	17
Şekil 4.1 CRIPS-DM metodolojisi.....	19
Şekil 4.2 Hastane veri tabanından alınmış anamnez dosya görüntüsü.....	21
Şekil 4.3 Hastane veri tabanından alınmış laboratuvar dosya görüntüsü.....	22
Şekil 4.4 Düzenlenmiş anamnez dosya görüntüsü	23
Şekil 4.5 Düzenlenmiş laboratuvar dosya görüntüsü	23
Şekil 4.6 Aykırı değerlerin tespiti.....	27
Şekil 4.7 Performanslarına göre ROC eğrileri.....	34
Şekil 4.8 Medium Gaussian SVM Yöntemi için ROC eğrisi ve AUC değeri.....	35
Şekil 4.9 Optimizable Tree Yöntemi için ROC eğrisi ve AUC değeri.....	35
Şekil 4.10 Linear Discriminant Yöntemi için ROC eğrisi ve AUC değeri.....	35

TABLO LİSTESİ

Tablo 1.1 Çeşitli ülkelerde belirlenmiş hipospadias oranları.....	2
Tablo 1.2 Türkiye'nin çeşitli bölgelerinde belirlenmiş hipospadias oranları.....	3
Tablo 4.1 ICD kodları ve sağlık problemleri.....	25
Tablo 4.2 İstatistiksel analiz test sonuçları.....	28
Tablo 4.3 Maliyet matrisi.....	30
Tablo 4.4 Model başarımlar oranları.....	30
Tablo 4.5 Medium Gaussian Yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri.....	31
Tablo 4.6 Optimizable Tree Yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri.....	32
Tablo 4.7 Coarse Tree Yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri.....	32
Tablo 4.8 Sınıflandırma modellerinde başarımlar kriterleri.....	33

Hipospadias Olgularında İnflamatuvar Sürecin Veri Madenciliği Yöntemleri ile Tespiti

Hülya ACAR

Matematik Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Birol ASLANYÜREK

Eş-Danışman: Doç. Dr. Emrah AYDIN

Hipospadias çocuklarda görülen en yaygın genital bölge anomalilerinden biridir. Bu hastalık üretral plakada inflamasyona yol açmaktadır. Hemogramda bulunan parametreler inflamatuvar süreçler hakkında bilgi verebilmektedir. Ayrıca hemogram testi uygulanabilirliği ve maliyetinin düşük olması sebebiyle yaygın bir test yöntemidir. Literatüre baktığımızda, her geçen yıl bilgisayarların verileri depolama kapasiteleri artmaktadır. Bu durum çeşitli alanlarda devasa bilgi yığınlarının oluşmasına yol açmaktadır. Bu bağlamda önde gelen alanlardan birisi de tıp sektörüdür. Günümüzde hastane veri tabanları keşfedilmeyi bekleyen birbiriyle ilişkili çok miktarda veri barındırmaktadır. Bu tezde hipospadias olgularında veri madenciliği yöntemleriyle inflamatuvar süreç tespit edilmeye çalışıldı. Bunun için hastaneye sünnet operasyonu talebiyle gelmiş çocuklardan alınan veriler ile hipospadias tanısı konulmuş olguların verileri iki ayrı grupta incelenmiştir. İnflamatuvar süreci ortaya koyabilmek için öznitelik olarak hemogram test sonuçları kullanılmıştır. Öncelikle her bir hemogram parametre değerinin Kolmorogov Smirnov testi aracılığıyla normal dağılıma uyup uymadığı

tespit edilmiştir. Daha sonra hipospadiaslı olgularla sünnet talebiyle gelmiş çocuklar arasında Mann Whitney U testi kullanılarak anlamlı bir fark olup olmadığı incelenmiştir. İstatistiksel analiz sonrası veri madenciliği yöntemleriyle hipospadias olgularında inflamasyonu ayırt ettirecek belirteçler olup olmadığına dair sonuçlar hemogram değerleri üzerinden elde edilmiştir. Bunun için Matlab programının ‘Classification Learner’ (2022b) adlı aracında tanımlanmış denetimli makine öğrenmesi yöntemleri uygulanarak çeşitli modeller eğitilmiş ve test veri seti kullanılarak doğruluk oranları hesaplanmıştır. Modellerden çıkan sonuçlar karşılaştırıldığında, Medium Gaussian, Optimizable Tree ve Coarse Tree yöntemlerinin en yüksek doğruluk değerleri ile tahminleme yaptığı görülmüştür.

Anahtar Kelimeler: *Veri madenciliği, hipospadias, sınıflandırma, inflamasyon.*

Detection of Inflammatory Process in Hypospadias Cases with Data Mining Methods

Hülya ACAR

Department of Mathematical Engineering

Master of Science Thesis

Supervisor: Assoc. Prof. Dr. Birol ASLANYÜREK

Co-Supervisor : Assoc. Prof. Dr. Emrah AYDIN

Hypospadias is one of the most common genital anomalies in children. This disease causes inflammation in the urethral plate. The parameters in the hemogram can provide information about inflammatory processes. However, hemogram test is a common test method due to its feasibility and low cost. When we look at the literature, the data storage capacity of computers is increasing every year. This situation leads to the formation of huge stacks of information in various fields. In this context, one of the leading fields is the medical sector. Today, hospital databases contain a large amount of interrelated data waiting to be discovered. In this thesis, it was tried to detect the inflammatory process in hypospadias cases with data mining methods. For this, the data obtained from the children who came to the hospital with the request of circumcision operation and the data of the cases diagnosed with hypospadias were analyzed in two separate groups. To reveal inflammatory process, hemogram test results were used as features. Firstly, the value of each hemogram parameter was determined by using the Kolmogorov Smirnov test to fit the normal distribution. Then, the Mann Whitney U test was used to examine whether there was a significant difference between the hypospadias cases and the children who came with the request for

circumcision. After statistical analysis, data mining methods were used to determine whether there are markers to distinguish inflammation in hypospadias cases, based on hemogram values. For this, various models were trained by applying supervised machine learning methods defined in the 'Classification Learner' (2022b) tool of the Matlab program, and accuracy rates were calculated using the test data set. When the results from the models were compared, it was seen that the Medium Gaussian, Optimizable Tree and Coarse Tree methods made predictions with the highest accuracy values.

Keywords: *Data mining, hypospadias, classification, inflammation.*



1.1 Literatür Özeti

Hipospadias çocuklarda en fazla ortaya çıkan ürogenital anomalilerden biridir (Keays ve Dave, 2017). İnmemiş testisten sonra erkeklerde ikinci en sık görülen doğumsal anomalidir (Van Der Horst, 2017). Hipospadias idrar deliğinin penis ucunda olmamasına eşlik edebilen penis gövdesinin öne doğru eğik olması (kordi) ve sünnet derisinin bütünlüğünün tam olmaması durumlarının bütünü olarak tanımlanır (Anand ve Lotfollahzadeh, 2022). Hipospadias yaklaşık olarak 300 doğumda 1 görülmektedir (Baskin ve Ebberts, 2006). Dünya üzerinde hipospadias oranlarını tespit edebilmek üzere farklı ülkelerde çeşitli çalışmalar yapılmıştır.

2009 yılında yayınlanan Danimarka'ya ait ulusal doğum kayıtlarının bulunduğu veri tabanından alınan çalışmaya göre 1977-2005 yıllarında doğan 921.745 erkek doğumunda hipospadias oranı %0,24 ile %0,52 arasında olduğu görülmüştür (Lund vd., 2009). Porter ve arkadaşlarının yayınladığı bir başka çalışmaya göre ise, Amerika Birleşik Devletleri'nin Vaşington Eyaleti'nde 12.930 erkek çocuk arasında hipospadiaslı erkek çocukların oranı %0,35 ile %0,56 arasında çıkmıştır (Porter vd., 2005). Yine ABD'nin Kaliforniya Eyaleti'nde 1984-1994 yılları arasında yapılan çalışmada bu oran %0,33 olarak bulunmuştur (Carmicheal vd., 2003).İngiltere'de 1993-2000 yılları arasında 165.000 erkek çocuk arasında yapılan araştırmada ise bu oranın %0,31 olduğu görülmüştür (Abdullah vd., 2007).1988-2013 yılları arasında Kanada'da yapılan bir diğer çalışmada ise 132.122 erkek çocuk arasında hipospadias patolojisine ait çocukların oranı %0,78 olarak belirlenmiştir (Lane vd., 2017), (Tablo 1.1).

Türkiye'de hipospadias görülme sıklığının ise çeşitli çalışmalarda %0,39 ile %0,83 arası değişen oranlarda tespit edildiği görülmüştür (Kayıkçı vd., 2005). Literatüre bakıldığında, hipospadias oranının tespitine yönelik çalışmaların büyük çoğunluğunun farklı il ve ilçe merkezlerindeki ilköğretim okullarında bulunan erkek öğrencilerden hareketle tespit edildiği görülmüştür.

Tablo 1.1 Çeşitli ülkelerde belirlenmiş hipospadias oranları

Çalışma	Çalışma Aralığı	Popülasyon	Bölge	Veri Tabanı	Hipospadias Oranı
Lund vd.	1977-2005	921.745	Danimarka	Ulusal Doğum Kaydı	%0,24- %0,52
Porter vd.	1987-2002	12.930	ABD	Vaşington Doğum Kayıt Belgesi	%0,35- %0,56
Carmicheal vd.	1984-1997	5.838	ABD	Kaliforniya Doğum Kayıt Belgesi	%0,33
Abdullah vd.	1993-2000	165.000	İngiltere	Kuzey Doğu Toplum Sağlığı Merkezi	%0,31
Lane vd.	1988-2013	132.122	Kanada	İl Yenidoğan Veri Tabanı	%0,78

Aralık 2009 tarihinde Cizre İlçe Merkezi'nde eğitim gören 7-14 yaş grubu toplam 500 erkek öğrenci arasında yapılan araştırmada hipospadiaslı çocukların oranı %0,6 olarak tespit edilmiştir (Koç vd.,2011). 1999 yılının Mayıs ve Haziran aylarını kapsayan bir diğer çalışmada ise Diyarbakır İl Merkezi'ndeki 65 ilköğretim okulundaki 10.350 erkek öğrenci arasından rastgele örneklem yöntemi ile seçilen 2.000 erkek öğrencinin %0,45'inde hipospadias gözlemlenmiştir. Hipospadias ve kordi hastalığının birlikte görülme oranı ise %0,15 olarak bulunmuştur (Akay vd.,2002). Kayıkçı ve arkadaşlarının yaptığı çalışma ise 2002-2003 Eğitim Öğretim Yılı'nda Düzce İl Merkezi'nde bulunan 25 adet ilköğretim okulunu kapsamaktadır. 13.420 erkek öğrenci arasından rastgele seçilen 1.534 erkek öğrencinin %0,39 oranında hipospadiaslı olduğu görülmüştür (Kayıkçı vd.,2005).2009 yılında Hatay'da yapılan araştırma ise, 6 ilköğretim okulundan 981 öğrenci arasında yapılmış, hipospadias oranı %0,4 olarak bulunmuştur (Danacı vd., 2010). Muğla İl Merkezi'nde Nisan 2011 tarihli çalışmada ise birinci ve ikinci sınıf öğrencisi olan toplam 285 erkek çocuk araştırmaya dahil edilmiş ve hipospadias oranı %1,1 olarak bulunmuştur (Deliktaş vd., 2014). 2009 yılının Mart ve Haziran ayları arasında yapılan araştırma ise Samsun İl Merkezi'ndeki 3 ilköğretim okulundan 6-11 yaşları arasında 977 erkek öğrenciyi kapsamaktadır. Hipospadias oranı ise %0,4 olarak belirlenmiştir (Varol vd., 2011). İlköğretim okullarında yapılan bu çalışmaların yanı sıra Mart 2011-Ocak 2012 tarihleri arasında Siirt Devlet

Hastanesi Üroloji Kliniği'ne sünnet talebiyle gelen 2 ay-12 yaş arası 260 erkek çocuk arasında hipospadias oranı %2,3 bulunmuştur (Yıldırım vd., 2011). Üzerinde çalışma yapılan örneklemin Türkiye'nin genel dağılımını temsil etmesi bakımından göze çarpan araştırma ise İzmir'de askerliğini yapmakta olan 2.061 genç erkeği kapsayan çalışmadır. %0,9 oranı ile 18 bireyin hipospadiaslı olduğu tespit edilmiştir. Bu hipospadiaslı olgulardan mevcut durumunun farkında olanların sayısının 8 kişi olduğu, %53,6'sının ise patolojilerinin farkında olmadığı görülmüştür. Hipospadiaslı 8 kişiden opere olan hasta sayısının 6, ortalama operasyon yaşlarının 12,6 olduğu bulunmuştur (Söylemez vd., 2011), (Tablo 1.2).

Tablo 1.2 Türkiye'nin çeşitli bölgelerinde belirlenmiş hipospadias oranları

Çalışma	Çalışma Aralığı	Bölge	Yaş	Örneklem	Hipospadias Sayısı	Yüzdesel Oran
Koç vd.	Aralık 2009	Cizre	7-14	500	3	%0,6
Akay vd.	Mayıs, Haziran 1999	Diyarbakır	7-8	2000	9	%0,45
Kayıkçı vd.	2002-2003 Öğretim Yılı	Düzce	7-15	1534	6	%0,39
Danacı vd.	Mayıs, Haziran 2009	Hatay	7-11	981	4	%0,4
Deliktaş vd.	Nisan 2011	Muğla	7-9	285	3	%1,1
Varol vd.	Mart-Haziran 2009	Samsun	6-11	977	4	%0,4
Yıldırım vd.	Mart-Ocak 2012	Siirt	0-12	260	6	%2,3
Söylemez vd.	Mart 2009	İzmir	19-22	2061	18	%0,9

Hipospadias ile alakalı yayınlanmış verilere dayanan çalışmaların büyük bir kısmı (Ortqvist vd., 2019) cerrahi yöntemler ve sonrasında ortaya çıkan işlevsel, morfolojik ve cinsel fonksiyonlar üzerindeki etkisine odaklanmıştır (Elmoghazy vd., 2020). İnflamasyon, patojenler ve hasarlı hücreler gibi uyaranlara karşı vücut dokularının vermiş olduğu kan damarlarını, bağışıklık hücrelerini ve moleküler araçları kapsayan koruyucu bir tepkidir. İnflamatuar sürecin işlevi zarar görmüş dokuları temizlemek ve doku onarımını başlatmaktır (Eming vd., 2007).

İnflamatuar süreç, kan hücrelerinin nitelik ve nicelik açısından incelenmesi olan başta hemogram olmak üzere birçok farklı test aracılığıyla izlenebilmektedir (Sikka ve Wang, 2008). Klinik pratikte, trombosit/lenfosit oranı (PLR), ortalama trombosit hacmi (MPV), nötrofil sayısının lenfosit sayısına oranı (PLR) gibi

hemogram parametreleri, çeşitli hastalıklarda inflamasyon belirteçleri şeklinde kullanılabileceği belirlenmiştir (Atar ve Aşkın, 2017). Buna ek olarak, erken dönemde yükselen lökosit sayısı (WBC), nötrofil (NE), lenfosit (LY) ve eritrosit dağılım genişliği (RDW) gibi hemogram parametreleri de inflamasyonun biyobelirteçleri arasındadır (Kaya vd., 2021).

Ekmark ve arkadaşlarının hipospadias anomalisini hücre bazda ele aldıkları çalışmada, hipospadias patolojisinin tekrarlayan özellikteki inflamatuvar hücreleri içerdiği ortaya konulmuştur (Ekmark vd., 2020). 2009 yılında yapılan bir çalışmada hipospadias ve çevresel ilişkiyi ortaya koymak için kayıtlı verilere dayalı bir kontrol çalışması yapılmıştır. Çalışmada çok değişkenli istatistiksel yöntemler ve veri madenciliği yöntemleri kullanılarak kayıtlı veriler analiz edilmiştir. 1.500'den fazla vakaya dayanan çalışmada tarımsal kirliliğin olduğu bölgelerde hipospadias oranının fazla olduğu ortaya konulmuştur (Amar vd., 2018).

2020 yılında yapılan başka bir çalışmada ise makine öğrenmesi ile görüntü işleme yöntemleri kullanılarak hipospadias tespiti ve sınıflandırılması amaçlanmıştır. Bunun için öncelikle hipospadias görüntü veri tabanından alınan 627 adet örnek ile eğitim yapılmıştır. Model, eğitiminde kullanılmamış olan 29 adet görüntü üzerinde test edildiğinde doğruluk oranı %60 olarak bulunmuştur. Daha sonraki eğitimde ise görüntü sayısı artırılarak 1.169 adet örnek kullanılmış ve 29 adet görüntü üzerinden test edildiğinde doğruluk oranı %90'a çıkmıştır (Fernandez vd., 2020).

1.2 Tezin Amacı

Bu tezin amacı hemogram test sonuçlarını kullanarak veri madenciliği yöntemleri aracılığıyla hipospadias olgusuna sahip hastalar ile sünnet talebiyle hastaneye başvurmuş hipospadiaslı olmayan grup arasındaki inflamatuvar farkı ortaya koymaktır. Bunun için 23.02.2021 tarihli 2021.58.02.21 protokol numaralı etik kurul izni ile Tekirdağ Namık Kemal Üniversitesi Çocuk Cerrahisi Anabilim Dalı'na ait veri tabanından alınan veriler kullanılmıştır. Hipospadias olgusuna sahip olan hasta verileri ile sünnet talebine sahip sağlıklı çocukların verilerine, iki sınıf olarak makine öğrenmesi işlemleri uygulanmıştır. Makine öğrenmesi modellerinden

ıkan sonular bařarı kriterlerine gre kıyaslanmıř, en iyi sonu veren algoritmalar tespit edilmeye alıřılmıřtır.

1.3 Orijinal Katkı

Literatr zetinde belirtildiėi zere hipospadiasa ait alıřmaların byk kısmında cerrahi yntemlere ve sonrasında ortaya ıkan komplikasyonlara odaklanılmıřtır. Veri madenciliėiyle alakalı alıřmaların ise ok fazla olmadıėı gzlemlenmiřtir. Yapılan literatr alıřması sonucu inflamatuvar sre ile hipospadias arasında veri madenciliėinin kullanıldıėı alıřmaya rastlanmamıřtır. Bu tez, hipospadias olgularında veri madenciliėi yntemleriyle inflamatuvar srecin ortaya konulmasına bir rnek teřkil etmektedir.



Hipospadias, peniste idrar deliğinin (meatus) normalde olması gereken yerde olmaması durumudur. Aynı zamanda bu olgularda penisin ön kısmında sünnet derisinin sıklıkla eksik olduğu da gözlemlenir. Vakaların büyük kısmında idrar deliği penisin uç kısmına yakın bölgededir. Sayıca az ve ağır vakalarda ise bu deliğin penisin orta ya da daha arka kısımlarında yer aldığı görülür. Meatusun penisin uç kısmına yakın bölgede olmasına balanik hipospadias, orta kısımlarda yer almasına midpenil hipospadias ve daha geri olan yumurtalara yakın kısımda bulunması durumuna ise penoskrotal hipospadias denir.

2.1 Etiyoloji

Hipospadiasın nedeni tam olarak tespit edilemese de gebelik esnasında penisin gelişimi için gerekli hormonların yeterli miktarda salgılanamaması, genetik yatkınlık, çevresel faktörler gibi çeşitli sebeplerden kaynaklandığına dair hipotezler bulunmaktadır. Dolayısıyla hipospadiasın ortaya çıkmasının multifaktöriyel olduğu söylenebilir (Baskin, 2000). Ailede hipospadias görülmesi, tüp bebek tedavisi ile oluşan gebeliklerde annenin diyabet hastalığının bulunması, düşük ağırlıktaki doğum, annenin 40 yaş üstü veya 24 yaşından önce hamile olması, erken doğum gibi durumlar hipospadias riskini artırmaktadır (Tekgül vd., 2016).

2.2 Genetik Faktörler

Olguların %7'si ailesel kümelenmeden etkilenmektedir. Ailesel faktörler midpenil hipospadias ve bulanık hipospadiasta daha fazla görülmektedir. Hipospadiasa sahip bir çocuğun erkek kardeşinde görülme yüzdesi %9 ile %17, tahmini olarak kalıtım oranı ise %55 ile %77 arasındadır (Van Der Zanden vd., 2012).

2.3 Hormonsal ve Çevresel Etkiler

Androjenler; embriyonik gelişim safhasında üreme sisteminin gelişmesini, metabolizmanın cinsiyete göre farklılaşmasını düzenleyen hormonlardır (Kimura vd., 1993). Androjenin genital gelişim safhasında kritik öneme sahip olduğu bilinmektedir. Androjen sentezindeki aksaklıklar ya da reseptör düzeyindeki problemler hipospadiası sebep olabilmektedir (Thankamony, 2016). Antiandrojenil etkinlik gösteren maddelerin ya da sentetik östrojenlerin hipospadiası sebebiyet verebildiği bazı araştırmalarla desteklenmektedir. Endüstriyel ürünlerin daha çok tüketildiği bölgelerde hipospadias oranının yüksek çıkması, antiandrojen ya da östrojenik etkiye sahip maddelere maruz kalan gebe hayvanların yavrularında hipospadias oranının yüksek çıkması çevresel faktörlere örnek olarak gösterilebilir (Tekgül vd., 2016).

2.4 Sendromlar ve Birlikte Gözlemlenen Anomaliler

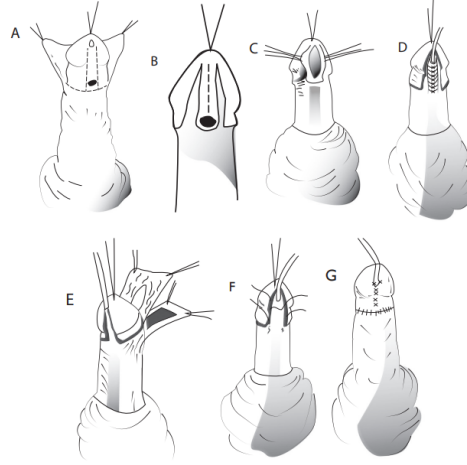
Hipospadias çeşitli sendromlarla birlikte gözlemlenebilmektedir. Bunlardan yaygın olanlar, WAGR sendromu, Opitz-6 sendromu ve Wolf-Hirschhorn sendromu olarak söylenebilir. Bu olgulara sahip hastaların %8 ile %10 arasında inmemiş testis, (Proksimal hipospadiaslı vakalarda bu oran %32'lere ulaşmaktadır) %9 ile %15 oranında inguinal herni gözlemlenmektedir (Tekgül vd., 2016).

2.5 Tedavi

Hipospadias tedavisinde amaç hem fonksiyonel işlevlerin normalleştirilmesi hem de kozmetik yönden normalleştirilmenin sağlanması olarak iki temel başlıkta toplanabilir. Hipospadias tedavisinin gerekliliği; oturarak idrar yapma zorunluluğu, cinsel birleşme esnasında eğriliğin oluşabilmesi sebebiyle problemler yaşama, meatusun normal yerinde olmaması sebebiyle spermin yumurtayı dölleyememesi, sperm birikiminde zorluk yaşanması ve sahip olunan genital görünümünden rahatsızlık olarak sıralanabilir (Rizedmiller, 2001).

Hipospadiasın tedavisi cerrahi yöntem ile yapılmaktadır. Düşük orandaki olgularda cerrahi operasyon öncesi hormon tedavisi yapılabilir. Temel yöntem idrar yolunda yeni kanal oluşturularak meatusun penisin uç kısmına getirilmesidir. Bu cerrahi yöntem meatusun bulunduğu yere ve penisin eğrilik

durumuna göre bir ya da birkaç seans olarak yapılabilir. Şekil 2.1 hipospadiaslı olguda cerrahi operasyonun adımlarını göstermektedir. Psikolojik sonuçlar, anestezi riskleri ve olgunun yaşına bağlı olarak doku boyutları göz önüne alındığında uygun operasyon zamanı 6 ile 18 ay arasında önerilmektedir (Manzoni, 2004).



Şekil 2.1 Hipospadias cerrahi operasyonu (Tekgül vd., 2010)

Cerrahi operasyon sonrası hastalarda kanama, yara ve idrar yolu enfeksiyonu, kordi (peniste eğrilik), üretral darlık, ameliyat hattında idrar sızıntısı ortaya çıkabilir (Stokowski, 2004). Bu komplikasyonların ortadan kaldırılmasına yönelik yeni cerrahi müdahaleler gerekebilir (Snyder vd., 2005).

Veri madenciliği çeşitli şekillerde tanımlanmıştır. Jacobs, ham verinin doğrudan sunamadığı bilgiye, veri analizi yöntemleriyle ulaşılabilme süreci olarak tanımlamıştır (Jacobs, 1999). Doğan ve Türkoğlu, büyük veri yığınlarından geleceğe yönelik tahminler de bulunması için bilgisayar programlarında bağlantıların aranması olarak tanımlamışlardır (Doğan ve Türkoğlu, 2008). Kısaca veri madenciliği değişken verilerden sakınarak tahminci, anahtar bilgilere ulaşmaya denir.

Bilgi, toplumlarda teknolojilerin ilerlemesinde daima yardımcı olmuştur. Gelişen teknoloji de aynı şekilde yeni bilgilerin ortaya çıkmasına katkı da bulunmuştur. Teknolojinin gelişmesi ile ortaya çıkan bazı bilgiler anlamlıyken bazı bilgiler ise çok fazla bir önem içermemektedir. Bilgi, verinin işlenmesi sonucunda elde edilen anlamlı bir haldir (Kaçtıoğlu vd., 1999). Veriyi bilgiye çevirme işlemine ise veri analizi denilmektedir. Bu analiz sonucunda elde edilen sonuç ise bilgiyi temsil etmektedir (Baykal, 2006). Ancak her bilgi değerli değildir. Bir bilginin değerli olabilmesi için bir amaca hizmet etmesi gerekmektedir. Bunun yanında test edilebilmesi, gerçekleşmesi ve güncelliğini koruması da önemli niteliklerdendir.

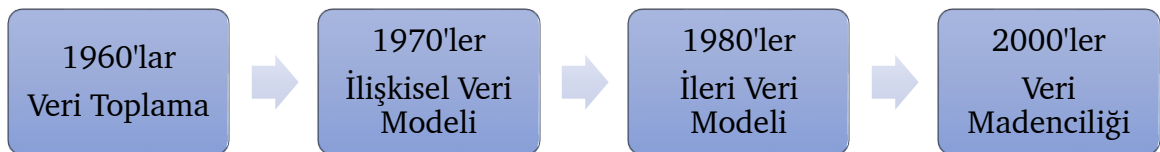
Tüm bu süreçlerin dijitale aktarılması ve bilginin düzenlenip saklanması ise bilişim sistemini ortaya çıkarmıştır. Organizasyondaki kararların verilmesi ve bilgi işleme sürecinin önemli bir parçası olan bilişim sistemi, formal bilgiyi göz önüne sermektedir.

Veri madenciliği çeşitli alanlarda kullanıldığı gibi tıp alanında da sıklıkla kullanılan bir yöntemdir. Sağlık alanında yapılan çalışmalar bilim alanına katkıda bulunurken veri kaynaklarından faydalanırlar (Yıldırım, vd., 2008). Potansiyel hastalıkların teşhisinde pek çok faktöre dikkat edilmesi gerekmektedir. Hastalığın teşhisinin hızlı ve doğru konulması, tedavinin bir an önce başlaması ve alınan kararların hızlı bir şekilde sonuç vermesi oldukça önemlidir (Çelik, 2020). Veri ambarından alınan bilgiler doğrultusunda gerek hastane yönetimi gerekse sağlık

alıřanları bu ynde alıřma politikaları geliřtirmektedirler. Bu alıřma stratejileri sonucunda hastalık karřısında en etkili m retilmeye alıřılmaktadır (Koyuncuđı ve zglbař, 2009).

3.1 Veri Madenciliđi Tarihi

Bilgisayar sistemleri 1950’li yıllardan itibaren kullanıma girerken, 1960’lı yıllardan itibaren veriler depolanmaya bařlanmıřtır. 1960’lı yılların sonuna dođru basit đrenmeli bilgisayarlar geliřtirilmiřtir. Minsky ve Papert, yapay sinir ađlarının ok basit bir řeklini kullanmıřlardır. İliřkisel veri tabanı yntemi ise 1970’li yıllarda kullanılmaya bařlanmıřtır (Adriaans ve Zantinge, 1997). 1980’li yıllarda veri tabanları sistemi daha yaygın hale gelmiřtir. Bu yıllardan itibaren veri tabanları piyasada zellikle mhendislik ve pazarlama alanlarında kullanılmaya bařlanmıřtır. 1980’li yıllardan itibaren bu verilerin analizi iin programlar dřnlmeye bařlanmıřtır. 1991’de yayımlanan “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop” makalesi (Piatetsky, 1990) veri madenciliđi srecine nemli katkı da bulunmuř ve bundan sonra daha bilimsel yaklařımlar benimsenmiřtir (Demirel ve Yakut, 2019). Temel kavramların tanınımının yapılması ve ihtiyalara ynelik programlamalar sayesinde veri madenciliđi sreci daha da hızlanmıřtır. 1992 yılında veri madenciliđi iin ilk yazılım hayata geirilmıřtir (Savař vd., 2012). 2000’li yıllardan itibaren dijitalleřme ile devasa veri yığınlarının iřlenmesi sz konusu olmuřtur (řekil3.1).

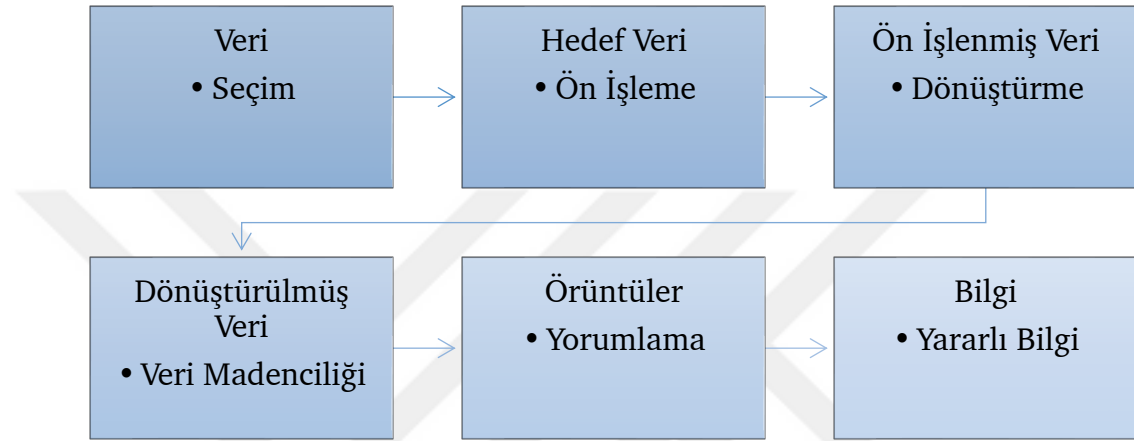


řekil 3.1 Veri madenciliđinin tarihi

3.2 Veri Madenciliđi Sreci

Veri madenciliđinde ilk ařama veri toplama ile bařlar. Bu adım ok nemlidir, nk analiz yapılıbilmesi iin veri ambarına ihtiya vardır. İlk ařamada elde edilen veriler ayıklanmamıř ve karıřık verilerdir. Veri temizleme ařamasında ise veriler, sınıflandırılır ve daha okunabilir hale getirilir. Bu n iřleme sayesinde modelin bařarısı artar. Veriler model oluřturmak iin hazır hale getirilir

(Silahtaroğlu, 2008). Verilere uygulanacak olan modelin seçim aşamasında ise pek çok faktör devreye girer. Burada çalışmanın amacına en uygun yöntem seçilir ve analize hazır hale getirilir. Bu aşamada farklı algoritmalar uygulanır ve algoritmalar kendi aralarında karşılaştırılarak en uygunu seçilir. Yararlı bilgiye ulaşmada oldukça önemli olan bu adımda araştırmacı dikkatli olmalıdır. Nihai adım olarak uygun model uygulanır ve en başarılı algoritma belirlenmiş olur. Çıkan sonuçlardan yararlı bilgi edinilmiş olur (Şekil 3.2).



Şekil 3.2 Veri tabanında bilgi keşif süreci

3.3 Veri Madenciliği Uygulama Alanları

Veri madenciliği, son yıllarda oldukça ihtiyaç duyulan güncel bir alandır. Dijitalleşen dünyada veri toplama kapasitesi artmış ve veri toplama biçimleri de çeşitlenmiştir. Kurumların ve toplulukların elinde veri yığınları birikmeye başlamıştır. Bu yığınlardan faydalı olan bilgiye ulaşmak için veri biliminden yardım alınmaktadır. Farklı disiplinlerde biriken bu işe yarar veriler, son yıllarda pek çok alanda kullanılmaya başlanmıştır (Savaş vd., 2012). Bu devasa bilgi havuzundan işe yarar bilgiye ulaşmak için makine öğrenmesine başvurulmaktadır. Sosyal bilimler, sağlık alanı, bankacılık sektörü gibi pek çok alanda kullanılan bu verilere ulaşmada farklı yollar izlenmektedir.

Veri madenciliğinin sağladığı faydaların başında bilgi kaybının önlenmesi ve geçmişte yaşanan problemlerin ileriye yönelik öngörülmesi gelmektedir (Savaş vd., 2012). Büyük miktardaki verilerin içine saklanmış örüntülerin keşfedilmesi ile söz konusu alan için geleceğe dair tahminlenme yapılabilmektedir. Giderek

önemi artan veri madenciliği, şu anda alternatif bir araç olarak kullanılsa da gelecekte kaçınılmaz bir uygulama yöntemi olacağı tahmin edilmektedir.

Veri madenciliğinin kullanıcılara sunduğu özellik veri tabanında bulunan benzer örüntü ve davranışları tespit edebilmesidir. Bilim alanlarının, kurumların ve işletmelerin genellikle faydalandığı kısım bu örüntülerdir. Özellikle pazarlama ve toplumsal fayda sağlayıcı alanlarda yoğun olarak kullanıldığı gözlemlenmektedir (İnan, 2003). Dinamiklerin hızla değiştiği ve risk faktörünün daima bulunduğu alanlarda sezgilere yönelik karar alma süreci kurumların ve bireylerin inisiyatifine bırakılmamalıdır. Riski en aza indirecek ve bilgiye dayalı ön görüde bulunacak olan veri bilimcilere bu noktada oldukça ihtiyaç duyulmaktadır.

Tıbbi literatürde veri madenciliğinin kullanım gerekçesine yönelik pek çok çalışma bulunmaktadır (Koyuncugil ve Özgülbaş, 2009). Tıbbi verilerin kesin olmayan ve geleneksel yöntemler ile analiz edilmesinin zorluğu, dijitalleşmeyle birlikte büyük veri yığınlarının okunamaz hale gelmesi, pek çok farklı belirtiler ile gelen hastaya teşhis konulmasının zorluğu, nüfusun artması ile birlikte sağlık hizmetlerine talebin de artması bunlardan birkaçıdır (Çarklı, 2010).

Veri madenciliği yalnızca biyolojik anlamda değil hastanede ve klinik yönetim sürecinde de kullanılmaktadır. Görüntüleme, teşhis koyma, prognoz gibi hastalığı doğrudan ilgilendiren alanların yanı sıra epidemiyolojik çalışmalar, tıbbi yönergelerde de kullanılmaktadır (Koyuncugil ve Özgülbaş, 2009), (Şekil 3.3).



Şekil 3.3 Veri madenciliğinin sağlıkta kullanım alanları

3.4 Veri Madenciliği Metotları

3.4.1 Sınıflandırma ve Regresyon

Veri madenciliğinde sınıflandırma yöntemi karışık halde duran verilerin okunmasında kullanılır. Mevcut verileri belli bir özelliğe göre sınıflandırmak analizi kolaylaştırmaktadır. Dolayısıyla yeni giren bir verinin hangi sınıfa ait olduğu kısa yoldan bulunmaktadır. Sınıflandırma yapılırken verinin örüntü geçmişine bakılır. Bu sayede hangi sınıfa ait olduğu belirlenir.

Öngörü yöntemi de sınıflandırmaya çok benzer ancak “bekle ve gör” prensibi kullanılır. Öngörü daha çok sınıflandırmayı test etmek amacıyla kullanılmaktadır. Deprem tahmini buna örnek olarak verilebilir.

Sınıflandırma yöntemini kullanabilmek için iki aşama gereklidir. İlk aşamada gözlenmiş eğitim verisinden sınıflandırma kuralları oluşturulur. Ardından sınıflandırma kurallarından yararlanılarak yeni test verisinin sınıflandırılması gerçekleşir. Öngörü de ise bu verilerden yola çıkılarak sonuca ulaşılır.

Regresyon ise bağımsız değişkenler ile amaç değişkeni arasındaki bağlantıyı matematiksel olarak formalize eden yöntemdir. Sınıflama kategorik veriler üzerinden tahminleme yaparken regresyon sürekli veriler için kullanılır. Örneğin

banka kredi uygulamalarının riskli ya da güvenli olma durumunu tespit etmede sınıflama yöntemi kullanılırken; regresyon modeli ise müşterinin gelir durumuna göre ne kadarlık bir ürün satın alabileceğinin tespit edilmesinde kullanılır (Özekes, 2003).

Bu modellerde karar ağaçları (decision tree) yapay sinir ağları (artificial neural network), K-en yakın komşuluk (K-nearest neighbours), genetik algoritmalar (genetic algorithm), Naive Bayes, bellek temelli nedenleme (memory based reasoning) ve hatayı geri yayma (back propagation) algoritmaları kullanılmaktadır (Baykal, 2003).

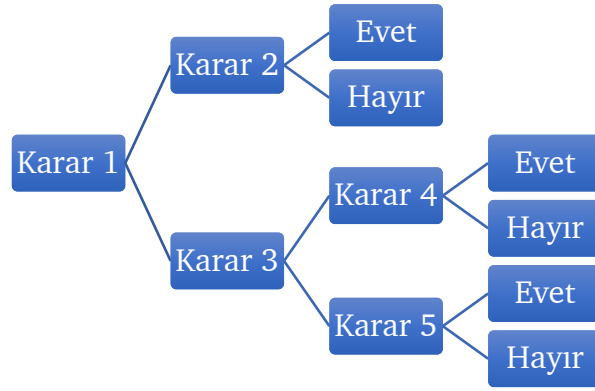
3.4.1.1 Navie Bayes

Sınıflandırma yönteminin bir algoritması olan Bayes ile sınıf üyeliği olasılığı hesaplanmaktadır. Naive Bayes ise tahmin etmenin yanı sıra verilerin sınıfını yani kategorisini tespit etmekte kullanılmaktadır. Bağımlı ve bağımsız değişkenler arasındaki ilişkiselliğin analizinde başvuru bir yöntemdir (Berrar, 2018).

3.4.1.2 Karar Ağaçları

Karar ağacı, bir amaca ulaşma olasılığı en yüksek olan yolu tespit etme üzerine kurulu bir modeldir. Makine öğrenmesinin gerçekleşmesiyle ihtimaller arasından en iyi yargıya ulaşırken genellikle başvuru bir yöntemdir. Pek çok parametrenin, ardışık ilişkinin ve örüntünün olduğu durumlarda kullanılan karar ağacı karmaşık verilerin analizini kolaylaştırmaktadır (Şekil 3.4). Karar ağacı yöntemi, makine öğrenmesi sürecinde birçok faktörü yan yana getiren ve bu faktörlere dayanan sonucu en kısa yoldan ortaya koymaktadır (Emel ve Taşkın, 2005).

Sınıflandırma uygulamasında kullanılan karar ağacı yönteminin güvenilir, basit, ucuz ve uyumlu olması tercih sebebidir. Bu yöntemi uygularken veri ambarındaki verilerin bir kısmı öğrenme için ayrılırken bir kısmı test aşamasında kullanılır. Adından da anlaşılacağı üzere karar ağacı yönteminde verilerin öğrenilmesi aşamasında ağaç dallanır ve istenilen düzeye gelindiğinde durdurulur. Bu da sürecin hassasiyetini ortaya koymaktadır. Doğru analiz yapılması ve verilerin okunabilir hale gelmesi için durdurma faaliyeti önemli bir adımdır. Aksi takdirde verilerin sınıflandırılması zorlaşacak ve araştırmacıya fazladan iş yükü çıkacaktır.



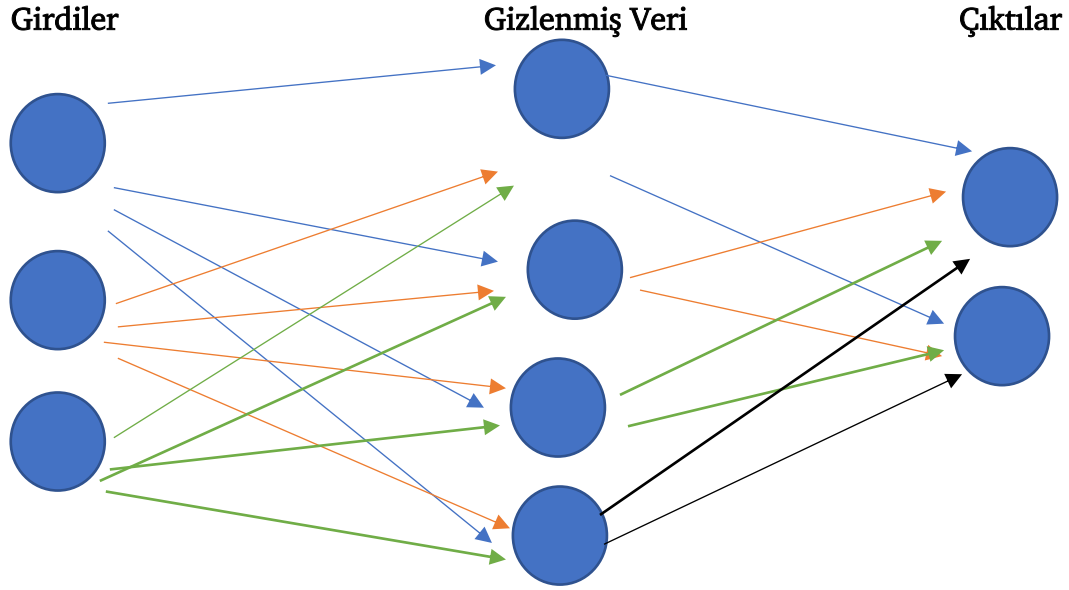
Şekil 3.4 Karar ağacı yapısı

3.4.1.3 K-En Yakın Komşu Algoritması

Sınıflandırma modelinin bir diğer algoritması ise k-en yakın komşudur. Bu algoritma sayesinde sınıfı henüz belirlenmemiş verilerin, eğitim setinde bulunan diğer veriler ile kıyaslanıp bir uzaklık ölçümü belirlenerek bu uzaklığa göre sınıflandırılması sağlanır. Daha önce sonuca varılmış durumlardan yola çıkarak yeni bir durum ile karşılaşıldığında en yakın komşu ile ilişkilendirilir. Geniş veri yığınlarında sınıflandırma yapılırken sıklıkla başvurulmuş bu yöntemde verinin komşusuna göre hareket etme özelliği söz konusudur. Örneğin bir veriden geleceğe yönelik tahminleme yapılması istendiği takdirde o veriye yakın komşuların davranışları incelenir, bu yakın komşuların davranışlarının ortalamasından bir sonuç çıkartılmaktadır. Buradaki “k” harfi komşu sayısını temsil etmektedir (Adriaans ve Zantinge, 1997).

3.4.1.4 Yapay Sinir Ağları

Yapay Sinir Ağları, tıpkı insan beyni gibi öğrenme sürecini temel alır. Öğrenme yolu ile yeni bilgiler ortaya çıkarmaktadır. Verilerden yola çıkarak keşfetme, yeni bilgi üretme gibi süreçleri sağlar. Yapay sinir ağları sınıflandırma uygulamasında daha karmaşık, belirsiz ve hata olasılığı yüksek verilerin işlenmesinde kullanılır. Şekil 3.5’te verilerin işlenmesine dair genel süreç resmedilmiştir. Geleneksel programların aksine daha gelişmiş bir yöntemdir. Özellikle bir matematiksel formül veya algoritmanın bulunmadığı durumlarda sıklıkla kullanılmaktadır. Sınıflandırma, örüntü, optimizasyon, ilişkilendirme gibi fonksiyonları yerine getirmektedir (Ergezer vd., 2003).



Şekil 3.5 Yapay sinir ağları

3.4.1.5 Genetik Algoritmalar

Genetik Algoritmalar, veri tabanından işe yarar veriyi bulmaya çalışırken kullanılan bir yöntemdir. Bu yöntem Darwin'in evrim kuramından esinlenilerek meydana gelen bir tekniktir. Bu tekniğe göre doğada en iyi uyum sağlayanın hayatta kalmasına benzer şekilde sanal dünyada da verilerin evrim geçirmesi söz konusudur. Genetik algoritmalar yapay sinir ağlarının eğitimi ve bellek tabanlı yöntemde kombinasyon işlemi için kullanılmaktadır. Genetik algoritma yönteminde beş aşama söz konusudur:

- Öğe grubunun başlatılması
- Çaprazlama tekniği
- Mutasyon algoritması
- Uygunluk işlevi
- Popülasyondaki en iyi sonuçları belirlemek için çaprazlama ve mutasyon tekniklerine uygulama (Baykal, 2003).

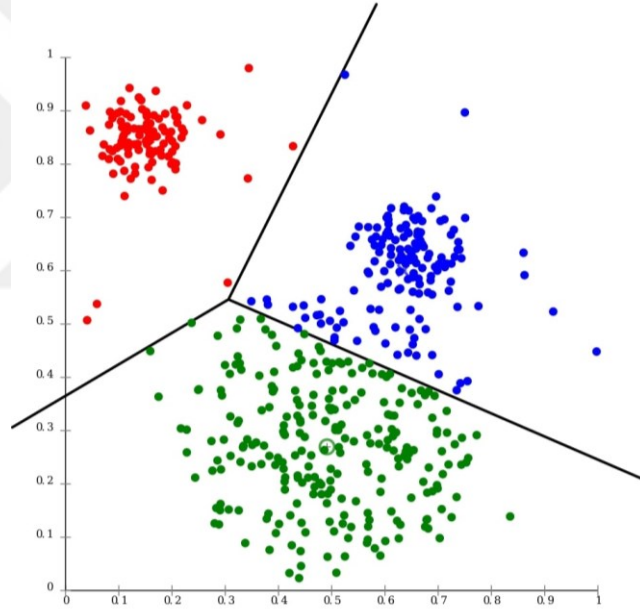
3.4.2 Destek Vektör Makinesi

Destek vektör makinesi (SVM) genellikle regresyon veya sınıflama problemlerinde kullanılan denetimli öğrenme yöntemlerinden biridir. Bu modelde örnekler, her bir öznelik belirli bir koordinata karşılık gelecek şekilde n boyutlu uzaya (burada n öznelik sayısı) nokta olarak yerleştirilir. Sonrasında bu noktaları ayıran hiper-

düzlemi tespit ederek sınıflama yapar. Bu hiper-düzlemi tespit ederken sınıflar arasındaki noktaları oldukça iyi ayırmayı amaçlar. Genellikle iki sınıfa ait veriyi ayırmada kullanılır (Kavzoğlu ve Çölkesen, 2010).

3.4.3 Kümeleme

Kümelemenin öncelikli amacı kendine benzer olan verileri bir araya getirerek gruplandırmaktır. Bu sayede değişkenler arası ilişki aranırken benzer parametrelerden faydalanılmış olunur. Dolayısıyla meydana gelen birimler homojendir. Kümeleme analizinde öncelikli hedef; dağınık, işlenemez haldeki veri yığınına çalışmanın parametrelerinden yola çıkarak benzer özelliklerini tespit edip ona göre gruplandırmaktır (Şekil 3.6). Bu sayede büyük veri yığınları, daha okunabilir ve işlenebilir hale getirilir (Akın, 2008).



Şekil 3.6 Kümeleme örneği (Mesci, 2020)

Kümeleme yönteminin pek çok algoritması bulunmaktadır. Bazıları şunlardır:

- Bölme yöntemleri
- Hiyerarşik yöntemler
- Yoğunluk tabanlı yöntemler
- Izgara tabanlı yöntemler
- Model tabanlı yöntemler

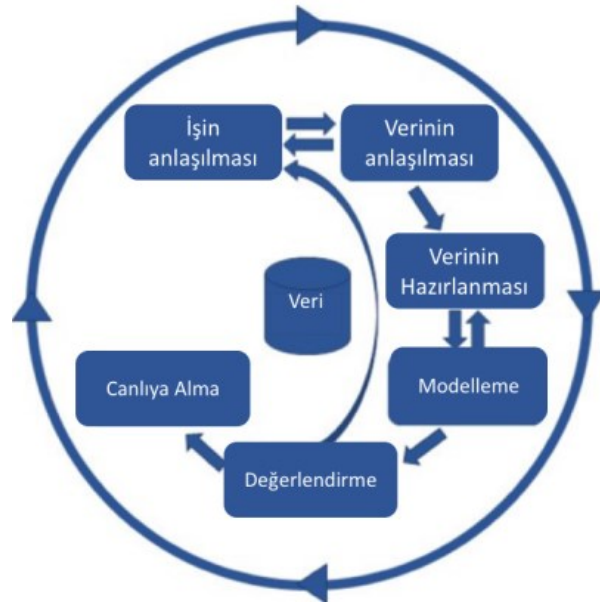
3.4.4 Birliktelik Kuralları

Birliktelik analizi veri ambarında sıklıkla yan yana gelen veriler için kullanılır. Birlikte oluşma, satın alınma, üretilme gibi durumlarda karşımıza çıkmaktadır. Örneğin bir market sepetinin içindeki ürünlerde kullanılabilir. Müşterinin hangi ürünleri birlikte aldığı sorusunun cevabı bu analiz ile verilmektedir. Bu hangi koşullarda hangi verilerin bir araya geldiği konusunda da araştırmacıya bilgi vermektedir. Pazar piyasasında sıklıkla kullanılan bu yöntem ile potansiyel müşteriler tespit edilmektedir (Emel vd., 2005).



HİPOSPADİAS OLGULARINDA İNFLAMATUAR SÜRECİN TESPİTİ

Bu çalışmada veri madenciliği metodolojilerinden olan CRIPS-DM metodolojisi kullanılmıştır. Bu metodolojinin ilk adımı işi anlamaktır. Problem ve iş süreçlerinin tasarımı bu aşamada yapılır. Sonrasında verinin anlaşılması safhasına geçilir. Bu adımda amaca yönelik veriler toplanır, nitelikleri anlaşılmaya çalışılır ve verinin probleme ne kadar çözüm getirdiği irdelenir. Daha sonra veri ön işleme adımına geçilir. Bu aşamada toplanan veriden veri seçme, tutarsız veya eksik verileri temizleme, veri dönüştürme gibi işlemler yapılır. Böylece modelleme aşamasına gelinmiş olur. Hazırlanmış olan veriler üzerinden veri madenciliği yöntemleriyle uygun modeller geliştirilir. Probleme uygun modeller tespit edildikten sonra test verileri oluşturularak bu modeller test edilir. Daha sonra değerlendirme adımına geçilir. Bu kısımda tespit edilen modellerin sonuçları problemin çözümüne katkı sağlamasına göre yorumlanır. Bütün bu aşamalardan sonra tespit edilmiş modeller çalıştırılarak amaca yönelik kullanılır. Şekil 4.1’de CRIPS-DM metodolojisinin basamakları gösterilmiştir.



Şekil 4.1 CRIPS-DM metodolojisi (Şeker, 2018)

4.1 İŖi Anlama

4.1.1 İŖin Hedeflerinin Belirlenmesi

Hipospadias olgularında üretral plakada inflamasyon gözlenmektedir. Hemogram testi ise inflamatuvar süreçler hakkında bilgi vermektedir. Hastane veri tabanından alınan bilgilerde hipospadias olgusuna sahip olan ve sünnet talebiyle gelmiŖ olan çocukların hemogram test sonuçları bulunmaktadır. Bu bağlamda hipospadias olgusuna sahip olan ve olmayan bireylerde hemogram test sonuçları aracılığıyla inflamatuvar sürecin ortaya konulması amaçlanmaktadır.

4.1.2 Durum Deęerlendirmesi

Hipospadias erkek çocuklarında genital muayene ile tespit edilebilmektedir. Genital muayeneden geçmeyen ya da ailenin anomaliyi fark etmemesi gibi durumlarda gerekli cerrahi operasyon gecikebilmekte ya da yapılamamaktadır. Hipospadiasın tedavisi cerrahidir ve önerilen operasyon zamanı çocuęun 6 ile 18 ay yaŖları arasındadır. Daha sonra yapılan cerrahi müdahale, vakalarda fonksiyonel ve psikolojik açıdan problemlerin artmasını beraberinde getirmektedir (Tekgöl vd., 2016; Jones vd., 2009). Ayrıca hipospadias ameliyatı için genellikle sünnet derisi cerrahi iŖlem sırasında idrar kanalı oluŖturmak için kullanılır. Sünnet operasyonu geęirmiŖ çocuklarda idrar kanalının oluŖturulabilmesi için vücudun baŖka bölgelerinden parçalar alınması söz konusu olabilmektedir. Dünya üzerinde yaygın olarak yapılan sünnet operasyonunun hipospadiasa sahip çocuklarda da yapıldığı gözlemlenmektedir. Bu bağlamda hipospadias onarımının uygun zamanlamayla yapılabilmesi önem arz etmektedir.

4.1.3 Veri Madencilięinin Amacının Belirlenmesi

Hemogram vücutta var olan herhangi bir inflamasyona yanıtı göstermekte yardımcı olan testlerden biridir. Hemogram testinin yaygın ve kolay yapılmasından dolayı hastanelere baŖvuran hastalardan sıklıkla talep edilmektedir. Bu sayede hastalara ait veri tabanlarında büyük miktarda veri saęlanmış olmaktadır. Aynı zamanda bu veri tabanları hastalara konulan tanı bilgilerini de içermektedir. Dolayısıyla veri madencilięi perspektifinde bakıldığında eğitim ve test için yeterli miktarda veriye ulaŖılabilir.

4.1.4 Proje Planı Üretme

Hastane veri tabanlarından alınan ve birbirinden ayrık veri yığınları Excel ve SQL programları aracılığıyla birleştirilerek laboratuvar sonuçlarındaki hemogram değerlerinden hareketle MATLAB programında veri madenciliği yöntemleri ile makine öğrenmesi yapılmıştır. Hipospadiasa sahip hastalar ile bu hastalığa sahip olmayan örneklerin sonuçları etiketlenerek denetimli öğrenme metoduyla inflamatuvar sürecin ortaya konulması amaçlanmaktadır.

4.2 Veriyi Anlama

4.2.1 İlk Verilerin Toplanması

Bu çalışma için, 23.02.2021 tarihli 2021.58.02.21 protokol numaralı etik kurul izni ile Tekirdağ Namık Kemal Üniversitesi Çocuk Cerrahisi Anabilim Dalı'na ait veri tabanından alınan hasta anamnezi ve laboratuvar sonuçlarından yararlanılmıştır. İki ayrı Excel dosyası şeklinde alınan bu bilgiler çalışmanın verilerini oluşturmaktadır. Anamnez dosyasında hastaların başvuru numarasına göre hikâye, şikâyet veya muayene sonuçlarını ve hastaların başvuru zamanlarını belirten 4 ayrı sütunda olmak üzere toplamda 48.789 satırlık Excel verisi mevcuttur (Şekil 4.2). Veriler incelendiğinde 48.789 satırlık veriden 15.532 farklı başvuru numarasına ait veri olduğu görülmüştür.

1	T_BASVURUNO	T_CESIDI	T_SAAAT	T_TEXT
2	17773	HİKAYESİ	4.01.2010 11:13:34	USG solda minimal genişleme
3	17773	ŞİKAYETİ	4.01.2010 11:13:34	Tekrarlayan UTİ
4	17799	MUAYENE	4.01.2010 11:30:06	Palpasyonla yabancı cisim ele geliyor
5	17799	ŞİKAYETİ	4.01.2010 11:30:06	Sol göz kenarında cam kesisi sonrası dikiş atılmış, bir parlama farketmişler
6	20144	ŞİKAYETİ	14.01.2010 11:59:49	Boyunda kitle, daha önce geldiklerinde ellerindeki USG higroma ?, Sn'de tiroid no
7	20178	HİKAYESİ	14.01.2010 12:14:47	5 aydır farketmiş
8	20178	MUAYENE	14.01.2010 12:14:47	Her 2 testis skrotumun üst tarafında, retraktil testis açısından liste yapacaklar
9	20178	ŞİKAYETİ	14.01.2010 12:14:47	Sağ tarafta şişlik
10	20156	ŞİKAYETİ	14.01.2010 12:25:58	Kabızlık tedaviden fayda görmüş Arada çiş kaçırması var
11	20145	ŞİKAYETİ	14.01.2010 12:44:13	Göbekte akıntı
12	20145	MUAYENE	14.01.2010 12:44:13	Yd döneminde granulom bağlanmış, ara ara minik kan tarzı akıntı oluyormuş
13	20274	ŞİKAYETİ	14.01.2010 16:02:29	yANAĞINA KALEM BATMIŞ
14	20672	ŞİKAYETİ	18.01.2010 11:19:21	Kabızlık
15	20672	HİKAYESİ	18.01.2010 11:19:21	1 ay önce tedavi başlamıştı fayda görmüş kaçırımları azalmış
16	20672	MUAYENE	18.01.2010 11:19:21	Batın rahat
17	20688	ŞİKAYETİ	18.01.2010 11:48:16	2 haftalık erkek bebek, sakral bölgede hairy lezyon akıntı yok, kızarıklık, ilk gün bol m
18	20688	MUAYENE	18.01.2010 11:48:16	Sakral bölgede kızarıklık lezyon Orifis görülemedi Anal stenoz var Fimoziş yok Testis
19	20688	HİKAYESİ	18.01.2010 11:48:16	Çocuk Cerrahisi ve NRŞ görmüş, ADBG'de özellik yok denmiş
20	20733	HİKAYESİ	18.01.2010 12:15:16	Marmarada takipli
21	20733	ŞİKAYETİ	18.01.2010 12:15:16	Hidroşefali, doğumdan 15 gün önce saptanmış

Şekil 4.2 Hastane veri tabanından alınmış anamnez dosya görüntüsü

Laboratuvar dosyası iki ayrı sayfadan oluşmakta ve sayfalar arası başvuru numaraları birebir uyuşmamaktadır. Sayfalar hastaların farklı türden laboratuvar sonuçlarına göre ayrılmıştır. Birinci sayfa 18 sütun ve 65.535 satırdan; ikinci sayfa ise 35 sütun ve 12.563 satırdan oluşmaktadır. Sütunlarda başvuru numarası, test adları, test sonuçlarının değerleri, kabul tarihi, sonuç tarihi gibi öznitelikler bulunmaktadır. Birinci sayfada 2.874, ikinci sayfa 2.879 adet farklı başvuru numarasına sahip veri bulunmaktadır (Şekil 4.3).

1	T_BASVURU	T_BARKOD	T_TESTAD	T_P_ADI	T_DEGER	T_BIRIM	T_KABULTARİH	SONUCTARİH	T_DOSYANUM	T_CINSIYE
2	1000883	00402322	APTT	APTT	35,800	sn	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
3	1000883	00402322	HEMOGRAM	BA#	0,090	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
4	1000883	00402322	HEMOGRAM	BA%	1,000	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
5	1000883	00402322	HEMOGRAM	EO#	0,380	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
6	1000883	00402322	HEMOGRAM	EO%	4,200	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
7	1000883	00402322	HEMOGRAM	HCT	29,400	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
8	1000883	00402322	HEMOGRAM	HGB	9,800	g/dl	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
9	1000883	00402322	HEMOGRAM	LY#	5,210	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
10	1000883	00402322	HEMOGRAM	LY%	57,800	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
11	1000883	00402322	HEMOGRAM	MCH	29,200	pg	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
12	1000883	00402322	HEMOGRAM	MCHC	34,000	g/dl	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
13	1000883	00402322	HEMOGRAM	MCV	87,000	fL	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
14	1000883	00402322	HEMOGRAM	MO#	0,940	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
15	1000883	00402322	HEMOGRAM	MO%	10,400	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
16	1000883	00402322	HEMOGRAM	MPV	7,600	fL	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
17	1000883	00402322	HEMOGRAM	NE#	2,400	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
18	1000883	00402322	HEMOGRAM	NE%	26,600	%	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
19	1000883	00402322	HEMOGRAM	PDW	12,000	fL	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
20	1000883	00402322	HEMOGRAM	PLT	400,000	10 ³ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek
21	1000883	00402322	HEMOGRAM	RBC	3,400	10 ⁶ /ul	24.03.2014 11:40:29	24.03.2014 13:28:56	146695	Erkek

Şekil 4.3 Hastane veri tabanından alınmış laboratuvar dosya görüntüsü

4.3 Veri Ön İşleme

4.3.1 Veri Seçme

İki ayrı dosyadaki verilerin birbirleriyle olan bağlantısı başvuru numarasıdır. İlk olarak anamnez dosyası ele alındığında, hastanın hastaneye gelerek şikayetlerini anlattığı muayene ve şikâyet metni sütunlar halinde yer almaktadır. Excel fonksiyonları aracılığıyla her bir satırdaki veriler sadece bir tane başvuru numarası içerecek şekilde 4 tane sütun ve 15.532 satırlık bir excel dosyasına indirgenmiştir (Şekil 4.4).

1	BASVURU_NO	HİKAYE	ŞİKAYET	MUAYENE	HİKAYE_TARİH
2	17773	USG solda minimal genişleme	Tekrarlayan UTİ	YOK	4.01.2010 11:13:34
3	17799	YOK	Sol göz kenarında cam kesisi son	Palpasyonla yabancı cisim ele gel	YOK
4	20144	YOK	Boyunda kitle, daha önce geldi	YOK	YOK
5	20178	5 aydır farketmiş	Sağ tarafta şişlik	Her 2 testis skrotumun üst tarafı	14.01.2010 12:14:47
6	20156	YOK	Kabızlık tedaviden fayda görmüş	YOK	YOK
7	20145	YOK	Göbekte akıntı	Yd döneminde granulom bağlan	YOK
8	20274	YOK	YANAĞINA KALEM BATMIŞ	YOK	YOK
9	20672	1 ay önce tedavi başlamıştı	Kabızlık	Batın rahat	18.01.2010 11:19:21
10	20688	Çocuk Cerrahisi ve NRS görmüş, A12	haftalık erkek bebek, sakral böl	Sakral bölgede kızamık lezyon	Orif 18.01.2010 11:48:16
11	20733	Marmarada takipli	Hidroşefali, doğumdan 15 gün ö	YOK	18.01.2010 12:15:16
12	20777	YOK	Kabızlık	YOK	YOK
13	21434	YOK	Burun kanaması, sert kaka yapm	Batın rahat, solunum sesleri doğ	YOK
14	21564	YOK	1 GÜN ÖNCE SAĞKASIKTA ŞİŞLİK	YOK	YOK
15	21597	YOK	Bilateral inmemiş testis sünneti	YOK	YOK
16	21557	UTİ yok balanopostit yok rahat	ç Pipiden	Hafif fimotik Testisler skrotumda	21.01.2010 13:19:28
17	21783	YOK	Kabızlık tedavisi başlandı	Çok me	YOK
18	21835	YOK	Zayıflık	Testisler skrotumda, fimozis (+),	YOK
19	22743	YOK	İdrar ve kaka kaçırma	YOK	YOK
20	23319	AJİTE	5 AYDAN BERİ KARIN AĞRISI, BA	YOK	1.02.2010 11:17:07
21	23377	YOK	1. 5 AYDAN BERİ ÖKSÜRÜK GEÇİYOR	YOK	YOK

Şekil 4.4 Düzenlenmiş anamnez dosya görüntüsü

Laboratuvar dosyası iki sayfa içerdiği için ‘laboratuvar 1’ ve ‘laboratuvar 2’ olmak üzere iki ayrı Excel dosyası oluşturulmuştur.

‘Laboratuvar 1’ dosyası düzenlenerek 2.874 satırlık veri elde edilmiştir. Bu veriler arasından 3000905, 3003299, 650069 başvuru numaralı kayıtlar, ilgili hastaların T.C. kimlik numaraları olmadığı için silinmiştir. Son durumda laboratuvar 1 dosyası 2.871 satırlık 7 sütunluk dosyaya indirgenmiştir.

1	BASVURU_NO	APTT	APTT	HEMOGRAM_BA#	HEMOGRAM_BA#	HEMOGRAM_EO#	HEMOGRAM_EO#	HEMOGRAM_HCT	HEMOGRAM_HGI	HEMOGRAM_LY#	HEMOGRAM
2	1000883	35,8	0,09	1	0,38	4,2	29,4	9,8	5,21		
3	1000921	33,1	0,12	1,6	0,38	4,9	33,4	10,7	4,53		
4	1003039	29,7	0,11	1,2	0,32	3,5	38,4	12,4	5,37		
5	1003440	33,1	0,09	1	0,24	2,7	35	11,5	5,54		
6	1003559	31,1	0,07	1	0,39	5,3	25,5	7,5	3,79		
7	1003810	34,9	0,06	0,6	0,36	3,6	33,9	10,9	6,71		
8	1004981	YOK	0,04	0,5	0,13	1,6	39,5	12,7	2,57		
9	1009957	YOK	0,06	0,8	0,3	4	39,9	13	1,98		
10	1010116	32,7	0,12	1,2	0,46	4,6	36,6	11,5	7,33		
11	1010268	YOK	YOK	YOK	YOK	YOK	YOK	YOK	YOK		
12	101419	27,1	0,04	0,4	0,18	1,9	45,3	15,7	2,59		
13	1014210	30,2	0,11	1,5	0,48	6,7	36,6	12,1	4,91		
14	1014248	YOK	0,12	0,9	0,41	3,1	34	10,9	5,39		
15	1016675	YOK	0,07	1	0,33	4,9	42,5	13,8	2,99		
16	1016746	30,2	0,06	0,8	0,19	2,5	31,6	10,3	4,07		
17	1018257	28	0,07	0,8	0,09	1	37,1	11,5	2,84		
18	101868	27,6	0,03	0,4	0,28	4,1	33,8	11,2	3,54		
19	1019300	28,8	0,04	0,5	0,26	2,9	36,7	12,1	3,56		
20	1019322	34,6	0,03	0,4	0,12	1,9	YOK	11,9	3,45		
21	1020480	25,7	0,09	1,6	0,16	3,1	36,1	11,6	2,68		

Şekil 4.5 Düzenlenmiş laboratuvar dosya görüntüsü

Aynı şekilde 'laboratuvar 2' dosyası önce 2.879 örneğe daha sonra üç adet hasta verisi T.C. kimlik numarası sebebiyle silindiği için 2.876 örneğe düşürülmüştür (Şekil 4.5).

Daha sonra düzenlenmiş olan bu Excel dosyaları SQL programına aktarılarak her bir hastayı temsil eden başvuru numarası ile anamnez, 'laboratuvar 1' ve 'laboratuvar 2' dosyalarındaki düzenlenmiş olan sütunlar birleştirilip tek bir dosyada tüm veriler toplanmıştır. Dosyaları birleştirme işlemi sırasında anamnezdeki başvuru numaraları referans alınmıştır. Diğer dosyalar SQL programı aracılığıyla ortak ve eşsiz olan bu başvuru numaraları ortaklığında "left join" komutuyla birleştirilmiştir. Bu birleştirme işleminde diğer üç dosyada sonuçları bulunmayan veriler varsa o hücrelere "YOK" yazılmıştır.

Böylelikle her bir tanı tek bir başvuru numarası ile eşleştirilmiştir. Ve yine tek bir satırda ilgili başvuru numarasına ait laboratuvar dosyasındaki veriler birleştirilmiş oldu. Bu dosya birleşimi ile 15.532 satırlı ve 55 sütunlu tüm dosyaların birleştirildiği bir dosya elde edilmiştir. Daha sonra Excel programına tekrar tek bir dosya formatında aktarılmıştır.

Bu dosyadan başvuru numaralarına göre hemogram parametreleri ve tanı metinleri alınarak yeni bir Excel dosyası oluşturuldu. Elde edilen Excel dosyası anamnez başvuru numaralarına göre oluşturulduğu için 15.532 satırlık bir dosya idi. Bunların laboratuvar sonuçları olmayan 12.854 tanesi silindi. Böylelikle başvuru numarası, 24 hemogram parametresi ve tanı özniteliklerinin olduğu 2.677 satırlık veri seti elde edildi.

4.3.2 Veri Temizleme

Hemogram parametrelerini içeren özniteliklerinden, çok az miktarda veriye sahip olan PCT, P-LCR ve RDW-SD öznitelikleri silindi. Ayrıca tanı sütununda birden fazla hastalığa sahip hücreler, sahip olunan hastalık sayıları adedince (satır sabit kalmak şartıyla) farklı sütunlara yerleştirildi. Bunun için hastalıkların ICD (International Statistical Classification of Diseases and Related Health Problems) kodları referans alındı. ICD kodları irdelendiğinde veri grubunun toplamda 163 adet farklı hastalık türüne sahip olduğu tespit edildi. Tespit edilen ICD kodları, bu

kodlara karşılık gelen sağlık problemleri ve bu hastalığa sahip kişilerin sayısı Tablo 4.1’de gösterilmiştir.

Tablo 4.1 ICD kodları ve sağlık problemleri

ICD KODLARI	KODLARA KARŞILIK GELEN SAĞLIK PROBLEMLERİ	BAŞVURU SAYISI
N48	PENİSİN DİĞER BOZUKLUKLARI	581
K40	İNGUİNAL HERNİ	477
Q53	İNMEMİŞ TESTİS	304
Q54	HİPOSPADİAS	216
N47	SÜNNET DERİSİNİN FAZLALIĞI, FİMOZİS VE PARAFİMOZİS	165
N43.2	HİDROSEL, DİĞER	151
Q53.1	İNMEMİŞ TESTİS, UNİLATERAL	127
R10.4	KARIN AĞRISI	125
Q53.2	İNMEMİŞ TESTİS, BİLATERAL	92
K59.0	KABIZLIK	54
R82.7	İDRARIN MİKROBİYOLOJİK İNCELEMESİNDE ANORMAL OLGULAR	42
I88	LENFADENİT, NONSPESİFİK	41
K42	UMBİLİKAL HERNİ	34
Q52.5	LABİAL FÜZYON	31
K76	KARACİĞERİN DİĞER HASTALIKLARI	30
N39.4	ÜRİNER İNKONTİNANSLAR TANIMLANMIŞ	20
P83.5	KONJENİTAL HİDROSEL	17
K40.9	TEK TARAFLI VEYA TANIMLANMAMIŞ İNGUİNAL HERNİ	16
F98.0	ORGANİK OLMAYAN ENÜREZİS	15
Q38.1	ANKİLOGLOSSİ	15
N13.0	HİDRONEFROZ, ÜRETEROPELVİK BİLEŞKE OBSTRÜKSİYONU İLE	13
K14.8	DİLİN DİĞER HASTALIKLARI	11
L05	PİLONİDAL KİST	10
N62	MEMENİN HİPERTROFİSİ	10
Q44.6	KARACİĞER KİSTİK HASTALIĞI	10
N13.3	HİDRONEFROZ, DİĞER VE TANIMLANMAMIŞ	9
Q42.2	ANÜSÜN KONJENİTAL YOKLUĞU, ATREZİSİ VE STENOZU, FİSTÜLLÜ	9
Q54.1	HİPOSPADİAS, PENİL	9
D18	HEMANJİOM VE LENFANJİOM, HERHANGİ BİR YERDE	8
M43.6	TORTİKOLLİS	8
K35	AKUT APANDİSİT	7
K36	APANDİSİT, DİĞER	7
M79.9	YUMUŞAK DOKU BOZUKLUĞU, TANIMLANMAMIŞ	7
N43.3	HİDROSEL, TANIMLANMAMIŞ	7
I88.0	MEZENTERİK LENFADENİT, NONSPESİFİK	6
K40.2	BİLATERAL İNGUİNAL HERNİ	6
N51.1	TESTİS VE EPİDİDİMİSİN BOZUKLUKLARI	6
N63	MEMEDE TANIMLANMAMIŞ KİTLE	6
Q54.0	HİPOSPADİAS, BALANİK	6

Tablo 4.1 ICD kodları ve sağlık problemleri (devamı)

N39.9	ÜRİNER SİSTEMİN BOZUKLUĞU, TANIMLANMAMIŞ	5
K76.8	KARACİĞERİN DİĞER TANIMLANMIŞ HASTALIKLARI	4
L60	TIRNAK BOZUKLUKLARI	4
L92	DERİ VE SUBKUTANÖZ DOKUNUN GRANÜLOMATÖZ BOZUKLUKLARI	4
N35	ÜRETRA DARLIĞI	4
N39.0	ÜRİNER SİSTEM ENFEKSİYONU, YERİ TANIMLANMAMIŞ	4
N48.9	PENİSİN DİĞER BOZUKLUKLARI, TANIMLANMAMIŞ	4
Q42.3	ANÜSÜN KONJENİTAL YOKLUĞU, ATREZİSİ VE STENOZU, FİSTÜLSÜZ	4
Q54.2	HİPOSPADİAS, PENOSKROTAL	4
R10	ABDOMİNAL VE PELVİK AĞRI	4
R11	BULANTI VE KUSMA	4
R22.1	BOYNUN LOKALİZE ŞİŞME, KİTLE VE YUMRUSU	4
Z00.0	GENEL TIBBİ MUAYENE	4
B48.4	PENİSİLLOZ	3
D18.0	HEMANJİOM, HERHANGİ BİR YERDE	3
E07.9	TİROİD BOZUKLUKLARI, TANIMLANMAMIŞ	3

Bu hastalıklardan Q54.0, Q54.1, Q54.2, Q54.8, Q54, N48 ICD kodlu hastalıklar hipospadias kapsamına girmektedir.

Ayrıca N47 ve N48.9 ICD koduna sahip veriler ise sünnet operasyonu için hastaneye başvuran çocuklara aittir.

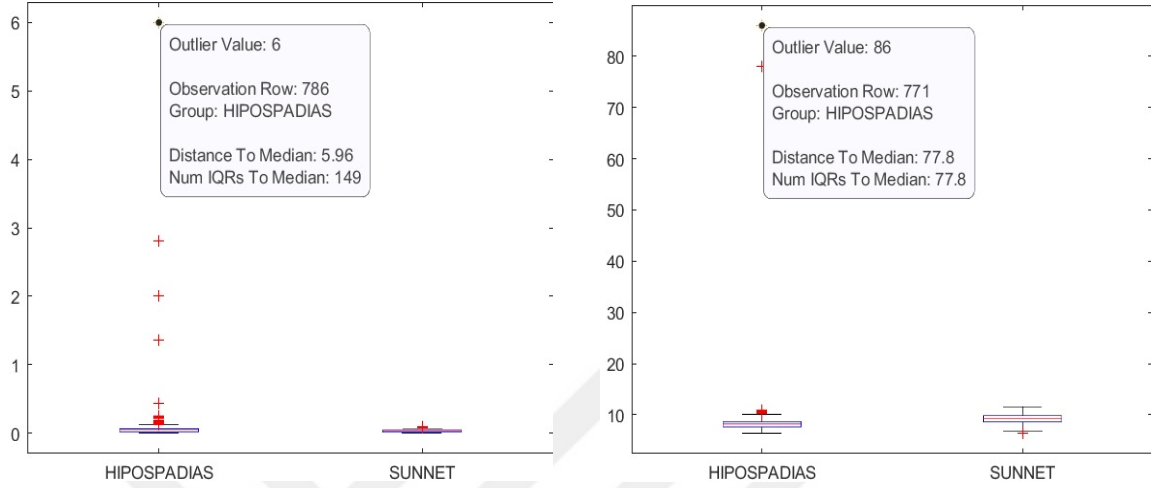
Hipospadias kapsamına giren anomalilere sahip çocuklara ait 803 tane örnek, sünnet operasyonu için hastaneye başvurmuş çocuklara ait 175 adet örnek ayrı bir Excel dosyasına alınmıştır. Oluşturulan son Excel dosyasında 978 satır veri ve başvuru numarası, 21 farklı hemogram parametresi ve olgunun hipospadias ya da sünnet olduğunu belirten “class” isimli öznitelik olmak üzere toplamda 23 sütun oluşturulmuştur.

Class sütunu oluşturulurken hem hipospadias hem de sünnet özneliğine sahip 10 adet örnek, sonuç itibarıyla hipospadias anomalisi taşıdığı için hipospadias grubuna dahil edilmiştir.

Geriye kalan 968 örnekten 45 tanesi hemogram değerleri tam olmadığı için silindi. Geriye 923 örnek veri kaldı. Bu verilerden 146 tanesi sünnet, 777 tanesi hipospadias etiketine sahiptir.

Elde edilen son dosya MATLAB programının içine aktarılmış, aykırı değerlere sahip iki adet başvuru numarası bulunmuştur. Bunlar 2974315 (771. satır) ve 3120769 (786. satır) başvuru numaralarına ait verilerdir. Bu iki başvuru

numarasının aykırı değer olarak tespit edildiği kutu diyagramları Şekil 4.6'da gösterilmiştir. 771. satırdaki örnek için gösterilen aykırı değer ortalama trombosit hacmi (MPV) olan hemogram parametresi iken, 786. satırdaki örnek için bazofil sayısı (BA) olan hemogram parametresidir (Şekil 4.6).



Şekil 4.6 Aykırı değerlerin tespiti

Tespit edilen bu iki aykırı örnek de silindikten sonra 921 satırlık ve 23 sütunluk veri seti makine öğrenmesi için son halini almıştır. Aşağıda sırasıyla temel istatistik analiz, makine öğrenmesi yöntemlerinin uygulanması ve elde edilen sonuçlar sunulmaktadır.

4.4 Modelleme

4.4.1 Özniteliklerin İstatistiksel Analizi

Ön işleme aşaması tamamlandıktan sonra 146 sünnetli ve 775 hipospadiaslı bireye ait hemogram parametrelerinden oluşan veri seti istatistiksel olarak analiz edilmiştir. Her bir hemogram parametresi için örneklem ortalamaları ve standart sapmaları Tablo 4.2'de verilmiştir. Her bir hemogram test sonucunu içeren özniteliklerin kutu diyagramları EK A'da verilmiştir.

Kolmogorov-Smirnov testi aracılığıyla hemogram parametrelerinin öncelikle normal dağılıma uyup uymadığına bakılmıştır. P değerinin 0,05'ten küçük olduğu değerler istatistiksel olarak anlamlı kabul edilmiştir. Bu test için hipotez testleri aşağıdaki gibidir.

- H_0 : Veriler normal dağılıma uymaktadır.
- H_1 : Veriler normal dağılıma uymamaktadır.

Kolmogorov-Smirnov testi uygulandığında elde edilen p değerleri Tablo 4.2’de verilmiştir. Her iki örneklem için p değerinin 0,05’ten büyük çıktığı hemogram parametresi yoktur. Dolayısıyla her iki örneklem için özniteliklerin hiçbirisi normal dağılıma uymamaktadır. Daha sonra her bir öznitelik için sünnetli ve hipospadiaslı hastaların hemogram değerleri arasında anlamlı bir fark olup olmadığını belirlemek adına non parametrik test yöntemi olan Man Whitney-U testi kullanılmıştır. Elde edilen p değerleri Tablo 4.2’de verilmiştir.

Tablo 4.2 İstatistiksel analiz test sonuçları

Öznitelik	Grup	Ortalama	Kolmogorov-Smirnov (p)	Mann Whitney-U (p)
BAZOFİL SAYISI	SÜNNET	0,028±0,016	0,000	0,000
	HİPOSPADİAS	0,051±0,110	0,000	
BAZOFİL YÜZDESİ	SÜNNET	0,327±0,204	0,000	0,000
	HİPOSPADİAS	0,631±1,510	0,000	
EOZİNOFİL SAYISI	SÜNNET	0,311±0,236	0,000	0,441
	HİPOSPADİAS	0,290±0,212	0,000	
EOZİNOFİL YÜZDESİ	SÜNNET	3,563±2,437	0,000	0,985
	HİPOSPADİAS	3,549±2,394	0,000	
HEMOTOKRİT	SÜNNET	35,979±3,36	0,039	0,966
	HİPOSPADİAS	36,150±3,322	0,000	
HEMOGLOBİN	SÜNNET	12,236±1,219	0,200	0,002
	HİPOSPADİAS	11,970±1,187	0,000	
LENFOSİT SAYISI	SÜNNET	4,433±1,992	0,000	0,265
	HİPOSPADİAS	4,137±1,713	0,000	
LENFOSİT YÜZDESİ	SÜNNET	49,711±13,320	0,006	0,782
	HİPOSPADİAS	49,644±13,203	0,003	
ORT. ERİTROSİT HEMOGLOBİN	SÜNNET	26,283±2,494	0,000	0,036
	HİPOSPADİAS	26,669±2,640	0,000	
ERİTROSİT HGB KONSANTRASYONU	SÜNNET	34,007±1,303	0,099	0,000
	HİPOSPADİAS	33,094±1,284	0,000	
ERİTROSİT HÜCRESİ ORTALAMA HACMİ	SÜNNET	77,269±6,775	0,000	0,000
	HİPOSPADİAS	80,536±7,306	0,000	
MONOSİT SAYISI	SÜNNET	0,805±0,274	0,000	0,000
	HİPOSPADİAS	0,652±0,297	0,000	
MONOSİT YÜZDESİ	SÜNNET	9,374±2,349	0,000	0,000
	HİPOSPADİAS	7,904±2,565	0,000	
ORTALAMA TROMBOSİT HACMİ	SÜNNET	9,165±0,862	0,200	0,000
	HİPOSPADİAS	8,241±0,815	0,000	
NÖTROFİL SAYISI	SÜNNET	3,179±1,436	0,014	0,444
	HİPOSPADİAS	3,166±1,594	0,000	

Tablo 4.2 İstatistiksel analiz test sonuçları (devamı)

TROMBOSİT SAYISI	SÜNNET	360,120±100,270	0,000	0,182
	HİPOSPADİAS	346,380±85,870	0,000	
ERİTROSİT SAYISI	SÜNNET	4,681±0,496	0,001	0,000
	HİPOSPADİAS	4,513±0,443	0,000	
ERİTROSİT DAĞ. GENİŞ.	SÜNNET	14,035±1,539	0,000	0,000
	HİPOSPADİAS	14,786±1,787	0,000	
LÖKOSİT SAYISI	SÜNNET	8,758±2,524	0,001	0,047
	HİPOSPADİAS	8,299±2,460	0,000	

Man Whitney-U testi için de p değerinin 0,05'ten küçük olduğu değerleri istatistiksel olarak anlamlı kabul edilmiştir.

- H_0 : Sünnetli ve hipospadiaslı bireylerin hemogram test sonuçları arasında anlamlı bir fark yoktur.
- H_1 : Sünnetli ve hipospadiaslı bireylerin hemogram test sonuçları arasında anlamlı bir fark vardır.

Tablo 4.2'ye bakıldığında bazofil sayısı, bazofil yüzdesi, hemoglobin, eritrosit hemoglobini, eritrosit hemoglobin konsantrasyonu, eritrosit hücresi ortalama hacmi, monosit sayısı, monosit yüzdesi, ortalama trombosit hacmi, trombosit dağılım genişliği, trombosit sayısı, eritrosit sayısı, eritrosit dağılım genişliği olmak üzere 13 tane hemogram testinin p değeri 0,05'ten küçük olduğu için H_0 hipotezi reddedilir. Dolayısıyla 13 adet hemogram test çeşidi sünnetli ve hipospadiaslı bireylere göre anlamlı bir fark içerir. Yine aynı şekilde, 8 tane hemogram test çeşidi için p değeri 0,05'ten büyük çıkması sebebiyle hemogram test sonuçları kişinin sünnetli veya hipospadiaslı olmasına göre anlamlı bir fark içermemektedir.

4.4.2 Modelin İnşası

Veri ön işleme safhasından geçmiş veri madenciliği yöntemlerine hazır olan veriye bakıldığında, hemogram değerlerinden hareketle vakanın hipospadias ya da hipospadias anomalisi taşımayan sünnet vakası olduğunun tespit etmesi amaçlanmaktadır.

Elimizdeki veri grubunda, modeli eğitmede kullanılacak girdi ve bunlara karşılık gelen çıktılar mevcuttur. Dolayısıyla denetimli makine öğrenmesi yöntemlerinden yararlanılmıştır. Bunun için MATLAB programının Makine Öğrenmesi Toolbox'ındaki 'Classification Learner' uygulaması kullanılmıştır. Verilerin %70'i

eğitim ve %30'u test için ayrılmıştır. Eğitim aşamasında 10 katmanlı çapraz doğrulama tekniği kullanılmıştır.

Geleneksel sınıflandırma yöntemlerinin çoğu eğitim veri setinin dengeli olduğunu, dolayısıyla her bir sınıfın yanlış sınıflandırma maliyetinin aynı olduğunu varsayar. Eğitim veri setinde sınıflardaki örnek sayılarında dengesizlik olduğunda, denetimli öğrenme algoritmaları doğruluk oranlarını arttırmak için örnek sayılarının fazla olduğu sınıfları seçme eğilimi içine girerler (He ve Garcia, 2009). Bu olumsuz durumu ortadan kaldırmak için yanlış sınıflandırma maliyetleri sınıflardaki örnek sayılarıyla orantılı bir şekilde seçilebilir. Elimizdeki 923 tane örneğin 146 tanesi sünnet ve 775 tanesi hipospadias olduğu için modellerde bu eğilimin önüne geçebilmek adına maliyet matrisi kullanılmıştır. Modeller çalıştırılmadan önce maliyet Tablo 4.3'te olduğu gibi oluşturulmuştur.

Tablo 4.3 Maliyet matrisi

	Tahmin Edilen Sınıf		
		Hipospadias	Sünnet
Doğru Sınıf	Hipospadias	0	146
	Sünnet	775	0

4.4.3 Modellerden Çıkan Sonuçların Değerlendirilmesi

Matlab programının 'Classification Learner' (2022b) adlı aracında tanımlanmış makine öğrenmesi yöntemleri uygulanarak çeşitli modeller eğitilmiş ve test veri seti kullanılarak doğruluk oranları hesaplanmıştır. Tablo 4.4'e bakıldığında en başarılı test sonuçlarının Medium Gaussian SVM (Gaussian çekirdekli DVM, çekirdek ölçeği=4.6), Optimizable Trees (Bayesian Optimization, iterasyon=30) ve Coarse Tree (maksimum bölme sayısı=4, bölme kriteri=Gini's Bölme) olduğu saptanmıştır.

Tablo 4.4 Model başarımlar oranları

Test Adı	Doğruluk (Eğitim)	Doğruluk (Test)
Medium Gaussian SVM	%89,8	%90,2
Optimizable Trees	%88,7	%90,2
Coarsa Tree	%87,8	%89,9

Tablo 4.4 Model başarımlar oranları (devamı)

Linear Discriminant	%87,6	%89,9
Coarsa Gaussian SVM	%89,1	%89,1
Optimizable SVM	%89,1	%89,1
Boosted Trees	%86,8	%89,1
Linear SVM	%89,0	%89,1
Quadratic SVM	%89,0	%89,1
RUS Boosted Trees	%87,0	%88,4
Opt Neural Networks	%89,0	%88,0
Bagged Trees	%88,5	%87,7
Kernel Naive Bayes	%87,4	%87,0
Opt. Naive Bayes	%87,4	%87,0
Subspace Discriminant	%87,3	%87,0
Optimizable Ensemble	%89,5	%86,2
Medium Tree	%86,0	%85,5
Bilayerd Neural	%85,9	%85,1
Optimizable KNN	%82,0	%84,8
Cubic SVM	%86,4	%84,4
Wide Neural Network	%89,5	%84,1
Medium Neural	%88,2	%84,1
Fine Tree	%85,9	%83,7
Trilayered Neural	%86,0	%83,7
Narrow Neural	%86,8	%83,3
Weighted KNN	%81,9	%83,3
Fine Gaussian SVM	%84,8	%83,0
Medium KNN	%80,8	%83,0
Cubic KNN	%81,4	%81,5
Fine KNN	%84,5	%80,8
Subspace KNN	%84,2	%80,8
Coarse KNN	%80,0	%79,7
Logistic Regression	%79,5	%79,3
Cosine KNN	%78,1	%79,0
SVM Kernel	%77,7	%76,1
Gaussian Naive Bayes	%68,5	%70,3

Tablo 4.5 Medium Gaussian *SVM* Yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri

		Tahmin Edilen Sınıf	
		Hipospadias	Sünnet
Doğru Sınıf	Hipospadias	210 (%91,7)	19 (%8,3)
	Sünnet	8 (%17,0)	39 (%83)

Medium Gaussian SVM modelinin karışıklık matrisi ve sınıflandırma sonuçları Tablo 4.5 'te verilmiştir. Algoritma; hipospadias olması gereken 19 örneği, sünnet

olarak sınıflamış ve %91,7 doğru tahminleme yapmıştır. Bununla birlikte sünnet sınıfında olması gereken 8 örneği de hipospadias olarak sınıflandırarak %83 doğruluk oranında tahminleme yapmıştır.

Tablo 4.6 Optimizable Tree yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri

		Tahmin Edilen Sınıf	
		Hipospadias	Sünnet
Doğru Sınıf	Hipospadias	205 (%89,5)	24 (%10,5)
	Sünnet	3 (%6,4)	44 (%93,6)

İkinci sırada yer alan Optimizable Tree modelinin karışıklık matrisi ve sınıflandırma sonuçları ise yukarıdaki iki tabloda verilmiştir. Hipospadias olması gereken 24 örnek, sünnet olarak sınıflandırılmış ve %89,5 doğruluk oranında tahminleme yapılmıştır. Bununla birlikte sünnet sınıfında olması gereken 3 örnek de hipospadias olarak sınıflandırarak %93,6 oranında doğru tahminleme yapılmıştır (Tablo 4.6).

Tablo 4.7 Coarsa Tree Yöntemi ile test verisi için elde edilen karışıklık matrisi ve yüzdeleri

		Tahmin Edilen Sınıf	
		Hipospadias	Sünnet
Doğru Sınıf	Hipospadias	205 (%89,5)	24 (%10,5)
	Sünnet	4 (%8,5)	43 (%91,5)

Bir diğer model olan Coarse Tree algoritmasında ise karışıklık matrisi ve yüzdeleri Tablo 4.7’de verilmiştir. Modelde, hipospadias olması gereken 24 örnek sünnet olarak sınıflandırılmış ve %89,5 doğru tahminleme yapmıştır. Bununla birlikte sünnet sınıfında olması gereken 4 örneği de hipospadias olarak sınıflandırarak %91,5 doğru tahminleme yapmıştır (Tablo 4.7).

Modellerden çıkan sonuçlara bakıldığında başarı oranı yüksek tahminlemelerin gerçekleştiği söylenebilir. Bu oran sınıflar arasında da birbirine yakın oranlarda homojen olarak dağılım göstermiştir.

Tablo 4.8 Sınıflandırma modellerinde başarı kriterleri

	Doğruluk	Duyarlılık	Özgüllük	Kesinlik
Medium Gaussian	%90,2	210/229=0,917	39/47=0,829	210/218=0,963
Optimizable Tree	%90,2	205/229=0,895	44/47=0,936	205/208=0,985
Coarsa Tree	%89,9	205/229=0,895	43/47=0,914	205/209=0,981

Doğruluk değeri en yüksek ilk üç yöntemi incelediğimizde duyarlılık oranında en yüksek değerin Medium Gaussian yönteminde olduğunu görüyoruz. Bu sonuç hipospadiaslı olan olguları doğru tespit etme oranını vermektedir. Dolayısıyla Medium Gaussian yöntemi ile hipospadias tahminlemesinde bulunan kişilerin hastalıklarının yaklaşık %92 oranında doğru tahmin edildiği görülmektedir.

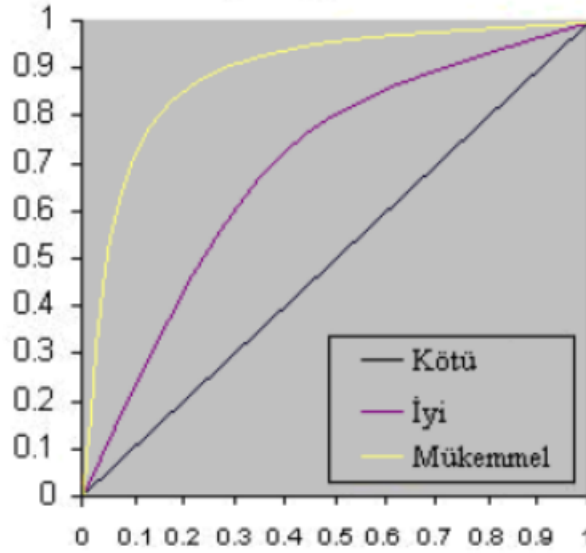
Özgüllük ise hipospadiaslı olmayan bireylerin doğru tespit edilme oranlarını vermektedir. Optimizable Tree yönteminin hipospadiaslı olmayan kişileri yaklaşık %94 oranında doğru tahminlendiği görülmüyor.

Hesaplanan kesinlik değerlerine baktığımızda yine Optimizable Tree yönteminin az farkla da olsa diğer iki yöntemden daha iyi sonuç verdiği görülmüyor. Kesinlik oranı hipospadias olarak tahminlenen hastalardan gerçekte kaçının hipospadiaslı olduğu bilgisini vermektedir. Bu bağlamda Optimizable Tree yöntemi ile hipospadiaslı olarak tahminlenen bireylerin yaklaşık olarak %99'nun gerçekte de hipospadiaslı olduğunu göstermektedir (Tablo 4.8).

Test sonuçlarının değerlendirilmesinde kullanılan doğruluk, duyarlılık, özgüllük ve kesinlik metrikleri tıp alanında tanı testi ölçütleri arasındadır. Tanı testi, bir teşhiste bulunmak amacıyla laboratuvar tekniklerine, klinik pratiklere ya da özgün gereç ölçümlerinden baz alınarak yapılan değerlendirme tekniklerini içeren genel bir kavramdır. Yukarıda konu edilen metriklerden hareketle tıpta tanı testlerinin performansının değerlendirilmesinde ROC eğrisi (Alicı İşlem Karakteristikleri, Reciever Operating Characteristic) testler arasında güvenilir bir karşılaştırma yapmaya olanak sağlamaktadır (Zweig ve Campbell, 1993).

ROC eğrisinin yerleştiği koordinat sisteminde X eksenini yanlış pozitif değeri (1- özgüllük) temsil ederken, Y eksenini ise duyarlılık değerini ifade eder. ROC eğrisi,

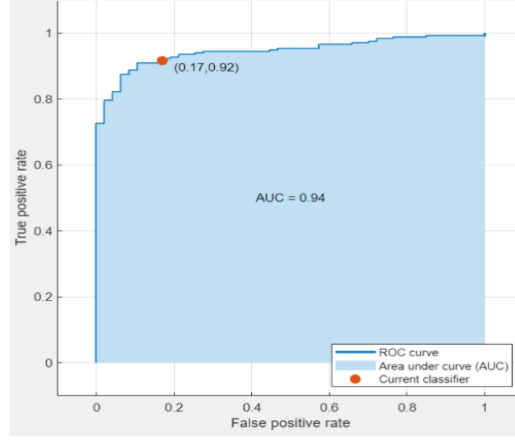
tanı koymak için kullanılan parametrenin deęişim genişlięi içinde yer aldığı bütün deęerlerinin sırasıyla X ve Y koordinat deęerlerinin kesin noktaları alınarak elde edilir (Sim vd., 2004). Şekil 4.7’de ideal ve performanslarına göre ROC eęrileri gösterilmiştir.



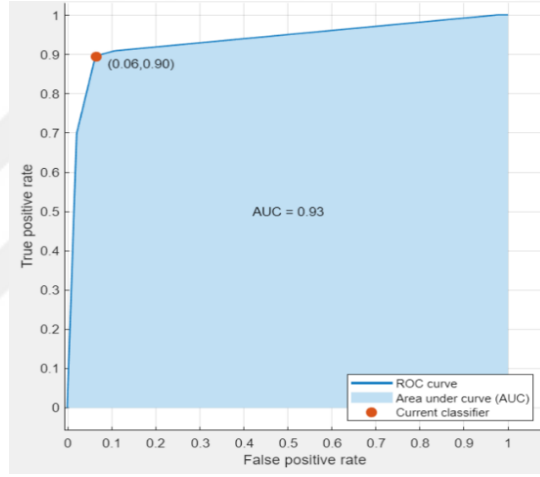
Şekil 4.7 Performanslarına göre ROC eęrileri (Koç, 2013)

Hanley ve Mcneil’in 1982 yılında yaptığı çalışmada ROC eęrisinin altında kalan alanın (AUC) tanı testlerinin performans deęerlendirmesinde kullanılabileceęi gösterilmiştir (Hanley ve Mcneil, 1982). ROC eęrisinin yapısı gereęi, koordinat deęerleri 0’dan 1’e kadar olan deęerleri alabilirler. Hatasız bir tahminlemede eęrinin altında kalan alan 1 iken, AUC deęeri 0,5 olduęunda ise tanı testinin tahminleme performansının olmadığı anlaşılr. AUC deęeri 1’e yaklaştıkça performansın mükemmele yaklaştığı sonucuna varılır. Elde edilen AUC deęerleri için aşıęıda verilen derecelendirme kullanılabilir (Obuchowski, 1994).

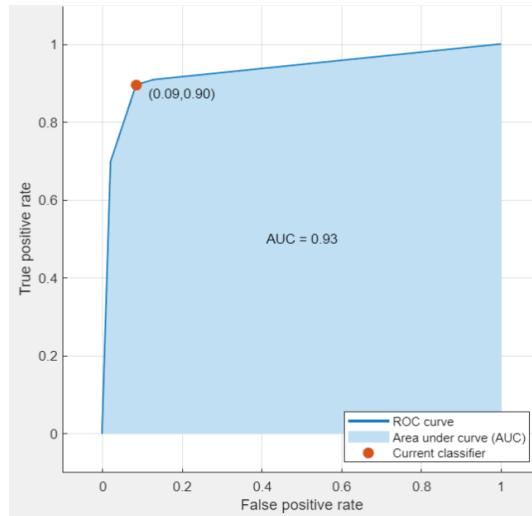
- 0,90-1.00 = Mükemmel
- 0,80- 0,90 = İyi
- 0,70-0,80 =Orta
- 0,60-0,70 = Zayıf
- 0,50-0,60 = Başarısız



Şekil 4.8 Medium Gaussian *SVM* Yöntemi için ROC eğrisi ve AUC değeri



Şekil 4.9 Optimizable Tree Yöntemi için ROC eğrisi ve AUC değeri



Şekil 4.10 Coarsa Tree Yöntemi için ROC eğrisi ve AUC değeri

Şekil 4.8, 4.9 ve 4.10'da doğruluk değeri en yüksek ilk üç modelin ROC eğrisi ve AUC değerleri gösterilmiştir. Medium Gaussian SVM yönteminde 0,94 Optimizable Tree yönteminde 0,93 Coarse Tree yönteminde ise yine 0,93 değerleri elde edilmiştir. AUC değer derecelendirmeleri baz alındığında model tahminlemelerinin mükemmel denilebilecek performanslara sahip olduğu sonucuna varılır.



Bu çalışma kapsamında hemogram parametreleri kullanılarak hipospadias olgularında inflamatuvar süreç ortaya koyulmaya çalışılmıştır. Ortaya çıkan sonuçların kıyaslanabilmesi için sağlıklı verileri temsilen sünnet operasyonu talebiyle gelmiş olan örnekler ele alınmıştır.

Yapılan literatür çalışması göz önüne alındığında Türkiye'nin çeşitli bölgelerinde ve genellikle ilköğretim okullarındaki erkek öğrencilerden hareketle hipospadias oranları tespit etmeye yönelik çalışmalara rastlanmıştır. Yine literatürde bulunan, dünyada belirli ülkelerdeki çalışmalar göz önüne alındığında, daha kapsamlı bir örneklem üzerinden elde edilebilecek sonuçların farkındalık oluşturabilmesi açısından önemli olduğu düşünülmektedir. Türkiye'de askerlik görevini yapmakta olan genç erkekler üzerindeki araştırmada, hipospadias olgusuna rastlanmış bireylerin farkındalık oranının düşük olduğu görülmüştür. Hipospadias sebebiyle operasyon geçirmiş bireylerin ise operasyon yaş ortalamasının fizyolojik ve psikolojik açıdan bakıldığında riskli bir değerde olduğu saptanmıştır.

Hipospadiaslı hastaların sünnet olmasının problemlere yol açacağı göz önünde bulundurulduğunda; bu çalışma veri madenciliğinde tespit edilen modeller yardımıyla sünnet talebiyle gelmiş hastaların karşılaşılabilecekleri olası risklerin önüne geçilmesine katkı sağlayabilir.

Çalışmada 21 çeşit hemogram parametresinin sonucu öznitelik olarak kullanılmıştır. Her bir öznitelik için ayrı ayrı istatistiksel analiz yapıldığında bunlardan bazofil sayısı, bazofil yüzdesi, hemoglobin, ortalama eritrosit hemoglobin, eritrosit hemoglobin konsantrasyonu, eritrosit hücresi ortalama hacmi, monosit sayısı, monosit yüzdesi, ortalama trombosit hacmi, trombosit dağılım genişliği, trombosit sayısı, eritrosit sayısı, eritrosit dağılım genişliği olmak üzere 13 çeşit hemogram parametresinin hipospadiaslı olgularla sağlıklı olan kişiler arasında anlamlı fark içerdiği görülmüştür. 8 adet hemogram parametresinde ise iki sınıf arasında anlamlı bir fark gözlenmemiştir. Veri

madenciliğinde kullanılan yöntemlerde ise, hemogram test çeşitlerinin tamamı birlikte değerlendirilerek yüksek oranda başarı içeren tahminleme sonuçları elde edilmiştir. Modellerden çıkan sonuçlar karşılaştırıldığında, Medium Gaussian yönteminin %90,2 Optimizable Tree algoritmasının %90,2 ve Coarse Tree algoritmasının %89,9 ile en yüksek başarı oranlarıyla tahminleme yaptığı görülmüştür. Bu sonuç veri madenciliği yöntemlerinin daha kapsamlı bir şekilde verileri değerlendirerek yüksek seviyede tahminleme olanağı tanıdığını gösterebilir.

Literatürde makine öğrenmesi sonuçlarının kıyaslanabileceği benzer bir çalışmaya rastlanamadığından, çıkan sonuçları yorumlamak için ROC analizinde bulunan derecelendirmeler kullanılmış (Obuchowski, 1994) ve sonuçların en başarılı kategoride bulunduğu tespit edilmiştir.

Veri madenciliğinde kullanılan özniteliklerin artırılması tahminleme başarı oranının yükselmesine imkân verebilir. Ayrıca inflamatuvar süreçle yeni özniteliklerin ilişkileri de gözlenmiş olur. Yine bu duruma paralel olarak kullanılan örnek sayılarının genişletilmesi de model başarı oranlarının artmasını sağlayacaktır.

- Abdullah, N. A., Pearce, M. S., Parker, L., Wilkinson, J. R., Jaffray, B., ve McNally, R. J. (2007). Birth prevalence of cryptorchidism and hypospadias in northern England, 1993–2000. *Archives of disease in childhood*, 92(7), 576-579.
- Adriaans, P. ve Zantinge, D. (1997). *Data Mining*, Boston, MA, USA Addison Wesley Longman Publishing.
- Akay, A. F., Şahin, H., Em, S., Kuru, A. F., Ayçiçek, S., ve Bircan, M. K. (2002). Diyarbakır il merkezindeki ilkököl çağı erkek çocuklarında dış genital organ anomali oranları. *Türk Üroloji Dergisi/Turkish Journal of Urology*, 28(1), 76-79.
- Akın, Y. K. (2008). Veri madenciliğinde kümeleme algoritmaları ve kümeleme analizi. Doktora Tezi, Marmara Üniversitesi, Türkiye.
- Amar, E., Michon, L., ve Gnansia, E. (2018). Hypospadias and environment: A registry-based case-control study. *Revue d'Épidémiologie et de Santé Publique*, 66, S333-S334.
- Anand, S., Lotfollahzadeh S. (2022). *Epispadias*, StatPearls Publishing.
- Atar, E., Aşkın, A. (2017). Diz osteoartrit hastalarında nötrofil/lenfosit oranı, trombosit/lenfosit oranı ve ortalama trombosit hacminin değerlendirilmesi. *Cukurova Medical Journal*, 42(2), 329-336.
- Baskin, L. S. (2000). Hypospadias and urethral development. *The Journal of Urology*, 163(3), 951-956.
- Baskin, L. S., ve Ebberts, M. B. (2006). Hypospadias: anatomy, etiology, and technique. *Journal of Pediatric Surgery*, 41(3), 463-472.
- Baykal, A. (2006). Veri madenciliği uygulama alanları. Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, (7), 95-107.
- Baykal, N. (2003). Veri tabanı ve veri madenciliği adlı sunum. Tıp Bilişimi Güz Okulu Ders Notları.
- Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 403.
- Carmichael, S. L., Shaw, G. M., Nelson, V., Selvin, S., Torfs, C. P., ve Curry, C. J. (2003). Hypospadias in California: trends and descriptive epidemiology. *Epidemiology*, 14(6), 701-706.
- Çarklı, B. (2010). sağlık sektöründe bir primoriliğisi ile bir veri madenciliği uygulaması, Yüksek Lisans Tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.
- Çelik, S. (2020). determination and classification of importance of attributes used in diagnosing pregnant women's birth method. *Alphanumeric Journal*, 8(2), 261-274.

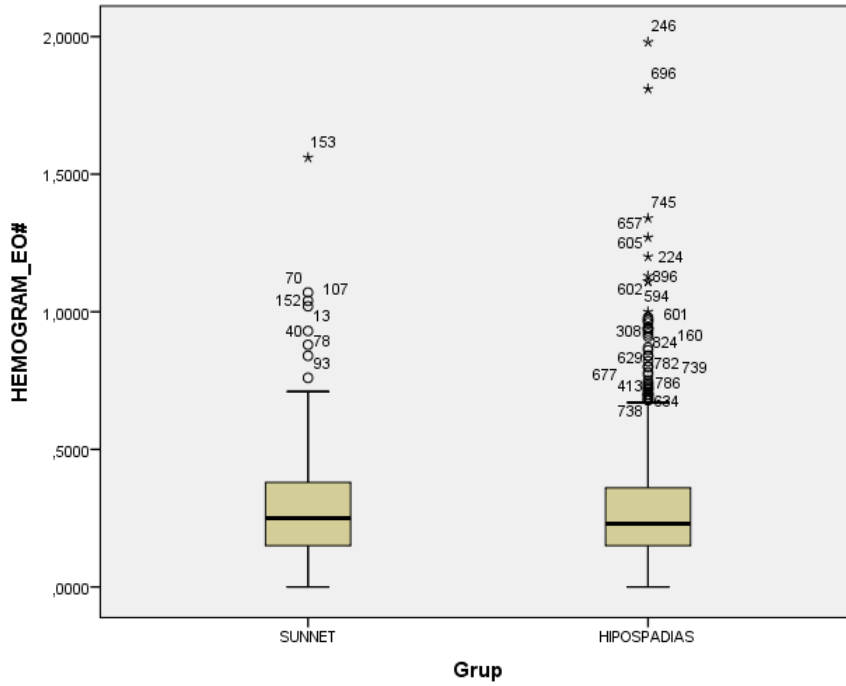
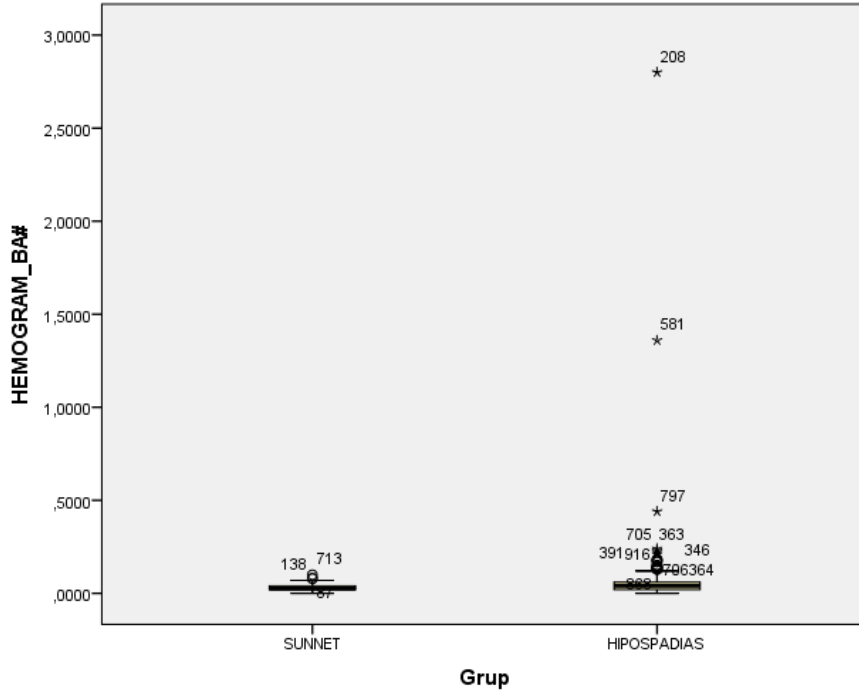
- Danacı, M., Çelik, M., ve Akkaya, A. E. (2010). Veri madenciliği yöntemleri kullanılarak meme kanseri hücrelerinin tahmin ve teşhisi. Akıllı sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU'2010), 21-24.
- Deliktaş, H., Tetiker, H., Çetinkaya, M., Şahin, H. ve Uğuz, C. (2014). Muğla ili ilköğretim okullarında erkek öğrencilerde dış genital organ anomali prevalansı. Muğla Sıtkı Koçman Üniversitesi Tıp Dergisi, 1 (1), 21-24.
- Demirel, Ş., ve YAKUT, S. G. (2019). Karar ağacı algoritmaları ve çocuk işçiliği üzerine bir uygulama. Sosyal Bilimler Araştırma Dergisi, 8(4), 52-65.
- Doğan, S., ve Türkoğlu, İ. (2008). Iron-deficiency anemia detection from hematology parameters by using decision trees. International Journal of Science ve Technology, Cilt 3, No 1, 85-92.
- Ekmark, A., Grelaud, D., Hansson, E., Svensson, H., Arnbjörnsson, E., Gisselsson, D. (2020). The cellular architectures of hypospadias. Pediatric and Developmental Pathology, 23(6), 476-478.
- Elmoghazy, H., Saber, M., Mamdoh, A., Rashed, E., Badawy, A. A., ve Riayd, A. M. (2020). Cosmetic and functional outcome for the use of stitch by stitch technique in hypospadias surgery; results of 235 patients. Journal of Pediatric Urology, 16(4), 436-e1.
- Emel, G. G., Taşkın Ç. (2005). Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 6(2), 221-239.
- Emel, G. G., Taşkın, Ç., ve Tok, A. (2005). Pazarlama stratejilerinin oluşturulmasında bir karar destek aracı: birliktelik kuralı madenciliği.
- Eming, S. A., Krieg, T., ve Davidson, J. M. (2007). Inflammation in wound repair: molecular and cellular mechanisms. Journal of Investigative Dermatology, 127(3), 514-525.
- Ergezer, H., Dikmen, M., ve Özdemir, E. (2003). Yapay sinir ağları ve tanıma sistemleri. PIVOLKA, 2(6), 14-17.
- Fernandez N., Armando L., Rickard M., Jaime P., Braga L., Matava C. (2020). Digital pattern recognition for the identification and classification of hypospadias using artificial intelligence vs experienced pediatric urologist, Journal of Urology, Volume 147, Issue 1.
- Hanley, J. A., ve McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.
- He, H., ve Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.
- İnan, O. (2003). Veri madenciliği, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Jacobs, P. (1999). Data mining: what general managers need to know. Harvard Management Update, 4(10), 8.

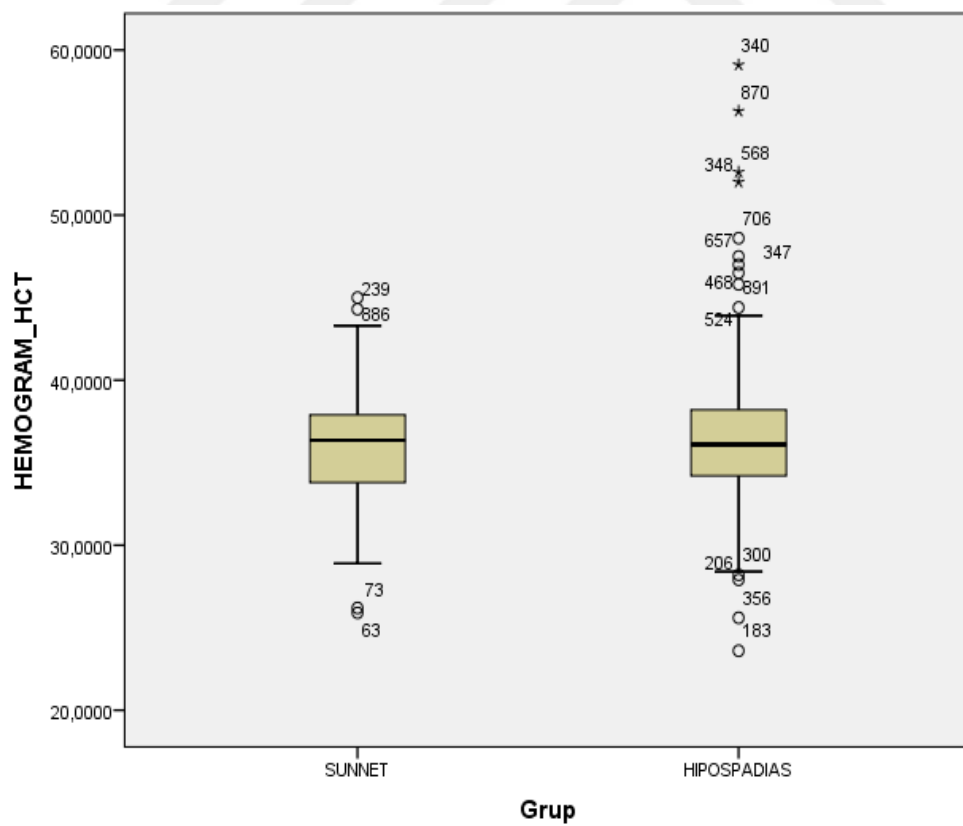
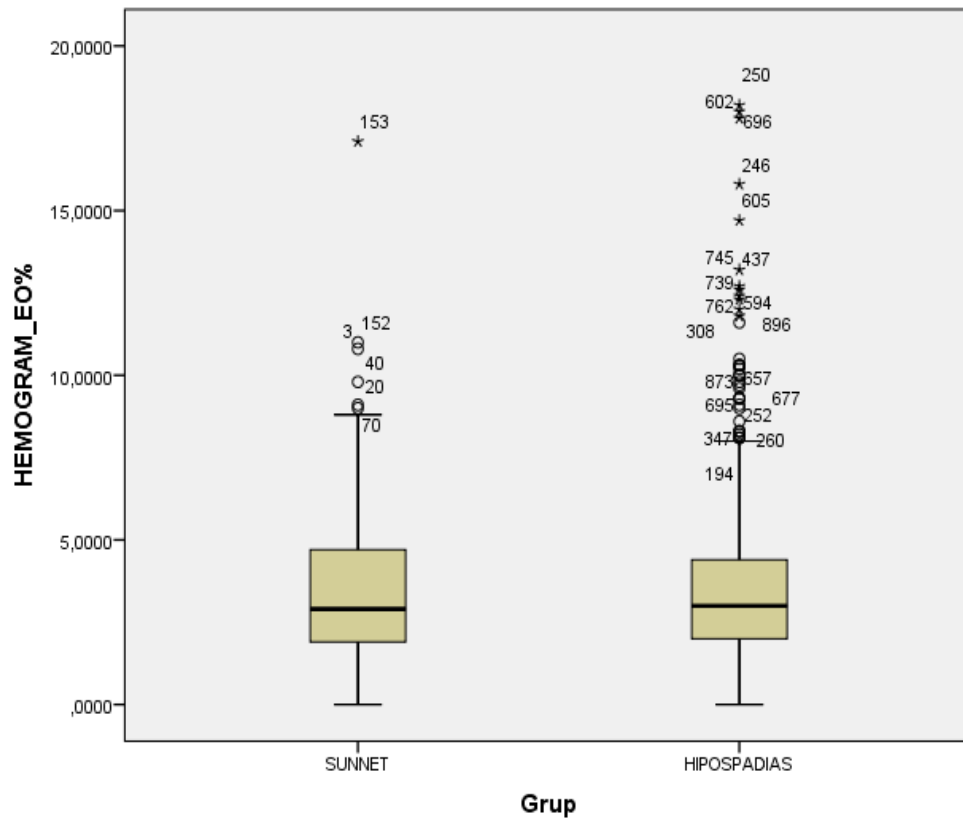
- Jones, B. C., O'Brien, M., Chase, J., Southwell, B. R., ve Hutson, J. M. (2009). Early hypospadias surgery may lead to a better long-term psychosexual outcome. *The Journal of Urology*, 182(4), 1744-1750.
- Kaçtıoğlu, S., Özen, Ü ve Yavuz U. (1999). *Bilgisayara Giriş ve Güncel Yazılımlar-1*, Aktif Yayınevi, Erzurum.
- Kavzoğlu, T., ve Çölkesen, İ. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. *Harita Dergisi*, 144(7), 73-82.
- Kaya, M. E., Payza, U., Bilgin S., Kayalı A., Yamanoglu A., Karakaya Z. (2021). hemogram parametrelerin, inflamatuvar barsak hastalığı olan hastalarda hastalık şiddeti, mortalite ve komplikasyonları öngörmedeki rolü. *Dokuz Eylül Üniversitesi Tıp Fakültesi Dergisi*, 35(2), 147-157.
- Kayıkçı M., Çam K., Akman R., ve Erol A. (2005). Düzce ilinde ilköğretim çağındaki erkek çocuklarda dış genital organ anomali oranları. *Türk Üroloji Dergisi/Turkish Journal Of Urology*, 31(1), 79- 81.
- Keays, M. A., ve Dave, S. (2017). Current hypospadias management: Diagnosis, surgical management, and long-term patient-centred outcomes. *Canadian Urological Association Journal*, 11(1-2Suppl1), S48.
- Kimura N., Mizokami A., Oonuma T., Sasano H. ve Nagura H. (1993). Immunocytochemical localization of androgen receptor with polyclonal antibody in paraffin-embedded human tissues. *Journal of Histochemistry and Cytochemistry*, 41:671- 678.
- Koç, A., Elaltuntaş, E. ve Ötünçtemur, A. (2011). Cizre’de 7-14 yaş arası erkek çocuklarda genital anomali oranları. *The New Journal Of Urology*, 6 (1), 20-23.
- Koç, E. (2013). Yöntem ve uygulama açısından klinik karar destek sistemleri. Yüksek Lisans Tezi, Okan Üniversitesi, Fen Bilimleri Enstitüsü
- Koyuncugil, A., ve Özgülbaş, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *Bilişim Teknolojileri Dergisi*, 2:21-32.
- Lane, C., Boxall, J., MacLellan, D., Anderson, P. A., Dodds, L., & Romao, R. L. (2017). A population-based study of prevalence trends and geospatial analysis of hypospadias and cryptorchidism compared with non-endocrine mediated congenital anomalies. *Journal of Pediatric Urology*, 13(3), 284-e1.
- Lund, L., Engebjerg, M. C., Pedersen, L., Ehrenstein, V., Nørgaard, M., ve Sørensen, H. T. (2009). Prevalence of hypospadias in Danish boys: a longitudinal study, 1977–2005. *European Urology*, 55(5), 1022-1026.
- Manzoni, G., Bracka, A., Palminteri, E., ve Marrocco, G. (2004). Hypospadias surgery: when, what and by whom?. *BJU international*, 94(8), 1188-1195.
- Mesci Y. (2020). <https://medium.com/deep-learningturkiye/k%C3%BCmeleme-teknikleri-ve-pythonda-k-means-62d93c624e5c>, 10.06.2022.

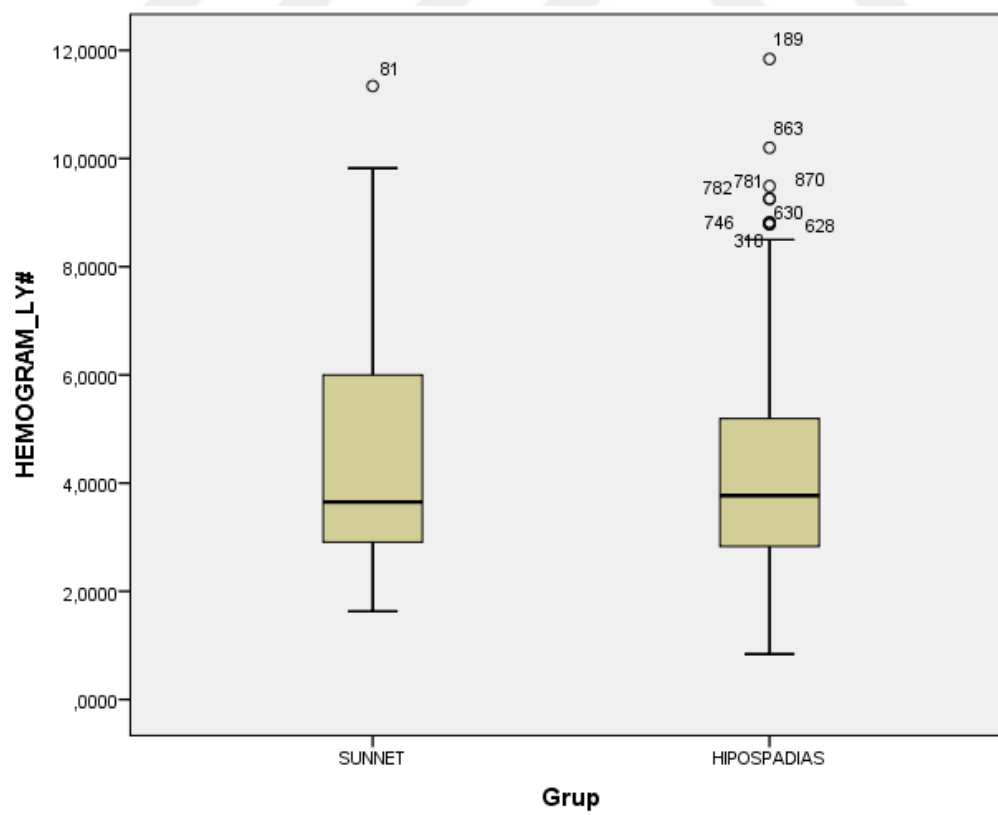
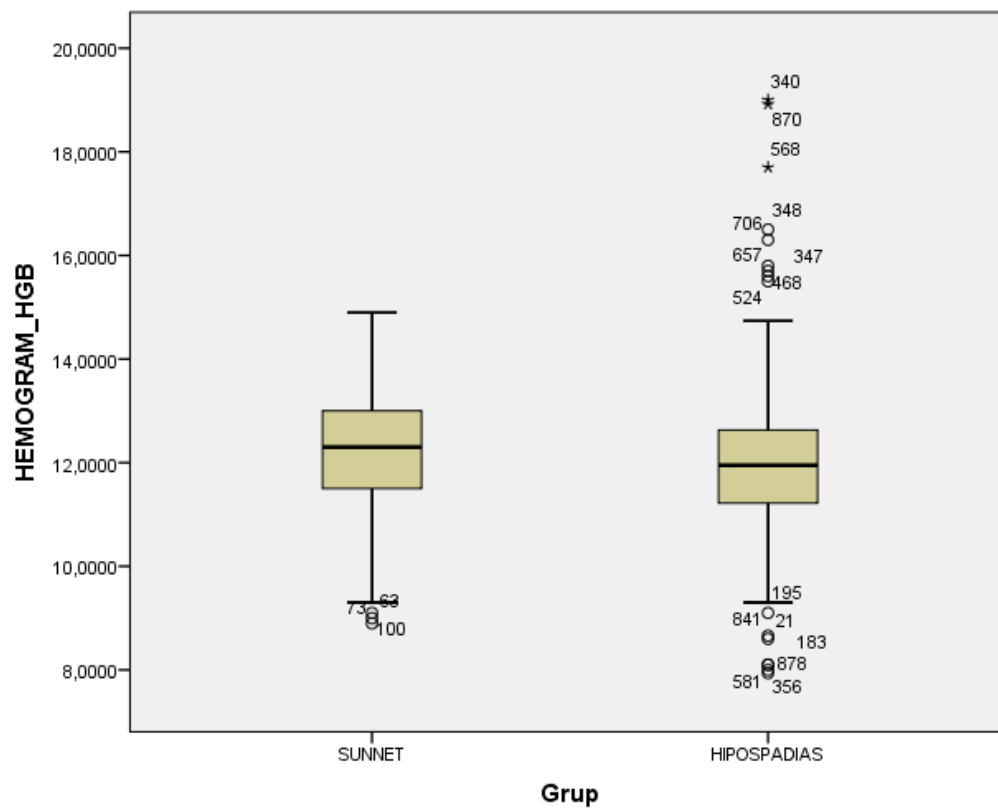
- Obuchowski, N. A. (1994). Computing sample size for receiver operating characteristic studies. *Investigative Radiology*, 29(2), 238-243.
- Ortqvist, L., Engberg, H., Strandqvist, A., Nordenström, A., Holmdahl, G., Nordenskjöld, A., ve Frisén, L. (2019). Psychiatric symptoms in men with hypospadias—preliminary results of a cross-sectional cohort study. *Acta Paediatrica*, 108(6), 1156-1162.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Dergisi*. 65-82.
- Piatetsky-Shapiro, G. (1990). Knowledge discovery in real databases: A report on the IJCAI-89 Workshop. *AI Magazine*, 11(4), 68-68.
- Porter, M. P., Faizan, M. K., Grady, R. W., ve Mueller, B. A. (2005). Hypospadias in Washington State: maternal risk factors and prevalence trends. *Pediatrics*, 115(4), e495-e499.
- Riedmiller, H., Androulakakis, P., Beurton, D., Kocvara, R., ve Gerharz, E. (2001). EAU Guidelines on Paediatric Urology1. *European urology*, 40(5), 589-599.
- Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- Sikka, S. C., ve Wang, R. (2008). Endocrine disruptors and estrogenic effects on male reproductive axis. *Asian Journal of Andrology*, 10(1), 134-145.
- Silahtaroglu, G. (2008). Kavram ve algoritmalarıyla temel veri madenciliği, Papatya Yayıncılık Eğitim AŞ. İstanbul, Türkiye.
- Sim, L., Manthey, K., Esdaile, P., ve Benson, M. (2004). Comparison of computer display monitors for computed radiography diagnostic application in a radiology PACS. *Australasian Physics ve Engineering Sciences in Medicine*, 27(3), 148-150.
- Snyder, C. L., Evangelidis, A., Hansen, G., Peter, S. D. S., Ostlie, D. J., Gatti, J. M., ... ve Murphy, J. P. (2005). Management of complications after hypospadias repair. *Urology*, 65(4), 782-785.
- Söylemez, H., Atar, M., Bozkurt, Y., Penbegül, N., ve Ali Sancaktutar, A. (2011). Askerlik aşındaki erkeklerde inguinal ve genital organ anomalisi sıklığı ve farkındalığı. *Türk Üroloji Dergisi/Turkish Journal Of Urology*, 37(2), 153-158.
- Stokowski, L. A. (2004). Hypospadias in the neonate. *Advances in Neonatal care*, 4(4), 206-215.
- Şeker, S. E. (2018). CRISP-DM: Endüstriler Arası Standart İşleme-Veri Madenciliği için (Cross Industry Standard Processing-Data Mining). *YBS Ansiklopedi*, 5(2).
- Tekgöl S., Türkeri L., Esen A., ve Alıcı B. (2016). Üroloji Masaüstü Başvuru Kitabı. s. 815-830, İris Yayınevi.

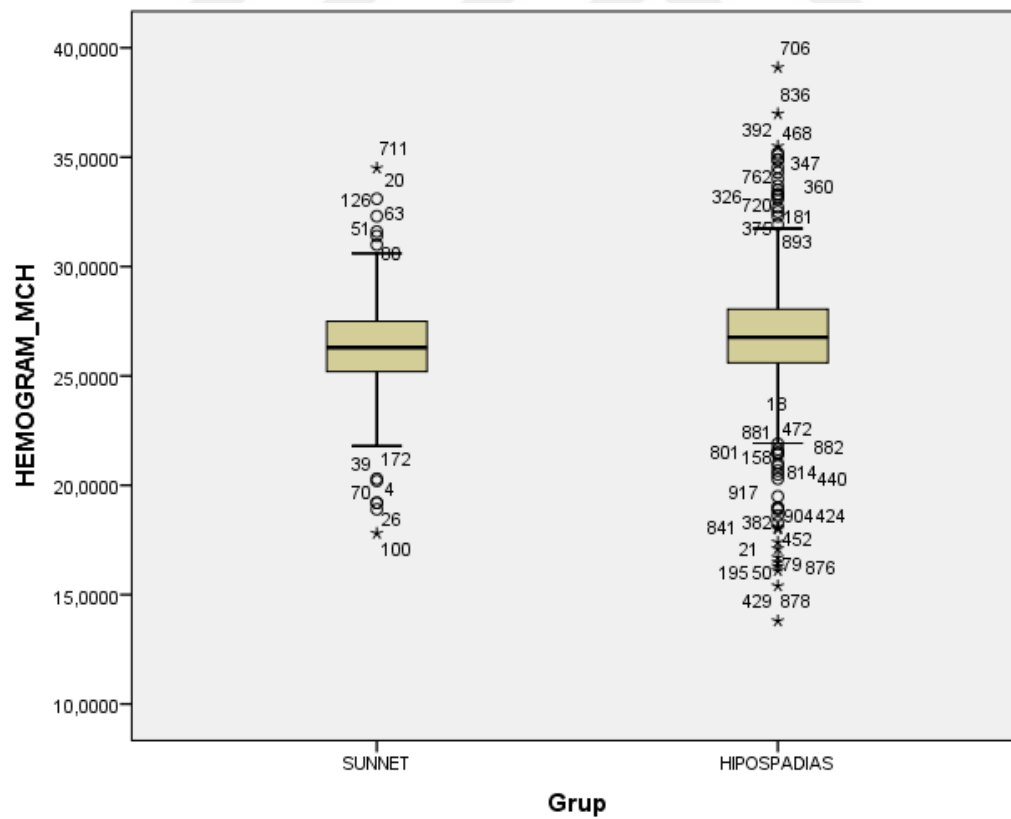
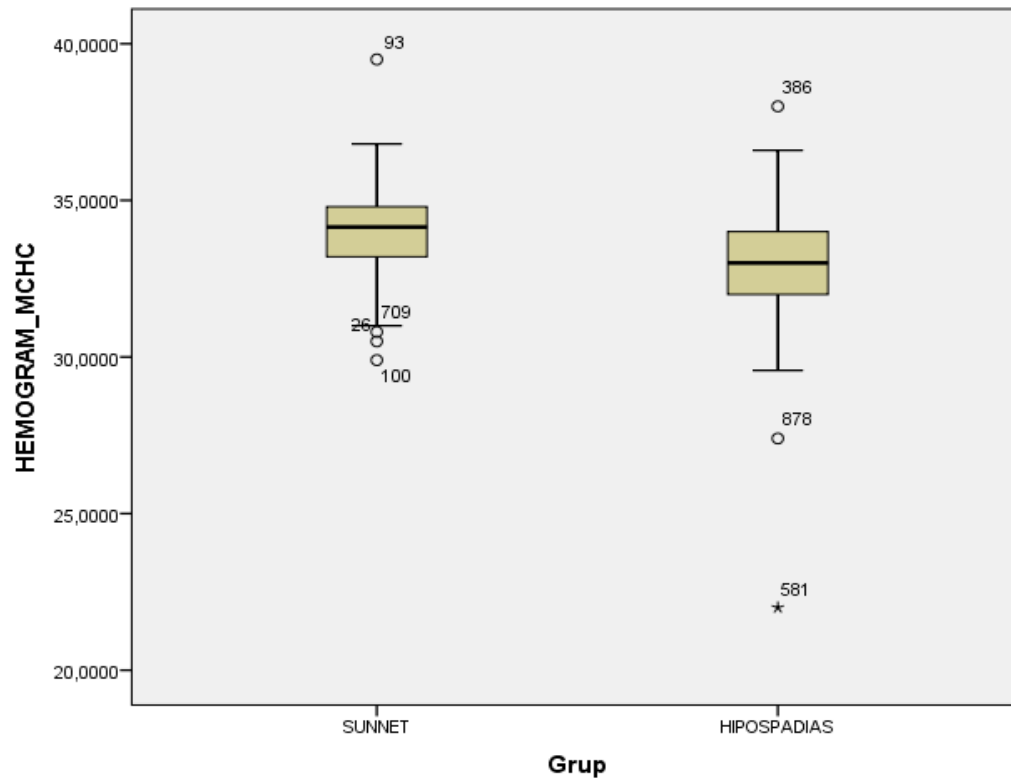
- Thankamony, A., Pasterski, V., Ong, K. K., Acerini, C. L., ve Hughes, I. A. (2016). Anogenital distance as a marker of androgen exposure in humans. *Andrology*, 4(4), 616-625.
- Van Der Horst, H. J. R., ve De Wall, L. L. (2017). Hypospadias, all there is to know. *European Journal of Pediatrics*, 176(4), 435-441.
- Van Der Zanden, L. F. M., Van Rooij, I. A. L. M., Feitz, W. F. J., Franke, B., Knoers, N. V. A. M., ve Roeleveld, N. (2012). Aetiology of hypospadias: a systematic review of genes and environment. *Human reproduction update*, 18(3), 260-283.
- Varol, S., Özden, E., Bostanci, Y., Yakupoğlu, Y. K., Yalman, C., ve Sarıkaya, Ş. (2011). Samsun ilinde ilköğretim çağı erkek çocuklarda testis ve penis boyutları ve dış genital organ anomali oranları. *Türk Pediatri Arşivi*, 46(2), 59-63.
- Yıldırım, M. E., Yanaral, F., ve Akçin, S. (2013). Siirt ilinde sünnet yapılan çocuklarda genital anomali oranları, penis boyu ve testis hacimleri. *Yeni Üroloji Dergisi*, 8(2), 6-9.
- Yıldırım, P., Uludağ, M., ve Görür, A. (2008). Hastane bilgi sistemlerinde veri madenciliği. *Çanakkale On Sekiz Mart Üniversitesi Akademik Bilişim 2008 Konferansı*, Çanakkale.
- Zweig, M. H., ve Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4), 561-577.

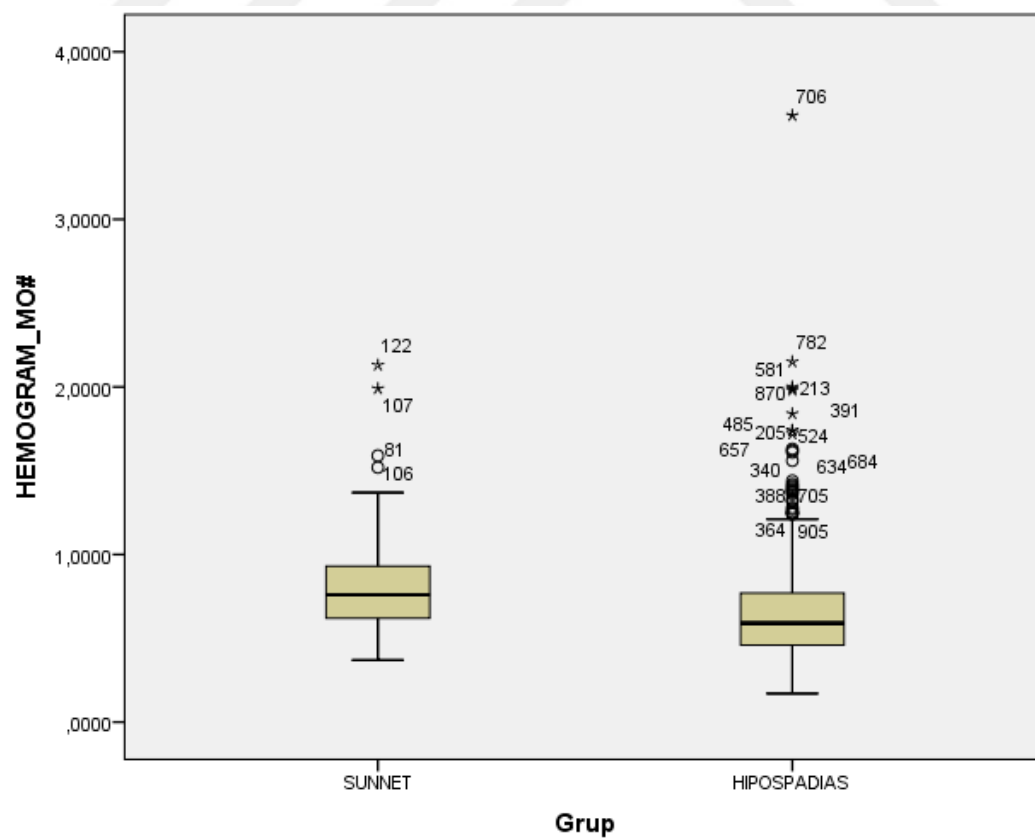
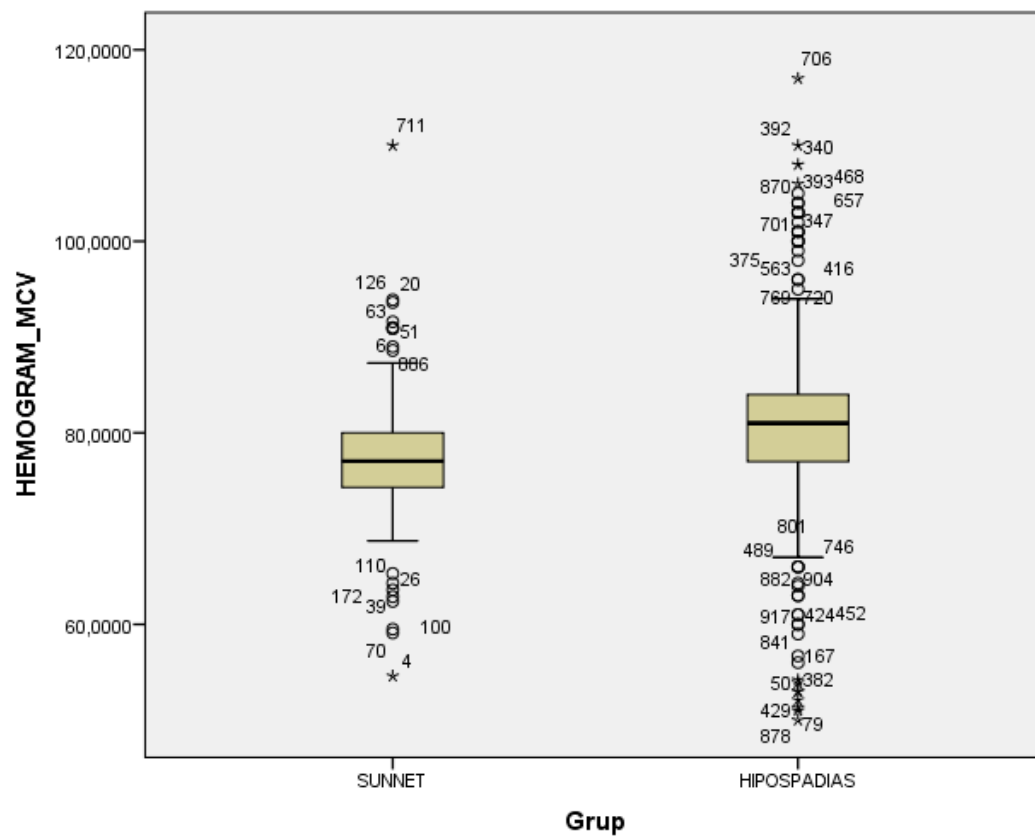
ÖZNİTELİKLERİN KUTU DİYAGRAMALARI

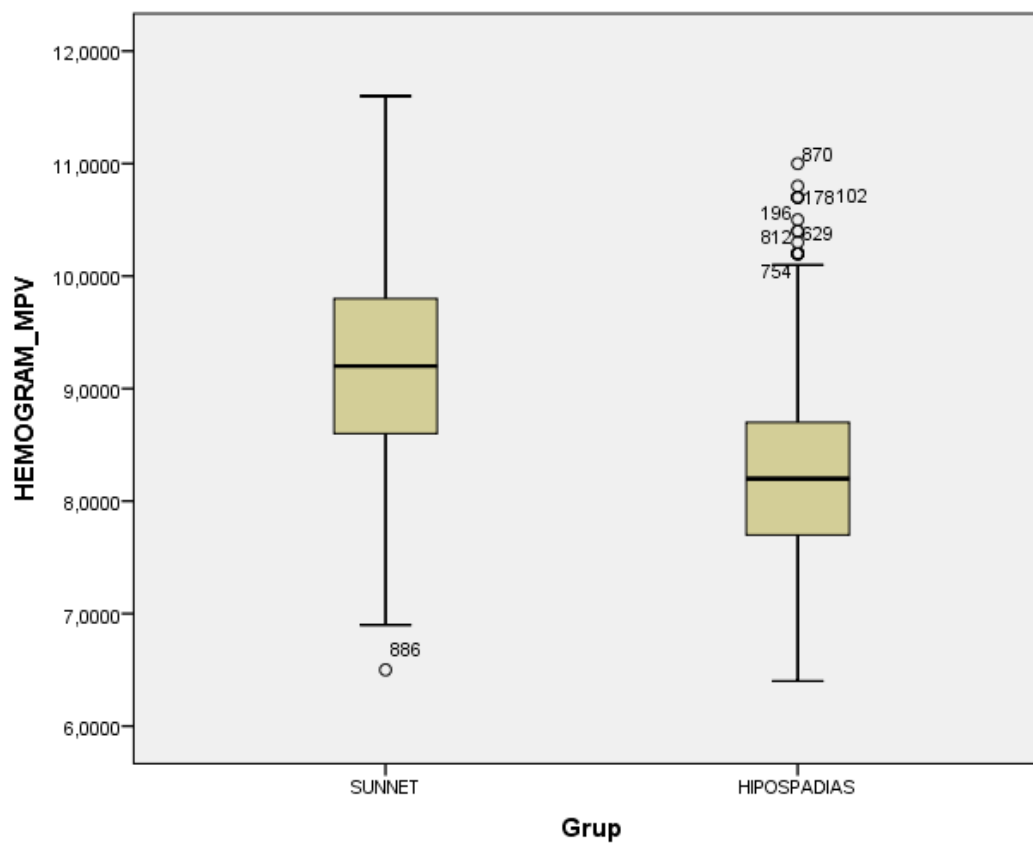
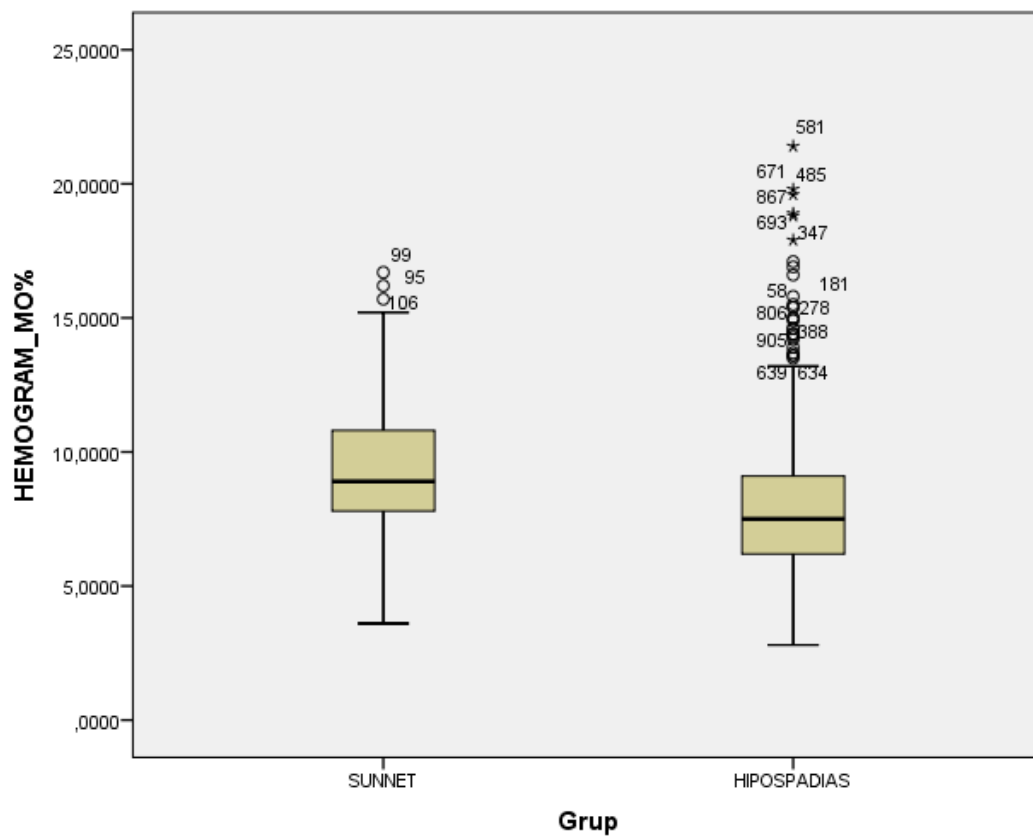


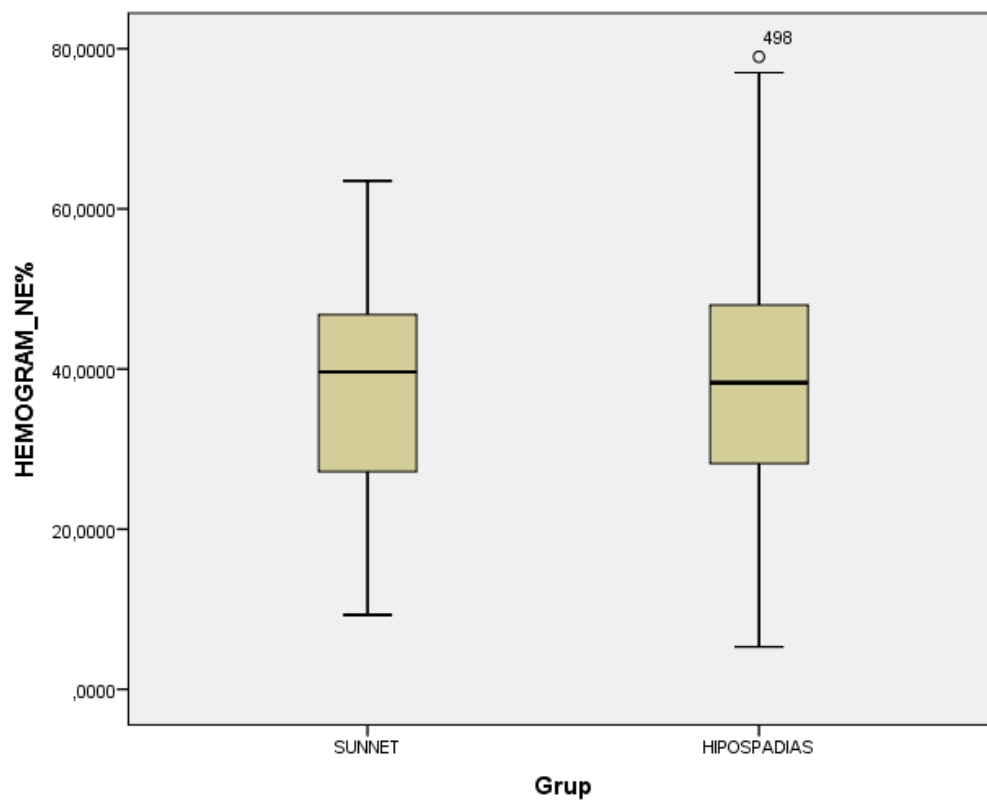
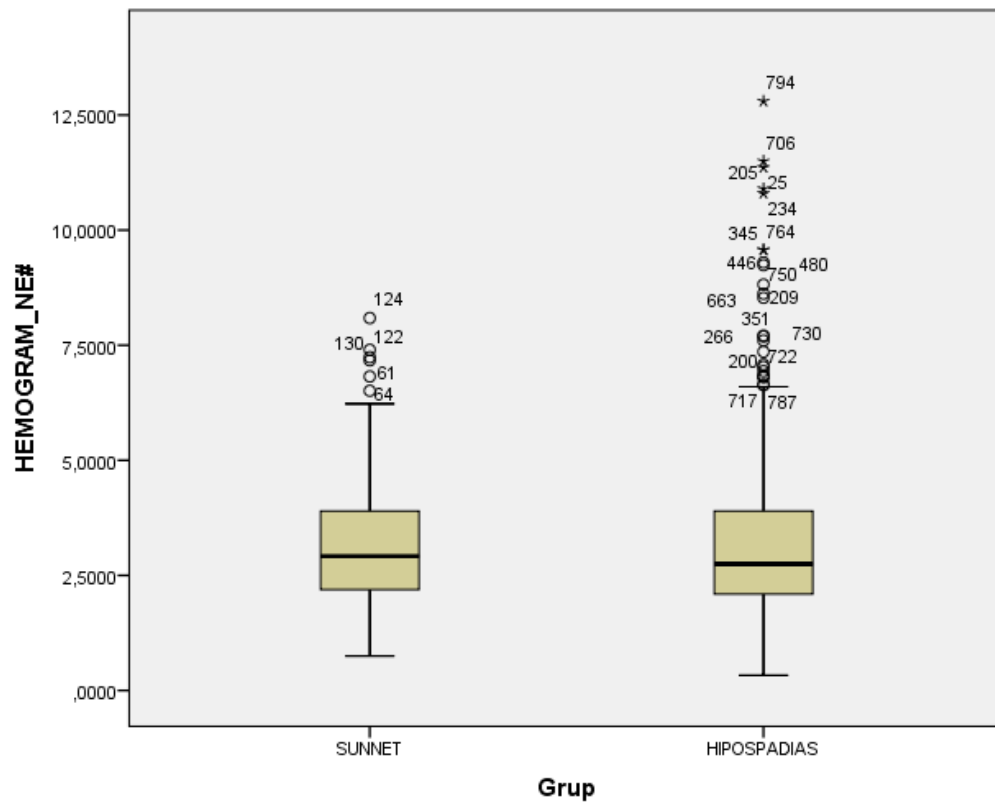


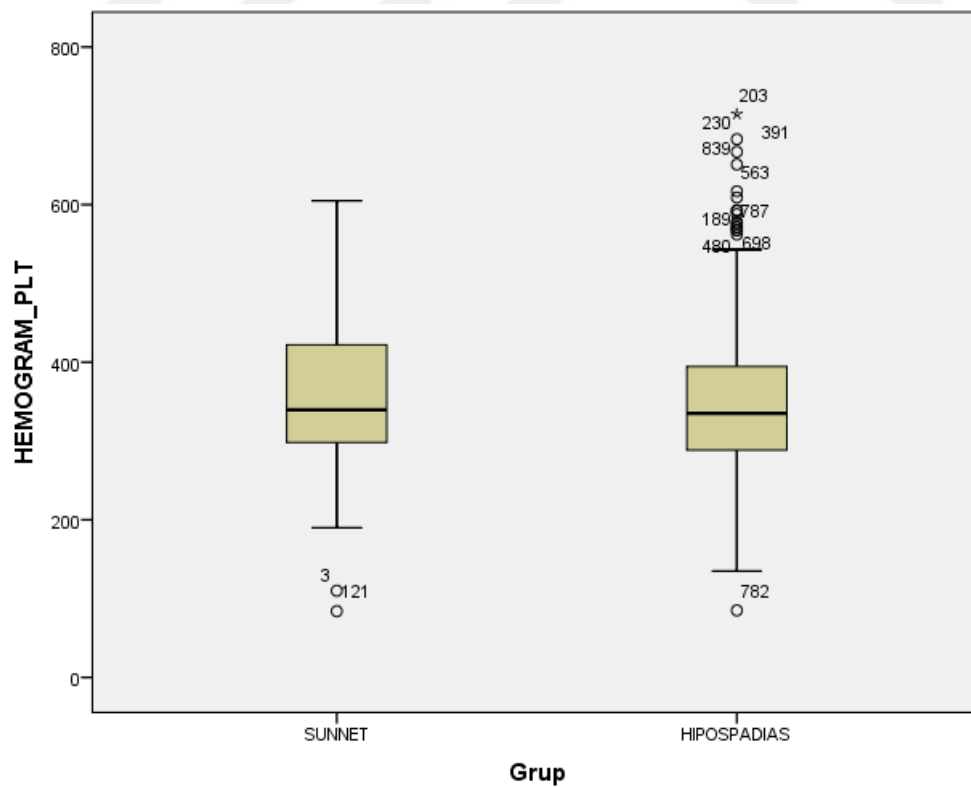
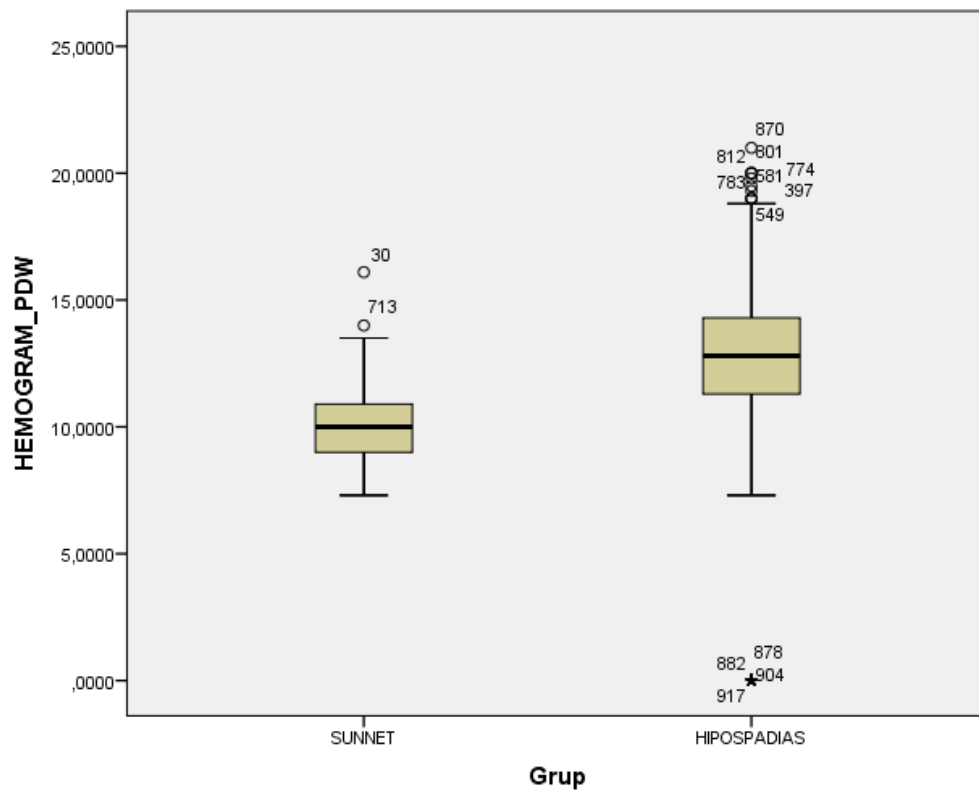


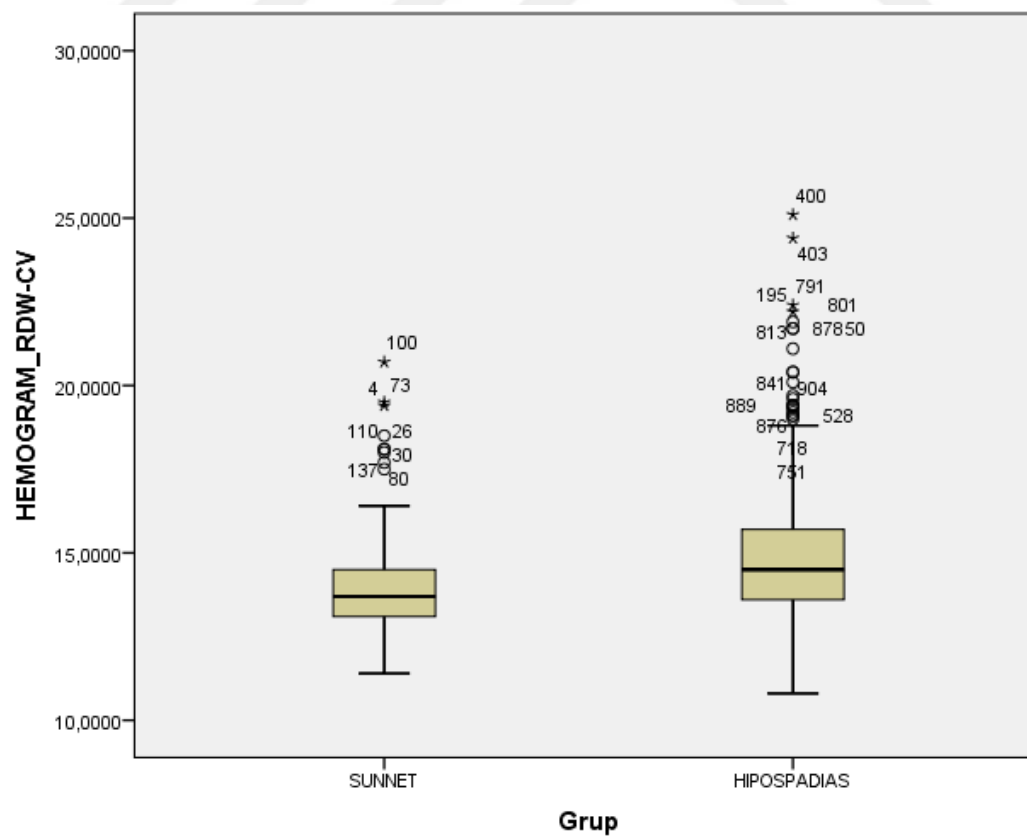
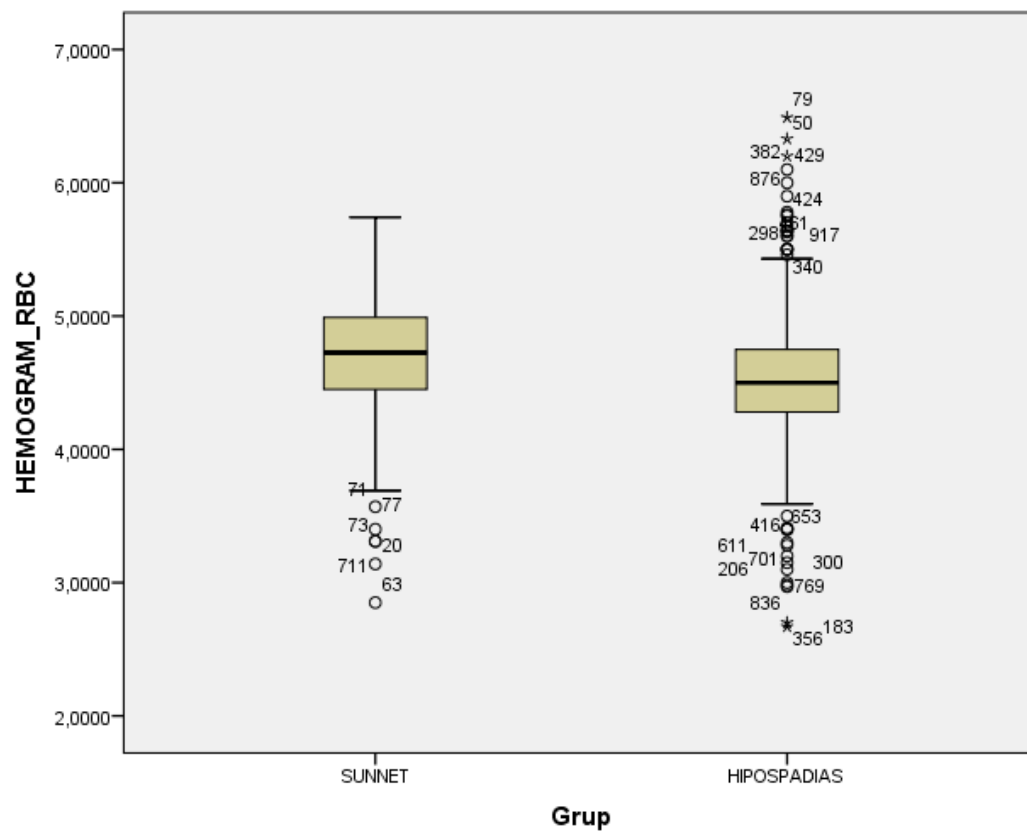


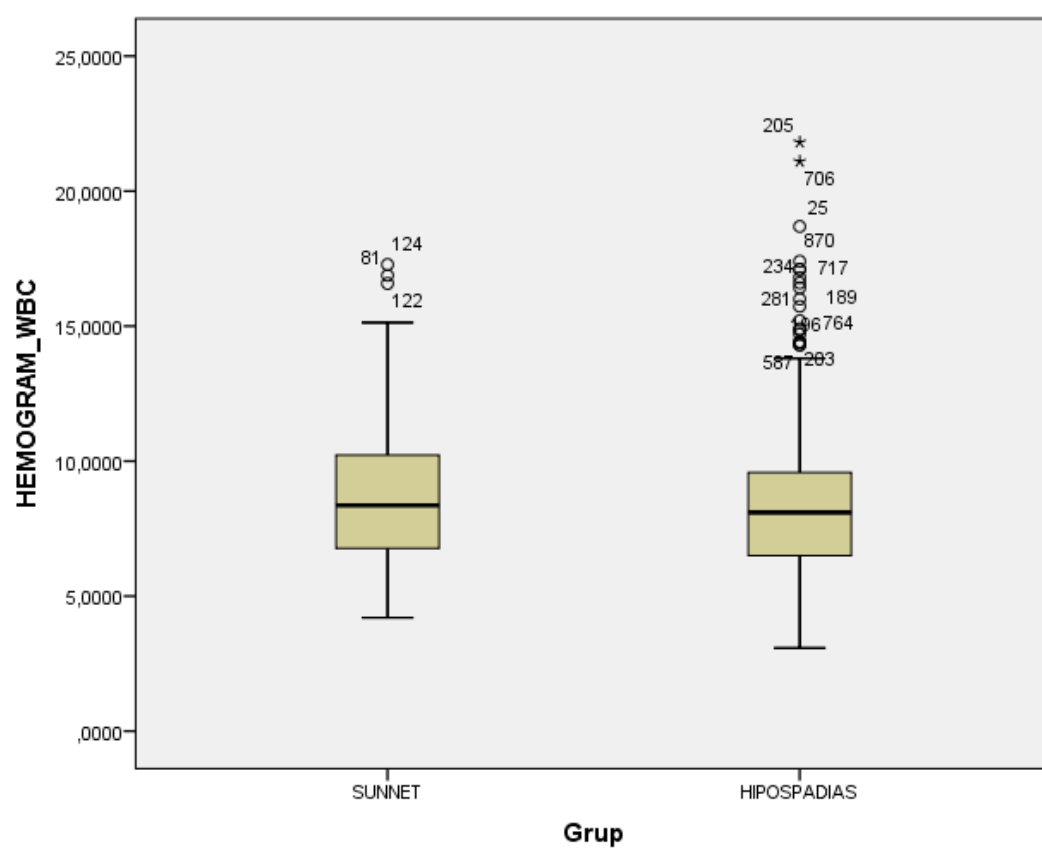












Konferans Bildirileri

1. Acar H., Aslanyürek B., Aydın E., (2022). Data Mining Applications Based On The Laboratory Results Of Kids Who Diagnosed With Hypospadias Or With Chordee Or Had Circumcision Operation, III. International Science and Innovation Congress 2022 (INSI 2022), (9-12 Haziran 2022).

