

**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**SAĞLIK BİLİMLERİ ENSTİTÜSÜ**  
**BİYOİSTATİSTİK VE TIBBİ BİLİŞİM ANABİLİM DALI**

**AMİYOTROFİK LATERAL SKLEROZ (ALS)**  
**HASTALIĞININ GENETİK VE KLİNİK VERİ**  
**İLİŞKİSİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE**  
**İNCELENMESİ**

Nesrin ÇELİK

YÜKSEK LİSANS TEZİ

2018-ANTALYA

**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**SAĞLIK BİLİMLERİ ENSTİTÜSÜ**  
**BİYOİSTATİSTİK VE TIBBİ BİLİŞİM ANABİLİM DALI**

**AMİYOTROFİK LATERAL SKLEROZ (ALS)**  
**HASTALIĞININ GENETİK VE KLİNİK VERİ**  
**İLİŞKİSİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE**  
**İNCELENMESİ**

Nesrin ÇELİK

YÜKSEK LİSANS TEZİ

**DANIŞMAN**  
**Doç. Dr. Uğur BİLGE**

“Kaynakça gösterilerek tezimden yararlanılabilir”

2017-ANTALYA

**Sağlık Bilimleri Enstitüsü Müdürlüğüne;**

Bu çalışma jürimiz tarafından Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı Tıp Bilişimi Programında Yüksek Lisans tezi olarak kabul edilmiştir. 30/03/2018

İmza

Tez Danışmanı : Doç. Dr. Uğur BİLGE  
Akdeniz Üniversitesi



Üye : Prof. Dr. Ahmet YARDIMCI  
Akdeniz Üniversitesi



Üye : Yrd. Doç. Dr. Melike ŞİŞECİ ÇEŞMELİ  
Mehmet Akif Ersoy Üniversitesi



Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulu'nun ...../...../..... tarih ve ...../..... sayılı kararıyla kabul edilmiştir.

Prof. Dr. Narin DERİN

Enstitü Müdürü

## ETİK BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün safhalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı beyan ederim.

Öğrencinin

Nesrin ÇELİK



Tez Danışmanı

Doç. Dr. Uğur BİLGE



## TEŐEKKÜR

Akademik geliřimim iin atmiř olduđum bu ilk adımı tamamlamıř olmanın mutluluđu ve gururu ierisinde, yapmam gereken ok alıřma ve đrenmem gereken ok Őeyin olduđunun farkında olarak;

Tez alıřmam sũresince deđerli bilgilerini benimle paylařarak alıřmamı tamamlamamı sađlayan, sayın danıřmanım Do. Dr. Uđur BİLGE'ye ve kendi alıřmalarından elde ettiđi verileri ve uzmanlık bilgilerini paylařan deđerli hocam Prof. Dr. Hilmi UYSAL'a teőekkũrũ bir bor bilirim.

Tez alıřmalarım sırasında anlayıřını esirgemeyen bařta Mũdũrũm Do. Dr. Olgun KİTAPCI olmak üzere tũm AKİŐMER TTM ailesine teőekkũr ederim.

Ayrıca alıřmam boyunca yorulmadan bana moral ve motivasyon sađlayan, daima yanımda olan ve teknik bilgilerini bana aktaran ruh eřime ok teőekkũr ederim.

Bũyũk fedakãrlıklar gũstererek eđitim ve đretim hayatım boyunca beni maddi ve manevi olarak destekleyen, her konuda arkamda duran, daima sonsuz sevgilerini hissettiđim ve bař tacım olan ok kıymetli babacıđım, ok kıymetli anneciđim ve biricik kardeřime sonsuz teőekkũrlerimi sunarım.

## ÖZET

**Amaç:** Amiyotrofik lateral skleroz (ALS), beyindeki ve omurilikteki sinir hücrelerini etkileyen, istemli kas hareketini kontrol eden bir motor nöron hastalığıdır. Veri madenciliği, veri tabanlarından veya örtük verilerden, bilinmeyen, anlamlı bir takım sonuçlar çıkarılmasını sağlayan disiplindir. Bu çalışma, veri madenciliği yöntemlerini kullanarak; ALS hastası olan kişilerin klinik özellikleri ile mutasyon tipleri arasında bir ilişkinin olup olmadığının incelenmesi amacı ile yapılmıştır.

**Yöntem:** Bu çalışmada, ALS teşhisi konulmuş 65 hastaya ait veriler kullanılmıştır. Veri setinde her hasta için klinik, genetik ve demografik olmak üzere 80 özellik bulunmaktadır. Özellikler arasındaki ilişkiyi incelemek için, R paket programı ve Java programlama dilinde geliştirilen yazılımlar kullanılmıştır. Veriler üzerinde, hiyerarşik kümeleme ve karar ağacı veri madenciliği yöntemleri uygulanmıştır.

**Bulgular:** ALS hastalarında yaşadığı yer, cinsiyeti, sigara kullanımı ve egzersiz durumu gibi çevresel ve demografik faktörler bakımından ilişki tespit edilememiştir. Klinik özellikler bakımından anlamlı ilişkiler tespit edilmiştir. Karar ağacı yönteminin veri seti üzerinde anlamlı sonuçlar üretmediği tespit edilmiştir. Genetik faktörler açısından hastalar SOD1 mutasyonlular ve diğerleri olarak gruplandırılmıştır. Mutasyon tiplerinde ALS tutulumu ve klinik özellikleri bakımından anlamlı sonuçlar elde edilememiştir. Yine demografik faktörler üzerinde yapılan analizlerde istatistiksel anlamlı ilişkiler ortaya çıkmamıştır.

**Sonuç:** Sağlık alanında elde edilen veriler üzerinde veri madenciliği yöntemleri kullanılarak yapılan çalışmalar oldukça yaygındır. Çalışmada klinik bulgular arasındaki ilişkilerin anlamlılığı veri madenciliği yöntemlerinin, kişinin oluşturabileceğinden fazla alternatifleri oluşturarak çeşitli ilişkileri tanımlayabileceğini göstermektedir. Bu nedenle hastalığın tespiti şuan için klinik değerlendirmeler ile yapılırsa da mevcut veya yeni veri madenciliği yöntemlerinin süreç içerisinde daha önemli bir hale gelmesi kaçınılmazdır.

**Anahtar Kelimeler:** amiyotrofik lateral skleroz, veri madenciliği, hiyerarşik kümeleme, R paket programı

## ABSTRACT

**Objective:** Amyotrophic lateral sclerosis (ALS) is a motor neuron disease that controls voluntary muscle movement, affecting nerve cells in the spinal cord and spinal cord. Data mining is a discipline that allows extraction of unknown, meaningful sets of results, implicit or explicit in the database. This study was conducted with the aim of investigating whether there is a relationship between the clinical characteristics of individuals with ALS and mutation types, using data mining methods.

**Method:** In this study, data from 65 patients diagnosed with ALS were used. The dataset contains 80 features, including clinical, genetic and demographic, for each patient.. R package program and software developed in Java programming language were used to examine the relationship between the features. On the data, hierarchical clustering and decision tree data mining methods were applied.

**Results:** No relationship was established between environmental and demographic factors such as place of residence, gender, smoking status, and exercise status in ALS patients. Significant relationships were found in terms of clinical features. It has been determined that the decision tree method does not produce meaningful results on the data set. In terms of genetic factors, patients were grouped as SOD1 mutations and others. Significant results were not obtained in terms of ALS involvement and clinical features in mutation types. Again, statistically significant relationships did not appear in the analyzes made on demographic factors.

**Conclusion:** Studies on data obtained in the field of health using data mining methods are quite common. The significance of the relationships between clinical findings in the study suggests that data mining methods can define various associations by creating more alternatives than the person can create. For this reason, it is inevitable that current or new methods of data mining become more important in the process, even though clinical evaluation is done for the diagnosis of the disease.

**Key words:** amyotrophic lateral sclerosis, data mining, hierarchical clustering, R package program

# İÇİNDEKİLER

<b>ÖZET</b>	i
<b>ABSTRACT</b>	ii
<b>İÇİNDEKİLER</b>	iii
<b>ŞEKİLLER DİZİNİ</b>	v
<b>TABLolar DİZİNİ</b>	vii
<b>SİMGELER ve KISALTMALAR</b>	viii
<b>1. GİRİŞ</b>	1
<b>2. GENEL BİLGİLER</b>	3
2.1. Amiyotrofik Lateral Skleroz (ALS) Kavramı	3
2.1.1. Etiyolojisi	7
2.1.2. Çevresel Faktörler	8
2.1.3. Genetik Faktörler	10
2.1.4. Epigenetik Faktörler	12
2.2. Tıp Bilişimi	13
2.3. Veri Madenciliği Kavramı	16
2.3.1. Veri Madenciliği Süreci	18
2.3.2. Veri Madenciliği Modelleri	23
2.3.3. Veri Madenciliğinin Uygulama ve Kullanım Alanları	27
2.3.4. Tıp ve Sağlık Hizmetlerinde Veri Madenciliği Çalışmaları	29
<b>3. GEREÇ ve YÖNTEM</b>	36
3.1. Araştırmanın Amacı	36
3.2. Araştırma Soruları	36
3.3. Araştırmanın Evren ve Örneklemi	36
3.4. Araştırmanın Etik Yönü	36
3.5. Veri Toplama Araçları	37
3.6. Veri Seti Genel Bilgileri	37
3.7. Verilerin Analizi	41



<b>4. BULGULAR</b>	44
4.1. Mutasyonu Tanımlı Olmayan Hasta Grubu	44
4.2. Mutasyonu Tanımlı Hasta Grubu	53
4.3. Toplam Hasta Grubu	65
4.3.1. Toplam Hasta Grubu - Ki kare Analizi	68
4.3.2. Toplam Hasta Grubu - Değişken Matrisi	69
4.3.3. Toplam Hasta Grubu - Hasta Matrisi	71
<b>5. TARTIŞMA</b>	74
<b>6. SONUÇ VE ÖNERİLER</b>	77
<b>KAYNAKLAR</b>	83
<b>EKLER</b>	
<b>EK-1. Terimler Sözlüğü</b>	
<b>ÖZGEÇMİŞ</b>	

## ŞEKİLLER DİZİNİ

	<b>Sayfa</b>
<b>Şekil 2.1.</b> Gözden geçirilmiş El Escorial tanı kriterleri	6
<b>Şekil 2.2.</b> Avrupa ve Asya popülasyonlarında FALS ve SALS' in genetik mimarisi	11
<b>Şekil 2.3.</b> Hastalığa neden olan faktörlerin ağırlığı	12
<b>Şekil 2.4.</b> Tıp bilişiminin ilgi alanları	14
<b>Şekil 2.5.</b> Bilgi keşfi süreci	18
<b>Şekil 2.6.</b> CRISP-DM süreci	19
<b>Şekil 2.7.</b> Veri kalitesinin kavramsal bir çerçevesi	21
<b>Şekil 3.1.</b> 65 hastanın yaş aralığına göre dağılımı	37
<b>Şekil 3.2.</b> 65 hastanın cinsiyete göre dağılımı	38
<b>Şekil 3.3.</b> 65 hastanın doğum yeri büyüklüklerine göre sınıflandırılması	38
<b>Şekil 3.4.</b> 65 hastanın doğum yeri özelliği	38
<b>Şekil 3.5.</b> 65 hastanın hayatını kaybettiği dönem aralığı	39
<b>Şekil 3.6.</b> 65 hastanın sigara kullanım durumu	39
<b>Şekil 3.7.</b> 65 hastanın mutasyon bilgileri	40
<b>Şekil 3.8.</b> 65 hastanın düzenli fiziksel egzersiz yapma durumu	40
<b>Şekil 3.9.</b> 65 hastanın meslek bilgileri	41
<b>Şekil 4.1.</b> 38 hastanın tüm verilerinin dendogram görüntüsü	45
<b>Şekil 4.2.</b> 38 hastanın tüm verilerinin ısı haritası (heatmap) görüntüsü	46
<b>Şekil 4.3.</b> 38 hastanın tüm verilerinin kümeleme grafiği görüntüsü	47
<b>Şekil 4.4.</b> 38 hastanın demografik verilerinin dendogram görüntüsü	48
<b>Şekil 4.5.</b> 38 hastanın demografik verilerinin ısı haritası (heatmap) görüntüsü	49
<b>Şekil 4.6.</b> 38 hastanın demografik verilerinin kümeleme grafiği görüntüsü	50

<b>Şekil 4.7.</b> 38 hastanın klinik verilerinin dendogram görüntüsü	51
<b>Şekil 4.8.</b> 38 hastanın klinik verilerinin ısı haritası (heatmap) görüntüsü	52
<b>Şekil 4.9.</b> 38 hastanın klinik verilerinin kümeleme grafiği görüntüsü	53
<b>Şekil 4.10.</b> 27 hastanın tüm verilerinin dendogram görüntüsü	54
<b>Şekil 4.11.</b> 27 hastanın tüm verilerinin ısı haritası (heatmap) görüntüsü	55
<b>Şekil 4.12.</b> 27 hastanın tüm verilerinin kümeleme grafiği görüntüsü	56
<b>Şekil 4.13.</b> 27 hastanın demografik verilerinin dendogram görüntüsü	57
<b>Şekil 4.14.</b> 27 hastanın demografik verilerinin ısı haritası (heatmap) görüntüsü	58
<b>Şekil 4.15.</b> 27 hastanın demografik verilerinin kümeleme grafiği görüntüsü	59
<b>Şekil 4.16.</b> 27 hastanın klinik verilerinin dendogram görüntüsü	60
<b>Şekil 4.17.</b> 27 hastanın klinik verilerinin ısı haritası (heatmap) görüntüsü	61
<b>Şekil 4.18.</b> 27 hastanın klinik verilerinin kümeleme grafiği görüntüsü	62
<b>Şekil 4.19.</b> 27 hastanın genetik verilerinin dendogram görüntüsü	63
<b>Şekil 4.20.</b> 27 hastanın genetik verilerinin ısı haritası (heatmap) görüntüsü	64
<b>Şekil 4.21.</b> 27 hastanın genetik verilerinin kümeleme grafiği görüntüsü	65
<b>Şekil 4.22.</b> 65 hastanın tüm verilerinin dendogram görüntüsü	66
<b>Şekil 4.23.</b> 65 hastanın tüm verilerinin ısı haritası (heatmap) görüntüsü	67
<b>Şekil 4.24.</b> 65 hastanın tüm verilerinin kümeleme grafiği görüntüsü	68
<b>Şekil 4.25.</b> Değişken matrisi hasta yoğunluk analizi	70
<b>Şekil 4.26.</b> Hasta matrisi değişken yoğunluk analizi	72

## TABLÖLAR DİZİNİ

	<b>Sayfa</b>
<b>Tablo 4.1.</b> Ki kare sonuç tablosu	69
<b>Tablo 4.2.</b> Değişken matrisi hasta yoğunluk analizi sonuç tablosu	71
<b>Tablo 4.3.</b> Hasta matrisi değişken yoğunluk analizi sonuç tablosu	73
<b>Tablo 6.1.</b> Kümeleme analizi bulguları özet tablosu	78



## SİMGELER ve KISALTMALAR

<b>ALS</b>	: Amyotrofik Lateral Skleroz
<b>CRISP-DM</b>	: Cross-Industry Standard Process for Data Mining
<b>C9ORF72</b>	: Chromosome 9 open reading frame 72
<b>EMG</b>	: Elektromiyografi
<b>FALS</b>	: Familial ALS
<b>LMN</b>	: Lower Motor Neurone
<b>NCS</b>	: Nerve Conduction Studies
<b>SALS</b>	: Sporadik ALS
<b>SOD 1</b>	: Süperoksit dismutaz 1
<b>UMN</b>	: Upper Motor Neurones
<b>YZ</b>	: Yapay Zeka

## 1.GİRİŞ

Amyotrofik Lateral skleroz (ALS) beyin ve spinal kordun üst ve alt motor nöronlarında dejenerasyonla karakterize edilen ilerleyici ve ölümcül nöromusküler bir hastalıktır. ALS ABD’de Amerikalı ünlü beyzbolcu Lou Gehrig’de görülmesinden dolayı yaygın olarak Lou Gehrig hastalığı olarak, Avrupa’da Motor Nöron Hastalığı olarak da bilinir. ALS hastalarının çoğunluğu teşhis edildikten sonra 2-5 yıl içerisinde hayatını kaybetmektedir. Hastalığın kalıtsal bir şekli olan ailesel ALS, olguların %5-10'unu oluştururken, kalan sporadik olguların etiyojisi hala açıkça tanımlanmamıştır. ALS için kanıtlanmış etkili tedavi edici müdahalelerin eksikliği devam etmekte olan bir sorundur. Mevcut tedavilerin ALS'yi iyileştirmediği ancak hastalığın ilerlemesini yavaşlattığını göstermiştir (Mehta ve ark., 2018).

ALS'nin kesin nedeni bilinmemekle birlikte, kimyasala maruz kalma, mesleki maruziyet, askerlik, beslenme, fiziksel aktivite ve travmanın da dâhil olduğu ALS'ye neden olan faktörlerle ilgili birçok hipotez öne sürülmüştür. Hasta özellikleri bakımından genel olarak ALS tüm ırk ve etnik kökenleri etkilemektedir; Bununla birlikte, beyazlarda, erkeklerde, İspanyol olmayanlarda, 60 yaşın üzerindeki kişilerde ve aile öyküsü olanlarda ALS hastalığının gelişme olasılığı daha yüksektir (Mehta ve ark., 2014).

Bugüne kadar ailesel ALS'nin yaklaşık üçte ikisinin genetik etiyojisi ve sporadik hastalığın yaklaşık % 10'u tespit edilmiştir (Renton ve ark., 2013). Genetik mutasyonlar ailesel hastalığın geri kalan üçte birinden sorumludur ancak geride kalan sporadik hastalığın ne kadarının genetik olduğu, çevresel maruz kalma, yaşlanma veya yaşam tarzı seçimleri gibi diğer faktörlerden ne kadar etkilendiği bilinmemektedir (Marangi ve Traynor, 2015).

ALS'nin patogeneze katkıda bulunan genetik ve çevresel faktörlerin çok fazla olması ve genetik ve/veya çevresel faktörler arasındaki karmaşık etkileşim nedeniyle de bu hastalıklara "kompleks hastalıklar" denilmektedir (Krakauer ve ark., 2002).

Günümüzde, tıp ve sağlık alanında bilişim teknolojilerinin kullanımının yaygınlaşması ve ALS gibi kompleks hastalıklar üzerinde bu araçların kullanılması klinisyenlere ve

hastalığa ait tanımlamaların yapılmasına yardım eder hale gelmiştir. Bu anlamda öne çıkan konu başlıklarından bir tanesi de veri madenciliğidir. Veri madenciliği, önceden bilinmeyen, gizli ve büyük ölçekli örtük verilerden anlamlı bilgiler elde edilmesini sağlayan veri tabanlarındaki bilgi keşfi süreci içinde yer alan bir adımdır ( Han ve Kamber, 2006). Günümüzde veri madenciliği bankacılıktan sağlığa, pazarlamadan biyolojiye kadar birçok alanda karşılık bulmaktadır.

Özellikle sağlık hizmetlerindeki büyük veri tabanı ve veri madenciliği uygulamaları değerlendirildiğinde yapılan çalışmalar geniş kategorilerden/başlıklardan (tedavi etkinliği, sağlık yönetimi, ilaç endüstrisi vb.) oluşmaktadır. Sağlık hizmetlerindeki verilerden anlamlı bilgilerin elde edilmesi için kullanılan birden fazla model bulunmaktadır. Veri madenciliğine ait bu yöntemler; Sınıflandırma, Kümeleme ve Birliktelik Kuralları olarak 3 ana başlık altında gruplandırılmıştır.

Sınıflandırma, veriyi daha önceden tanımlanmış sınıflara ayırma işlemidir. Veri madenciliğinde kümeleme ve karar ağacı en sık kullanılan yöntemler arasında yer almaktadır. Kümeleme yönteminde amaç gruplandırma yapmak ve önceden tanımlanmamış sınıflara ayırmaktır. Bu yöntem sınıf aralıklarının henüz belli olmadığı durumlara uygun bir araçtır. Yöntemde başlangıç kümelerinin sayısı belirli değildir ve analiz çıktıları dendogram adı verilen şekiller ile gösterilmektedir (Koyuncugil ve Özgülbaş, 2009). Karar ağacında ise amaç veri setini tanımlamak ve tahmin etmektedir. Karar ağacı bilinen örnek veriden tümevarım yöntemini kullanarak dallardan oluşan bir ağaç şeklini oluşturur. Bu yöntem, eğitim setinden elde ettiği kurallar doğrultusunda basit karar adımlarını uygulayarak büyük veriyi küçük gruplara bölmek için kullanılır (Albayrak ve Yılmaz, 2009). Veri madenciliği yöntemlerinin uygulanabilmesi için birçok paket program geliştirilmiş ayrıca programlama dilleri kullanılarak yeni programların geliştirilmesine de devam edilmektedir. Bu programlara R, WEKA, Rapid Miner, MOA örnek olarak verilebilir.

Bu tez çalışması ile ALS tanısı konmuş hastalara ait klinik, demografik ve genetik veri seti üzerinde, R paket programı ve Java programlama dilinde geliştirilmiş hiyerarşik kümeleme ve karar ağacı yöntemleri kullanılarak hastaların genetik ve klinik özellikleri arasındaki ilişkilerin belirlenmesi hedeflenmiştir.

## 2. GENEL BİLGİLER

### 2.1. Amiyotrofik Lateral Skleroz (ALS) Kavramı

İlk kez 1874 yılında Fransız nörolog Jean-Martin Charcot tarafından tanımlanmış olan Amiyotrofik Lateral Skleroz (ALS) kavramı “korteks, beyin sapı ve omurilikte yer alan motor nöronların seçici kaybıyla karakterize edilen ölümcül bir nörodejeneratif hastalıktır” (Eisen ve Krieger, 2006). ALS, beyindeki ve omurilikteki sinir hücrelerini (nöronları) etkileyen, istemli kas hareketini kontrol eden nadir bir nörolojik hastalıktır (Office of Communications and Public Liaison, 2013). ALS'nin patogenezinde katkıda bulunan genetik ve çevresel faktörlerin çok fazla olması göz önüne alındığında, genetik ve/veya çevresel faktörler arasındaki karmaşık etkileşim nedeniyle bunlara "kompleks hastalıklar" da denir (Krakauer ve ark., 2002).

Motor nöronlar, beyinden omurilik (üst motor nöron, UMN veya 1. motor nöron) ve omurilikten vücuttaki kaslara kadar uzanan (alt motor nöron, LMN veya 2. Motor nöron) sinir hücreleridir. Bu motor nöronlar çiğneme, yürüyüş, nefes alma ve konuşma gibi istemli kasları başlatır ve hayati iletişim bağlantılarını sağlar.

Beyindeki motor nöronlardan gelen mesajlar spinal korddaki motor nöronlara, beyin sapı motor çekirdeklere ve omurilikteki motor çekirdeğinden belli bir kasa veya kaslara iletilir.

ALS'de hem üst motor nöronlar hem de alt motor nöronlar bozulur veya ölür ve kaslara mesaj göndermeyi durdurur. Kaslar giderek zayıflar, seğirmeler başlar (fasikülasyonlar), yapı ve görev bakımından zayıflar (atrofi). Sonunda, beyin istemli hareketleri başlatma ve kontrol etme kabiliyetini kaybeder (Office of Communications and Public Liaison, 2013).

ALS'nin sporadic ve familial (ailevi veya genetik) olmak üzere 2 formu olduğu bilinmektedir.

**Sporadik ALS (SALS):** ALS vakalarının %90 veya daha fazlası sporadik olarak kabul edilir. Bu da hastalığın aile hikâyesi veya belirgin bir şekilde ilişkili risk faktörü olmaksızın ortaya çıktığı anlamına gelmektedir. Her ne kadar Sporadik ALS hastalarının aile üyeleri için de bir risk olsa da genel risk çok düşüktür.



**Familial ALS (FALS):** Tüm ALS vakalarının yaklaşık %5 ila %10'u aileseldir. Bu da bir kişinin hastalığı ailesinden miras aldığı anlamına gelmektedir. Ailevi vakaların yaklaşık yüzde 25 ila 40'ı "kromozom 9 açık okuma çerçevesi 72" veya C9ORF72 olarak bilinen bir gendeki bir kusurdan kaynaklanmaktadır. Ayrıca ailevi olguların yüzde 12-20'si, bakır-çinko Süperoksit dismutaz 1 (SOD1) enziminin üretimi için talimatlar sağlayan gendeki mutasyona bağlı olarak ortaya çıkmaktadır (Office of Communications and Public Liaison, 2013).

Klinik olarak ALS fenotipik sınıflandırma belirtilerin ilk ortaya çıktığı yere göre sınıflandırılır. Olguların %65'lik yüksek oranı kol ve bacakların tutulumu, güçsüzlüğü ile ortaya çıkan klasik formudur. Olguların %30'luk kısmında hastalık, dizartri (konuşurken kullanılan kasların kontrol ya da koordine edilmesinde güçlük yaşanması), disfaji (yutma güçlüğü) veya her ikisi ile başlayabilirken Olguların %5'inde ise solunum yetmezliği ile başlar (Zufiri 'a ve ark., 2016).

Her ne kadar ALS, yetişkinlikte herhangi bir zamanda ortaya çıkabilse de genellikle ellili yaşların ortalarındaki insanları etkilemektedir. ALS insidansı 100.000 kişide 1,2 – 4 olgu olduğu tahmin edilmektedir. Hastalık, hastalık başlangıcını takiben hastalık seyri boyunca tüm motor nöronlara yayıldığı için ALS hastalığından etkilenen bireylerin çoğu 2-5 yıl içinde hayatlarını kaybetmektedirler (Du ve ark., 2017). Sadece hastaların %35'inin 5 yıl veya daha uzun süre hayatta kaldığı belirtilmiştir (Shaw ve ark., 2001).

Başlangıcından itibaren giderek daha belirgin semptomlar halinde gözlenen ALS'nin erken belirtilerinden bazıları aşağıda verilmiştir (Office of Communications and Public Liaison, 2013).

- Kol, bacak, omuz veya dilde fasikülasyonlar (seyirme)
- Kas krampları
- Kaslarda tonus artışı (spastisite)
- Bir kol, bacak, boyun veya diyaframı etkileyen kas güçsüzlüğü,
- Belirsiz ve kısık ses ile konuşma
- Çiğneme ve/veya yutma güçlüğü

ALS aşamalı olarak artan bir hastalık olmakla birlikte ilk belirtileri kişilerde oldukça değişiklik gösterebilir. Bir kişi konuşma esnasında ses tonunda değişiklik ile karşılaşabilirken diğer bir kişi kalemi kavrayamayabilir veya bir fincanı kaldırmada sorun yaşayabilir. ALS motor nöronları etkilediğinden görme, dokunma, işitme, tat, koku hisleri, göz kasları ve mesane genellikle etkilenmez (The ALS Association, 2005).

ALS hastalığının teşhisinde tanı amacı ile kullanılan biyokimyasal, biyolojik bir belirleyici ve laboratuvar testi yoktur. Tanı, ilerici ekstremite, bulbar veya solunum kas zayıflığının tarihsel bulguları ve üst ve alt motor nöronun (UMN ve LMN) fiziksel belirtilerine dayanarak klinik olarak yapılır. Bazı hastalıkları dışlamak için elektromiyografi (EMG) ve sinir iletim çalışmaları (NCS) yapılmaktadır. El-Escorial Dünya Nöroloji Federasyonu, ALS için tanı kriterlerini, klinik muayene ve EMG nedeniyle vücut bölgelerinin sayısına göre derecelendirilmiştir (Shaw ve ark., 2001).

ALS teşhisi için ilk kez 1990 yılında Madrid yakınlarında toplantının yerinin isminin verildiği “*El Escorial Kriterleri*” oluşturulmuştur. Bu kriterler Şekil 2.1’de sıralanmıştır.

**ALS tanısı için gerekli olanlar:**

**A-1** Klinik, elektro fizyolojik veya histopatolojik olarak alt motor nöron tutulumunun kanıtlanması

**A-2** Klinik muayene ile üst motor nöron tutulumunun kanıtlanması

**A-3** öykü veya muayene ile belirti ve bulguların bir vücut bölgesi içinde ya da diğer bölgelere ilerleyici özellikte yayılması

**ALS tanısı için olmaması gerekenler:**

**B-1** Alt motor nöron ve/veya üst motor nöron tutulumunu açıklayabilecek başka bir hastalığı telkin eden elektro fizyolojik ve histopatolojik veriler

**B-2** Var olan klinik ve elektro fizyolojik bulguları görüntüleme yöntemleriyle açıklayabilecek veriler

**Bu ilkeler çerçevesinde;**

**Kesin ALS**

Üç bölgede üst ve alt motor nöron bulguları

**Olası ALS**

İki bölgede üst ve alt motor nöron bulgularıyla birlikte en azından bazı üst motor nöron tutulumu bulgularının alt motor nöron tutulumu bulgularının daha üstünde yer alması

**Olası ALS-Laboratuvar destekli**

Bir ya da birden çok üst motor nöron tutulumu bulgusu ve EMG ile en az iki bölgede alt motor nöron bulgusuna rastlanması

**Mümkün ALS**

Bir bölgede üst ve alt motor nöron bulgularının birlikte olması,

İki ya da daha çok bölgede üst motor nöron bulgularının olması veya

Üst motor nöron bulgularının alt motor nöron bulgularının üstünde yer almaksızın, iki bölgede üst ve alt motor nöron bulgularının olması

**Şekil 2.1.** Gözden geçirilmiş El Escorial tanı kriterleri (Şimşek Erdem, 2013)

El Escorial Tanı Kriterleri'nin yayınlanmasının ardından düzenlenmesi gerekliliği ile ilgili olarak tartışmalar meydana gelmiştir. Bunun üzerine, ALS Araştırma Grubu, Ocak-Nisan 2014 ve Ocak-Mart 2015 tarihleri arasında genel yorumlamalar için ALS sınıflandırılmasına ilişkin bir belge geliştirilmiştir. Bu belgeye göre halen bilinen ALS fenotipleri:

- Progresif bulber felci (PBP)
- Flail kol (Vulpian Bernhard) sendromu ve Flail bacak sendromu
- Progresif kas atrofisi (PMA)
- Primer lateral skleroz (PLS)

olarak verilmiştir (Ludolph ve ark., 2015).

Gelecekte beyin omurilik sıvısının (BOS), serumun ve idrarın proteomik (Proteinler ile ilgili bilgiyi topluca sağlamayı amaçlayan disiplin) ve metabolomik (Metabolitler içeren kimyasal süreçlerin bilimsel çalışması) profilleri tanıya yardımcı olabilmesi ve hastalık ilerlemesinin gözlemlenebilmesi için bir belirteç görevi görebileceği düşünülmektedir (Shaw ve ark., 2001).

ALS için henüz tedavi yöntemi bulunmamakla birlikte belirtileri kontrol altına almaya, komplikasyonları önlemeye yönelik tedaviler mevcuttur. Bunların başında sinir hücreleri ve motor nöronlar arasındaki iletileri taşımaya yönelik üretilen ilaçlar, bireyin bağımsızlığını arttırmaya ve kaslarını güçlendirmeye yönelik yapılan fizik tedavi yöntemleri gelmektedir. Ayrıca iletişim becerisini sürdürmeye yardımcı konuşma egzersizleri, özellikle yutma güçlüğü çeken bireyler açısından yemeklerin planlanması, yeterli miktarda besin alınması amacıyla beslenme desteği ve nefes almadan sorumlu kasların zayıflamasına bağlı olarak rahat nefes almalarına yönelik olarak nefes egzersizleri de hastalar için uygulanabilir tedavi yöntemleri olarak belirtilmektedir. Bu tedaviler tek bir hekim tarafından değil disiplinler arası ekipler (aile hekimi, terapist, beslenme uzmanları, evde bakım hemşireleri vb.) tarafından sağlanmaktadır. (Office of Communications and Public Liaison, 2013).

### **2.1.1. Etiyolojisi**

ALS tüm ırk ve etnik kökenden insanları etkilemekte birlikte ALS'nin etiyolojisi henüz tam olarak tespit edilememiştir, fakat ALS için birkaç potansiyel risk faktörü bulunmaktadır. Bunlar;

**Yaş:** Hastalık her yaşta olsa da, semptomlar en çok 55 ila 75 yaşları arasında gelişir.

**Cinsiyet:** Erkeklerin hastalığa yakalanma riski kadınlara göre biraz daha fazladır ancak, yaşlandıkça erkekler ile kadınlar arasındaki fark kaybolur.

**İrk ve Etnik Köken:** Kafkasyalılar ve Hispanik (Amerikan Nüfus Dairesi tarafından kültürel kökenleri Meksika, Porto Riko, Küba, Orta Amerika ve diğer Latin Amerika ülkeleri olanlar kişiler için kullanılmış olan terimdir.) olmayan kişilerin hastalığa yakalanma olasılığı yüksektir (Office of Communications and Public Liaison, 2013).

Daha önce de belirtildiği üzere vakaların %90'lık kısmı sporadiktir bu nedenle ALS hastalığının nedeni tam olarak tanımlanamamakla birlikte “sigara kullanımı, atletizm, ileri yaş, siyano bakteriler, kurşun, bakır, böcek ilacı, tarımsal kimyevi maddeler, sigara, yoğun fiziksel aktivite, kafa travması, elektromagnetik alan perinatal etkenler sorumlu tutulmaktadır (Koca, 2015)”. Bunların yanı sıra hastalığın ortaya çıkmasında;

- Genetik nedenler,
- Glutamat eksitotoksitesisi,
- Viral enfeksiyonlar,
- Otoimmün reaksiyonlar,
- Ağır metal zehirlenmeleri (kurşun, cıva ve alüminyum vb.)

gibi çevresel etkenlerin rol oynadığı ileri sürülmüştür (Hastings ve Goedert, 2013).

### 2.1.2. Çevresel Faktörler

**Ağır Metaller:** ALS gelişiminde kurşun, cıva, kadmiyum ve selenyum gibi ağır metaller rol oynamaktadır (Sutedja, ve diğerleri, 2009). Özellikle metil cıva ve etil cıva balık, kabuklular ve besin zinciri boyunca insanlarda birikir ve hücre metabolizmayı bozarak hücre ölümüne ve klinik hastalığa sebep olur. Cıva toksisitesinin (Bir kimyasal maddenin zarar verme kapasitesi) biyokimyasal mekanizmaları, bireysel genetik yatkınlığa, metalin kimyasal formlarına ve maruz kalmanın uzunluğuna bağlıdır (Yu ve Pamphlett, 2017).

ALS ile kurşun arasında da bir ilişki önerilmiştir (Kamel ve ark., 2002). Kaynak veya lehim materyallerine maruz kalan işçiler ALS gelişiminde risk altında görünmektedir (Strickland ve ark., 1996). Kurşunun sinir sistemi üzerindeki toksik etkileri, öncelikle çocuklarda beyin dokusundaki dejeneratif değişikliklerin görülmesini ve öncelikli olarak erişkinlerde motor sinirlerdeki hastalıkları içermektedir. (Yu ve Pamphlett, 2017). Aynı zamanda iskelet kası kurşun için bir saklanma yeri olabilmektedir. (Condradi ev ark., 1978).

Bunun ile birlikte mesleki olarak kurşuna maruz kalma ile ALS arasındaki ilişkiyi incelemek için yapılan gözlemsel meta analiz çalışma sonucunda, kurşuna mesleksi

maruz kalma öyküsü olan kişilerde maruz kalmayan kişilere oranla daha yüksek oranda ALS gelişme ihtimalinin olduğu ortaya koyulmuştur (Wang ve ark., 2014).

**Fiziksel Aktivite:** Fiziksel aktivitenin ALS hastalığı ile ilişkisini tespit etmeye yönelik yapılan çalışmalardan birçoğunda atletizm veya yoğun fiziksel aktivite ile ALS arasında bir ilişki olduğu kabul edilmiştir. Yapılan bir çalışmada İtalyan profesyonel futbol oyuncularında ALS hastalığı ve oyun süresi arasındaki ilişkiye yönelik olarak ALS'de 6 kat artış bildirilmiştir (Chiò ve ark., 2005). Aşırı fiziksel aktivite, tekrarlanan kafa yaralanmaları ve pestisitlere, besin takviyelerine veya çeşitli maddelere maruz kalma, futbolcularda görülen ALS hastalığı riskinin altında yatan nedenler olabileceği düşünülmektedir (Pamphlett ve Jew, 2013; Yu ve Pamphlett, 2017).

Yine konu ile ilgili olarak yapılan meta analiz çalışmasında dokuz çalışmanın altı tanesinde futbol ile ALS arasında anlamlı bir pozitif ilişki olduğu ve fiziksel iş gerektiren meslekler ile ALS hastalığı arasında ilişki olduğu tespit edilmiştir (Hamidou ve ark., 2014).

**Sigara:** Çeşitli araştırmalarda sigara kullanımının ALS riskini arttırdığı öne sürülmüştür. Yapılan meta analiz çalışmalarında kadınların sigara kullanımının ALS riskini arttırdığı düşüncesi desteklenmektedir (Alonso ve ark. 2010). Azınlık bir çalışma, sigaranın daha düşük risk faktörü olduğunu belirtmekle birlikte sigara, nikotin, tütün dumanı ve diğer zehirli maddeler ALS risk faktörleri arasında görülmektedir (Oskarsson ve ark. 2016).

**Diğer Çevresel Faktörler:** Das ve arkadaşlarının 110 ALS'li olgu ve 240 kontrol grubu üzerinde yapmış olduğu vaka kontrol çalışmasında ALS gelişiminde ainsel, çevresel ve mesleki risk faktörleri incelenmiştir. Grupların aile geçmişi, mesleği, yaşadığı yer, içme suyu kaynağı, endüstriyel, kimyasal, zirai toksinler ve ağır metallere maruz kalma durumu, fiziksel ve elektriksel hasarlara maruz kalma ve 10 yıldan uzun süredir manyetik alan altında çalışmaya yönelik bilgiler alınmıştır. Yapılan analizler doğrultusunda; kırsal bölgede yaşayan kişilerde elektrik çarpması veya yaralanma ile ALS gelişimi arasında kuvvetli bir ilişki tespit edilmiştir ( Das ve ark., 2012).

Bu faktörlerin yanı sıra Pupillo ve arkadaşlarının yapmış olduğu çalışmada travmaların da ALS ile ilişkili olduğu belirtilmiştir. Yapılan çalışmada hasta grupları 185 olguda (% 49.1)

belirgin ALS, 151 olguda (% 40.1) muhtemel ALS ve 41 olguda (% 10.9) olası ALS olarak tanımlanmıştır. Özellikle semptom başlangıcından itibaren 5 yıl içinde meydana gelen travma olaylarının varlığı ALS ile ilişkilendirilmiştir (Pupillo ve ark., 2012).

ALS'nin başlangıcı ve ilerlemesi ile ilişkili risk faktörlerinin tanımlanması üzerine yapılan bir sistematik inceleme ve meta analiz çalışması da çevresel faktörlerin hastalık üzerindeki etkilerini değerlendirmiştir. Wang ve arkadaşlarının yapmış olduğu bu çalışmada pestisit ve ağır metallere maruz kalma, sigara kullanımı, E vitamini, fiziksel hareket, çay kahve tüketimi ile ALS arasında bir ilişki olduğu belirtilmiştir (Wang ve ark., 2017).

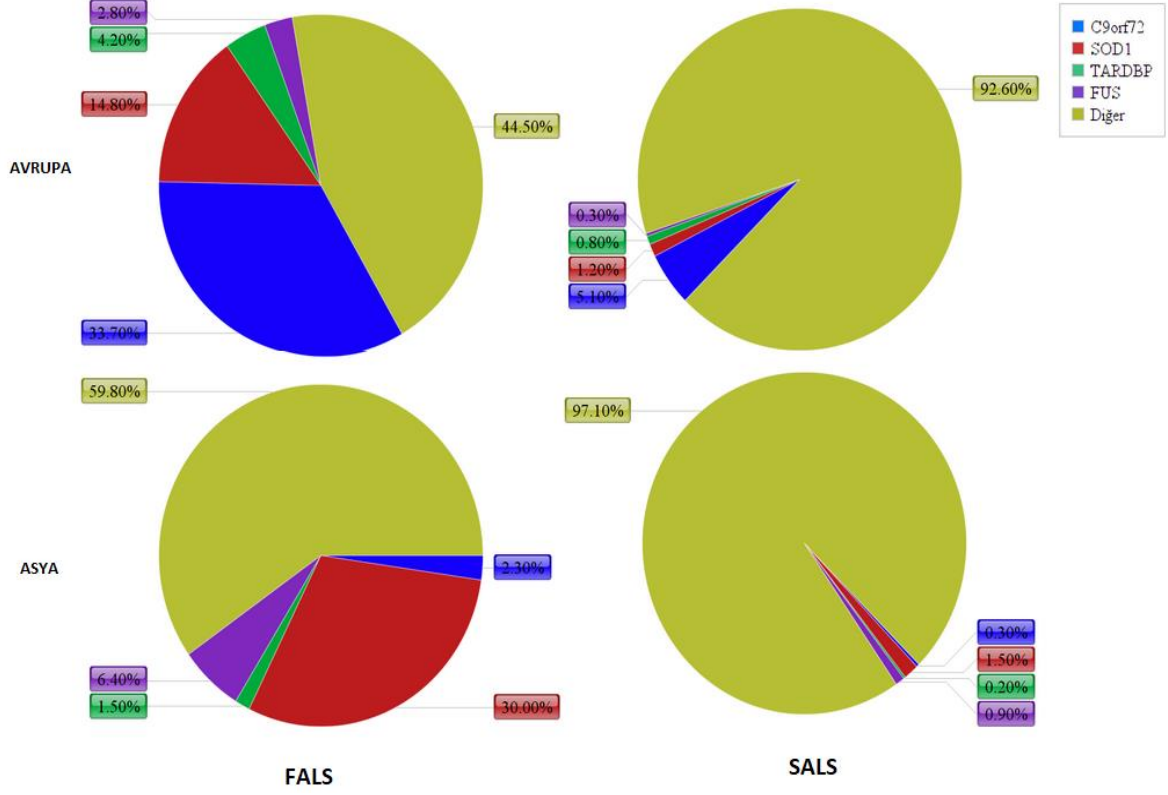
### **2.1.3. Genetik Faktörler**

Moleküler genetik araçlarının ALS'ye uygulanması yaklaşık 30 yıl önce meydana gelmiştir. Bu çabalar, ailesel ALS vakalarının % 20'sini oluşturan süperoksit dismutaz 1 (SOD1) geninin tanımlanmasına yol açmış ve SOD1'de 160'dan fazla mutasyon bildirilmiştir. 2006 yılında ALS'de ve sonrasında diğer nörodejeneratif hastalıklarda TAR DNA bağlayıcı proteinin (TDP-43 / TARDBP) keşfedilmesi önemli bir adım olmuştur. Sonrasında ALS ailelerinin taranması ile TARDBP'deki mutasyonların ailesel ALS'ye neden olduğu bulunmuştur. Ailesel ALS'ye neden olan TARDBP mutasyonlarının keşfinden kısa süre sonra başka bir RNA bağlayıcı protein FUS/TLS mutasyonların, ailesel ALS ve seyrek sporadik ALS vakalarının yaklaşık % 4'üne neden olduğu saptanmıştır. Yapılan çalışmalara C9ORF72'de mutasyonların tanımlanmasıyla destek verilmiştir ve ailesel ALS vakalarının % 30'undan, sporadik ALS vakalarının % 5'inden sorumlu olduğu tespit edilmiştir (Ajroud-Driss ve Siddique, 2014).

Bunu takip eden diğer çalışmalarda da ailesel olgulardaki ağırlıklı genetik tetikleyici mutasyonlar SOD1 (süperoksit dismutaz1), FUS, TARDBP, OPTN, VCP, UBQLN2, PFN1 ve C9orf72 (chromosome 9 open reading frame 72) olarak tanımlanmıştır (Leblond ve ark., 2014; Renton ve ark., 2015).

ALS ile ilişkili majör genlerin mutasyon frekanslarını belirlemek için yapılan meta-analiz çalışması sonucunda elde edilen sonuçlar Şekil 2.2'de gösterilmiştir. Çalışma, Asya popülasyonlarında ALS'nin genetik yapısının Avrupa popülasyonundaki genetik yapının

farklı olduğunu ve genetik testler yapılırken bu farklılıkların göz önünde bulundurulması gerekliliğini göstermiştir (Zou ve ark., 2016).



Şekil 2.2. Avrupa ve Asya popülasyonlarında FALS ve SALS' in genetik mimarisi (Zou ve ark., 2016)

Familial ile Sporadik ALS hastalarının klinik özellikleri ayırt edici olmamakla birlikte iki grup arasında küçük farklılıklar gözlemlenmiştir. Bunlardan bazıları hastalık başlangıç yaşı ve hastalık teşhisi konulan kadınların erkeklere oranlarıdır (Yılmaz, 2008). Bu değişikliklere neden olan genetik farklılıklar veya mutasyonlar çeşitli çalışmalarda tanımlanmıştır.

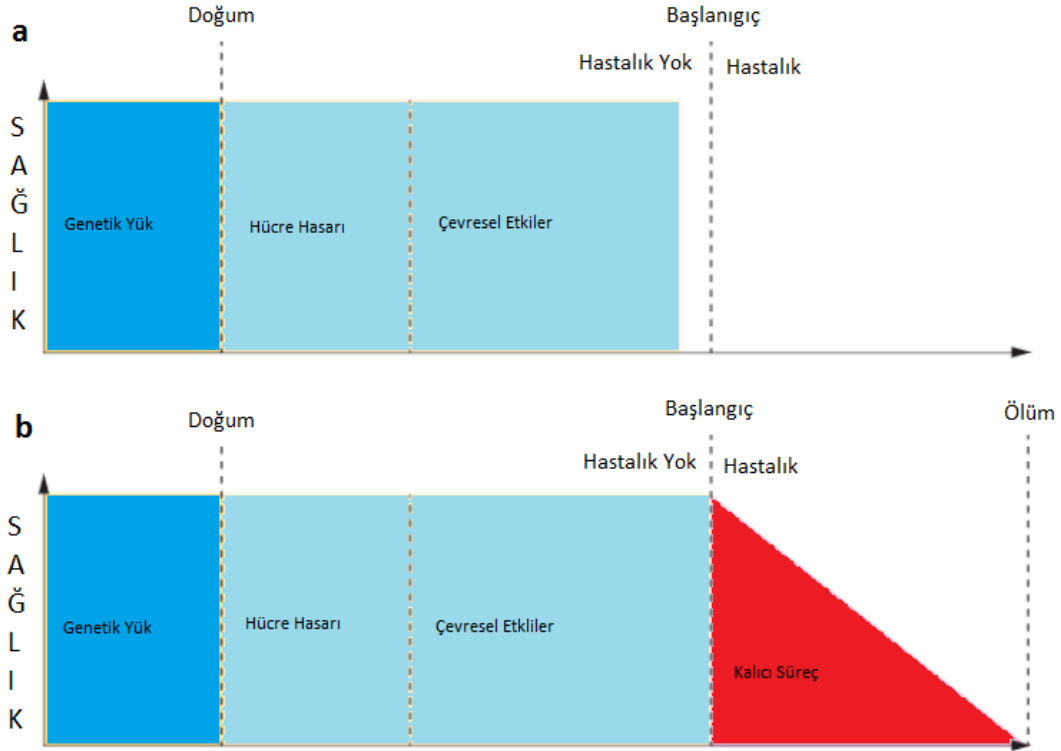
Corcia ve arkadaşları tarafından Mayıs 2017 yılında yayınlanan "ALS Genetiği" isimli yayında bu güne kadar 20 genin tanımlandığı belirtilmiş ve FALS ile ilgili olarak tespit edilen genlerin yanı sıra SALS hastalarında da bugüne kadar APOE, EPHA4, KIFAP3, UNC13A, NIMA, CCNF ve C21orf2 mutasyonların tespit edildiği belirtilmiştir (Philippe ve ark., 2017).



#### 2.1.4. Epigenetik Faktörler

Genler üstü genetik olarak tanımlanan Epigenetik, DNA'nın yapısına herhangi bir değişikliğe yol açmadan yani gen ifadesinde meydana gelen, DNA dizisinden bağımsız olan kalıtsal değişikliklere ve onları inceleyen bilime *Epigenetik* denir (Jiang ve ark., 2004).

Chalabi & Hardiman bireyde oluşan epigenetiğin etkisini ve gen-zaman-çevre modelini şu şekilde tanımlamıştır: Bir bireyde risk bileşenleri genetik faktörler, çevresel risk ve zaman ile oluşabilir. Sağlıklı bir bireyde doğumda sadece genetik faktör vardır ve zamanla çevresel faktörlerin birikmesi ile birlikte gen değişimleri hücre hasarıyla sonuçlanır. Ancak hastalık eşiğe ulaşmadığı sürece bireyde ALS oluşmayacak, eşiğe ulaşıldığında ise ALS meydana gelecektir (Şekil 2.3.). Bu nedenle çevresel etkenlerin DNA sekansını değiştirmeden genlerde kalıtsal bir değişikliğe neden olması ve ebeveynlerden çocuklara veya torunlarına ALS'nin taşınması olasıdır (Al-Chalabi ve Hardiman, 2013).



Şekil 2.3. Hastalığa neden olan faktörlerin ağırlığı (Al-Chalabi ve Hardiman, 2013)

## 2.2. Tıp Bilişimi

Tıp bilişimi ya da diğer bir ifade ile Tıbbi Bilişim bilgisayar teknolojileri ile tıp disiplini arasında bir kesişim oluşturmaktadır. Bu nedenle tıp bilişimi, tıp ve sağlık alanındaki bilgilerin kaydedilmesi, üretilmesi, değerlendirilmesi, analiz edilmesi, işlenmesi, yorumlanması, sunulması, saklanması ve ilerleyen dönemlerde kullanılmak üzere kayıtların tutulması süreçlerinin tamamıdır (Bemmel, 1984).

Bununla birlikte tıp bilişimini, İngiltere Tıp Bilişimi Derneği (British Medical Informatics Society) şu şekilde tanımlamıştır (Milli Eğitim Bakanlığı, 2011):

“Sağlık hizmetlerini yaygın olarak sunabilmek için var olan bilginin paylaşımını ve kullanımını sağlayacak araçların, becerilerin tümü” .

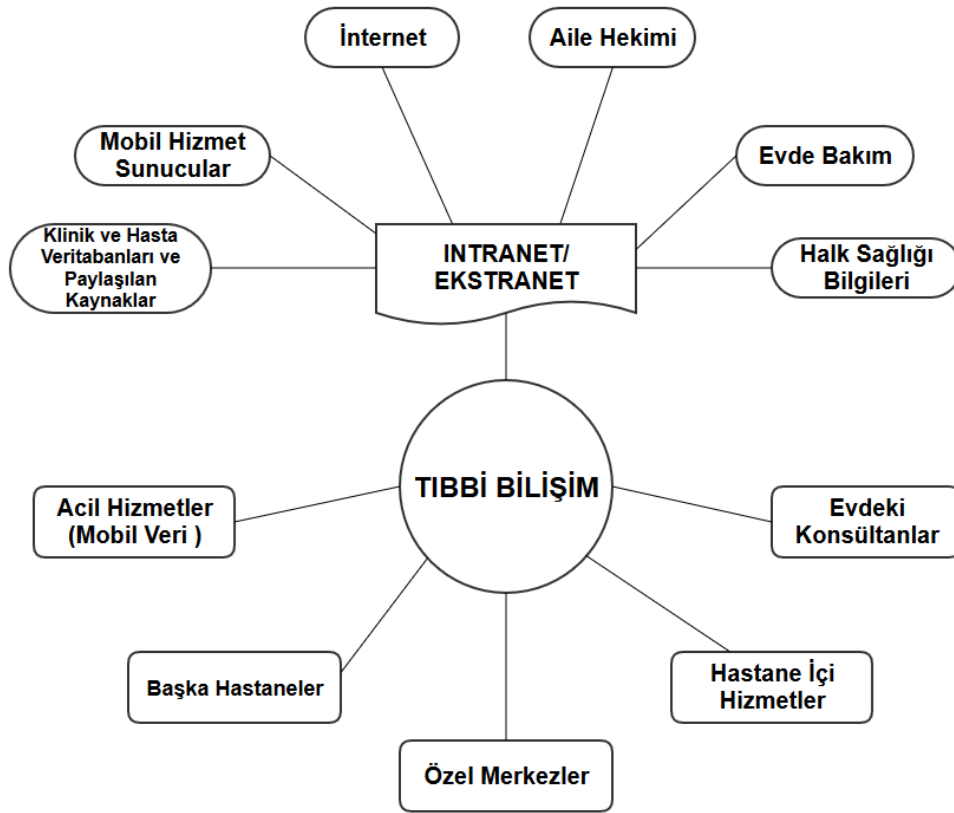
“Bilgi sistemlerinin sağlık hizmetlerine uygulanma yöntemlerini araştıran ve öğreten bir bilim dalı; sağlık, bilişim, psikoloji, epidemiyoloji ve mühendisliğin bulunduğu nokta” olarak tanımlamaktadır.

Milli Eğitim Bakanlığının Kaynak Yayın olarak sunduğu dokümanda tıp bilişimi kavramına bakacak olursak “Tıp alanında kullanılan bilgi ve iletişim teknolojilerini anlamaya, bu alandaki gelişmelere katkıda bulunmaya, katkıyı sağlayacak yöntem ve ilkeleri oluşturmaya ve sağlanan gelişmeleri kurumlar ve bireyler açısından değerlendirmeye odaklandığı” belirtilmiştir (Milli Eğitim Bakanlığı, 2011).

Aynı zamanda tıp bilişimi, tıp disiplininde ortaya çıkan bilgi yönetimi gereksinimlerine cevap vermeye çalışır. Bu doğrultuda ilgili olduğu alanlar aşağıdaki gibi sıralanabilir (Milli Eğitim Bakanlığı, 2011) (Şekil 2.4.):

- Veri ve bilgi kayıtlarının elde edilmesi, saklanması vb.
- Kodlama sistemleri
- Veri işleme
- Veri tabanı yönetim sistemleri
- Telekomünikasyon sistemleri
- Tıbbi sınıflandırma sistemleri

- Hasta kayıt sistemleri, elektronik hasta kayıtları
- Biyosinyal analizleri
- Tıbbi görüntüleme sistemleri
- Görüntü işleme ve analiz yöntemleri
- Klinik bilgi sistemleri
- Toplum hekimliği bilgi sistemleri
- Hemşirelik bilgi sistemleri
- Karar destek sistemleri
- Sağlık bilgi sistemleri
- Hastane bilgi sistemleri
- Bilgi sistemleri güvenliği”



Şekil 2.4. Tıp bilişiminin ilgi alanları (Milli Eğitim Bakanlığı, 2011)

Bilişim teknolojilerinin gelişmesi ile birlikte tıp bilişimin ilgili olduğu konular ve dolayısıyla tıp bilişiminin uygulama alanları artmaktadır. “Tıp bilişimi içerisinde yer alan

konu başlıkları ise aşağıdaki gibi belirtilebilir (Demirhan ve Güler, 2011; Milli Eğitim Bakanlığı, 2011);

- Elektronik sağlık kaydı sistemleri
- Sağlık organizasyonlarında bilgi yönetimi
- Tüketici Tıp Bilişimi ve Tele-Sağlık
- Halk sağlığı bilişimi ve sağlık bilgi altyapısı
- Hasta bakım sistemleri
- Hasta izleme sistemleri
- Bilgi erişimi ve sayısal kütüphaneler
- Klinik karar destek sistemleri
- Tıp eğitiminde bilgisayar kullanımı
- Biyoinformatik
- Sağlık bilgi sistemleri,
- Sağlıkta internet ve internet teknolojileri kullanımı,
- Biyoinformatik ve gen inforinatığı,
- Hasta kayıt ve sağlık kodlama sistemleri,
- Medikal görüntüleme sistemleri,
- Tıbbi sinyal işleme sistemleri,
- Sağlık iletişim sistemleri, teletıp, konsültasyon,
- Biyomedikal modelleme sistemleri”

En sade şekilde tıp bilişiminin veri, bilgi ve bilişimin sağlık açısından kaliteli olması için işleyişi olarak tanımlayan Haux, 2006 yılında yayınladığı Tıp Bilişiminin Geçmişi, Bugünü ve Geleceği makalesinde tıp bilişiminin 10 yıllık, 25 yıllık ve 50 yıllık süreçlerde geçmişten geleceğe olan bakışını sunmuştur. Ayrıca Haux makalesinde tıp bilişiminin gelişiminde bir takım kilometre taşları vermiştir. Bunlar;

- 1959 yılında Ledley ve Lusted’ın risk ve belirsizlik altında karar verme yöntemlerini ve sonraki bilgisayar temelli kararlar için gerekli olan metotları incelemesi,
- 1984-85 yıllarında tıp bilişimi gelişimini etkileyen makalelerin yayınlanması,

- 1999-2000 yıllarında tıp bilişiminin uluslararası anlamda bir disiplin olmasında uzlaşma sağlandığının ilk kanıtı olan IMIA (International Medical Informatics Association) tıp bilişimi eğitimi için ilk uluslararası önerilerinin yayınlanmasıdır.

Haux'un makalesinde söz ettiği bu kilometre taşları, tıp bilişiminin yeni sayılabilecek bir kavram olduğunu ve tıp bilişiminin bir disiplin olarak kabul edilmesinin 2000'li yılları bulduğunu göstermektedir (Haux, 2006). 17 yıllık gelişim dönemi, bu sürede yapılan yeni çalışmalar ve uygulama alanları göz önünde bulundurulduğunda tıp bilişiminin hızla ilerlediği, tıp disiplinindeki birçok alanda kullanım yerinin olduğu ve olmaya devam edeceği söylenebilir.

### **2.3. Veri Madenciliği Kavramı**

Veri madenciliği en geniş kapsamı ile büyük veri kümelerinden anlamlı bilgiler elde edilmesi olarak tanımlanmaktadır. Veri madenciliği her ne kadar son yıllarda yaygın olarak kullanılmaya başlansa da ilk kez 1980'li yıllarda literatüre girmiştir ve 3 temel ana başlık ile birlikte hızla ilerlemeye devam etmiştir. Bu konuların başında istatistik bilimi gelmektedir. Veri madenciliği, istatistik bilimi ile bu kapsamda yapılan standart istatistik, standart dağılım, standart sapma, regresyon analizi ve güven aralıkları gibi veriler arasındaki ilişkileri inceleyen yöntemlerin çalışılması ile gelişmeye başlamıştır (Poyraz, 2012).

Veri madenciliği çalışmalarına büyük katkı sağlayan ikinci başlık yapay zeka (YZ)'dir. Yapay zeka, insan benzeri düşünebilme prensibini temel alarak güçlü bilgisayarlar vasıtası ile çeşitli problemlere çözüm getirilmesini amaçlamaktadır (Poyraz, 2012).

Üçüncü alan ise istatistik ve yapay zekanın kesişimi ile oluşan makine öğrenmesidir. İstatistiksel yöntemler ile yapay zeka yöntemlerini kullanarak ileri düzey yapılan çalışmalar bütünüdür. Makine öğrenmesi verilerin işlenmesi, verilerden anlamlı sonuçlar çıkarılması ve bu sonuçlara yönelik karar alınmasını sağlamaktadır (Poyraz, 2012).

Veri madenciliği bir veri tabanı veya analiz yöntemi değil, farklı uygulama alanlarında kullanılan bilgisayarın keşfi ve gelişen teknoloji ile belirginleşen bir kavramdır.

Teknolojinin gelişim sürecindeki bu hız beraberinde çeşitli problemleri de doğurmuştur. Bu problemlerin başında her geçen gün artan veri miktarı gelmektedir (Gündoğdu, 2007).

Sürekli üretilen bu verinin yönetilmesi ve veriden bilgi elde edilmesi gerekliliği veri madenciliği kavramını ortaya çıkarmıştır. Bu nedenle veri madenciliği kavramı için görülmeyen bir takım örüntülerin, olasılıkların oluşturulması, bunların kullanılması, görsel olarak sunulması ve anlamlı sonuçlar oluşturulması süreçlerinin toplamı olarak söylemek mümkündür (Çam, 2014).

Bununla birlikte veri madenciliği için birçok tanımlama bulunmaktadır. Bu tanımlardan bazıları;

Hand ve arkadaşları veri madenciliğinin, umulmadık ilişkileri bulmak için gözlemsel veri kümelerinin analizi olduğunu ve veriyi anlaşılabilir ve yararlı yeni yollarla özetlediğini söylemektedirler (Hand ve ark., 2001).

Yine Hand ve arkadaşlarına göre veri madenciliği istatistik, veri tabanı teknolojileri, örüntü tanıma, makine öğrenimi ve diğer alanların birleşiminde olan bir ara yüzdür (Hand ve ark., 2001).

Mahdi Al-Azzawı veri madenciliğini, çok sayıda veri tabanından değerli bilgilerin başarıyla keşfedilmesinin temel adımları olarak görmekte ve büyük miktarda veri tabanından gömülü verilerin önemsiz olmayan bir şekilde çıkarılmasını kolaylaştıran bir süreç olarak tanımlamaktadır (Mahdi Al-Azzawı, 2017).

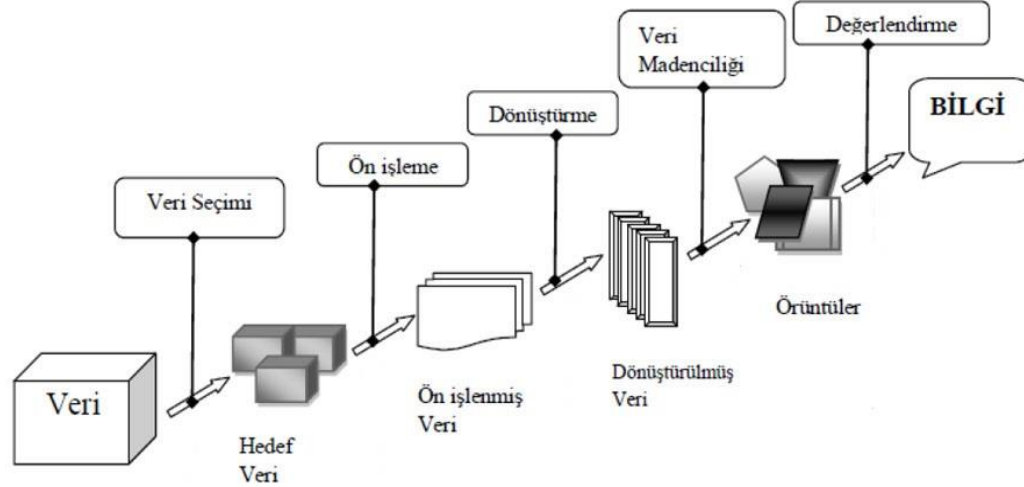
Dünyanın önde gelen araştırma şirketlerinden birisi olan Gartner Group veri madenciliğini, depolanan büyük miktardaki verileri eleyerek anlamlı korelasyonlar, desenler ve trendler keşfetme süreci olarak tanımlamaktadır. Ayrıca veri madenciliğinde, model tanıma teknolojilerinin yanı sıra istatistiksel ve matematiksel tekniklerin kullanıldığını bildirmektedir. (Gartner, 2007).

Lakshmanan ve arkadaşlarına göre veri madenciliği, biriken veri tabanlarından gizli bilgiyi keşfetme imkânı sunan teknoloji olarak tanımlanmaktadır (Lakshmanan ve ark., 2015).

Tüm bu tanımlamalardan yola çıkılarak genel bir tanımlama yapmak gerekirse “veri madenciliği, çeşitli disiplinleri, uygulama alanlarını ve yöntemleri bir araya getirerek, verinin işlenmesi ile büyük veri yığınlarından anlamlı sonuçlar ve bilgiler elde edilmesini sağlayan bir süreçtir” denilebilir.

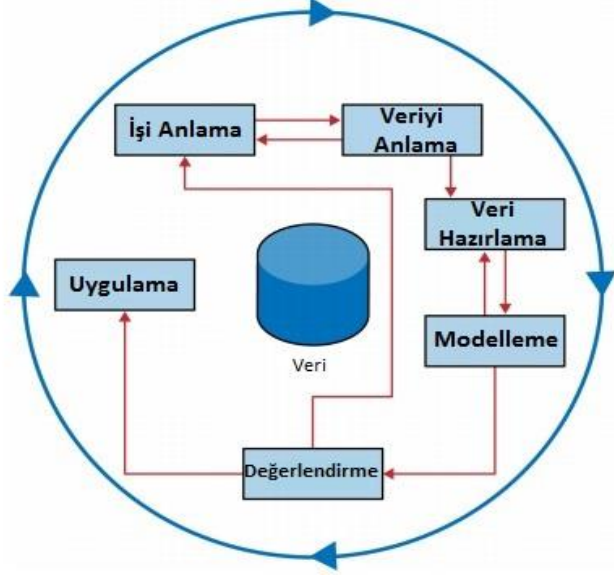
### 2.3.1. Veri Madenciliği Süreci

Bilgi keşfi örtük, önceden bilinmeyen ve potansiyel verilerin önemsiz veriden arındırılması olarak tanımlanmıştır. Bilginin keşfi tüm veri saklama, veri erişim süreçlerini kapsar. Bilgi keşfi veri madenciliğinin merkezinde yer alan bir alan olsa da veri madenciliğinin sadece bir bölümünü oluşturmaktadır (Bramer , 2013). Şekil 2.5.’de bilgi keşfinin süreci gösterilmiştir.



Şekil 2.5. Bilgi keşfi süreci (Bramer , 2013)

1996 yılında veri madenciliği için standart bir model olan CRISP-DM: Cross-Industry Standard Process for Data Mining, Türkçe adı ile “Endüstri Bağımsız Veri Madenciliği Süreci Standardı” geliştirildi. CRISP-DM veri madenciliğini, sadece bilgi keşfini değil iş sorunlarının belirlenmesi, veri toplama ve yönetimi, veri ön işleme, model oluşturma, model değerlendirmesi ve nihayetinde model dağıtımını yoluyla iletilmesini bütün bir süreç olarak görülmesi ve veri madenciliği sürecinin hızlı, verimli ve düşük maliyetli gerçekleştirilebilmesi için geliştirilmiştir. (Larose, 2005). Altı adımdan oluşan CRISP-DM süreci Şekil 2.6’da verilmiştir.



Şekil 2.6. CRISP-DM süreci (Clifton, 2004)

### İş Hedeflerinin Belirlenmesi (Business Understanding)

CRISP-DM'nin süreç anlama, süreç araştırma aşaması olarak da bilinen bu aşama veri madenciliğinin başlangıcı olması ve doğru hedeflerin belirlenmesindeki zorluklar nedeni ile sürecin en önemli adımıdır. Bu aşama çalışma kapsamını oluşturan sürecin ana unsurudur. Proje hedeflerini ve gereksinimlerini iş yaklaşımı açısından tanımlamak ve bu bilgileri, veri madenciliği çalışma ve hedeflere ulaşmak için problem tanımına aktarmak üzerine odaklanmıştır (Çetin, 2011).

İş hedeflerinin belirlenmesi aşamasının diğer bileşenleri personel, yazılım, bilişim kaynakları ve veri kümesi olarak tanımlanmaktadır.

Bu aşamanın sonunda genel olarak aşağıda yer alan başlıklar tamamlanır (Larose, 2005).

- Proje hedeflerini ve gerekliliklerini anlatmak,
- Bu hedefleri ve kısıtlamaları bir veri madenciliği problemi formuna dönüştürmek,
- Bu hedeflere ulaşmak için ön strateji hazırlamak



### **Veriyi Anlama (Data Understanding)**

İkinci aşama olan Veriyi Anlama aşaması, verinin kalitesinin, yeterliliğinin değerlendirilmesi ile ilgili süreçleri kapsar. Bu aşamada kullanılacak veriler ve hedefler arasında bir ilişki oluşturulur ve verilere bakarak hedefler netlik kazanır ve hedeflere bakarak veri tekrar tekrar değerlendirilir (Altun, 2017).

Bu aşamanın sonunda genel olarak aşağıdaki yer alan başlıklar tamamlanır (Larose, 2005).

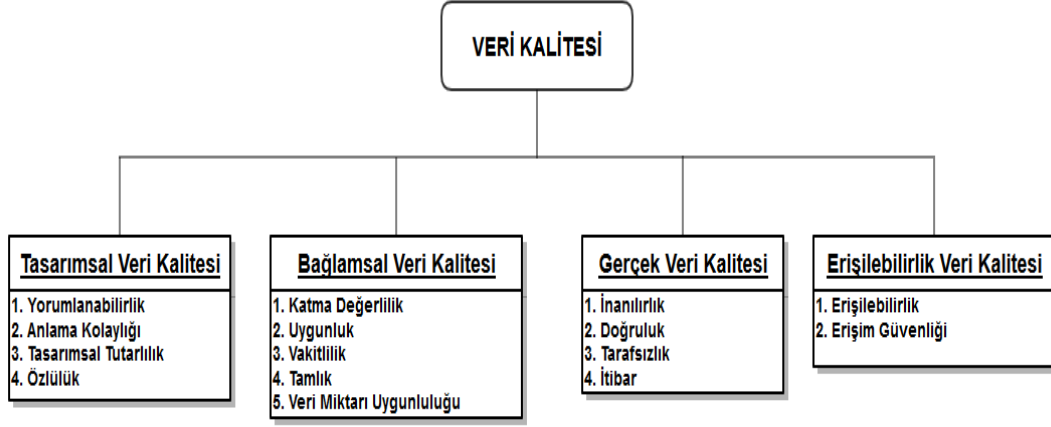
- Veriyi toplamak,
- Verileri anlamak ve keşfetmek için verileri kullanmak, değişkenleri belirlenmek ve tanımlamak,
- Verinin kalitesini değerlendirmek,
- Eğer istenirse uygulanabilir modeller için alt gruplar seçmek

### **Veri Hazırlığı (Data Preparation)**

Bu aşama tamamen veriyi nihai hale getirmeye yöneliktir. Verilerin temizlenmesi, dönüştürülmesi, değişkenlerin belirlenmesi, verilerin bütünleştirilmesi süreçlerini kapsar (Olson ve Delen, 2006). Ham veri/örnekleme, veri kalitesinin korunması, aykırı değerler, kayıp değerler, değişkenleri seçme ve dönüştürme veri hazırlığı aşamasında gerçekleştirilen ana faaliyet başlıklarıdır (Çetin, 2011). (Veri kalitesinin kavramsal bir çerçevesi Şekil 2.7.'de verilmiştir.) Veri içerisindeki farklı kayıtların tek bir formata dönüştürülmesi ve standartlaştırılması bu aşamada gerçekleştirilir.

Bu aşamanın sonunda genel olarak aşağıda yer alan başlıklar tamamlanır (Larose, 2005).

- Diğer tüm aşamalar için ham verilerden nihai veri kümesini oluşturmak,
- Analiz edilmek istenen ve analiz için uygun olan durumları, değişkenleri seçmek,
- Aynı durumu ifade eden değişkenleri dönüştürmek,
- Modellemeye katılması istenmeyen verilerin temizlenmesi



Şekil 2.7. Veri kalitesinin kavramsal bir çerçevesi (Wang ve ark., 1996)

### Modelleme (Modelling)

Tanımlanan problemin çözümü için en doğru modelin bulunması önem arz etmektedir. Bu nedenle uygulama safhasına geçilmeden önce çeşitli modellerin oluşturulması ve denenmesi gerekmektedir. Bu modellere uygun olarak, hazırlanan veri üzerinde veri madenciliği algoritmaları uygulanır. Bu aşamada değişken türlerine uygun algoritmaların tespit edilmesi gerekmektedir (Edelstein, 1999). Modelleme aşaması değerlendirme aşaması ile bütünleşiktir ve her ikisi de tekrar tekrar denenebilmektedir. Burada amaç en doğru ve en iyi sonucu elde etmektedir (Akküçük, 2015). Genel olarak Modelleme Tekniğinin Seçimi, Modelin Yapımı, Modelin Değerlendirilmesi ana başlıklarından oluşur (Çetin, 2011).

Bu aşamanın sonunda genel olarak aşağıda yer alan başlıklar tamamlanır (Larose, 2005).

- Uygun modelleme tekniklerini seçmek ve uygulamak,
- Sonuçları optimize etmek için modelleri geliştirmek,
- Aynı problemin çözümü için farklı teknikleri kullanabilmek,
- Gerekirse veri madenciliği tekniklerini uyumlu hale getirmek için veri hazırlama aşamasına dönmek.

## **Değerlendirme (Evaluation)**

Bu aşamada var olan modelin değerlendirmesi ve planlanan iş hedefleri ile uyumlu olup olmadığını kontrol etmek için testler yapılır. Sonuçları değerlendirmek için veriler ile model test edilir. Ayrıca projenin geldiği noktanın yeterliliği değerlendirilir (Argüden ve Erşahin, 2008).

Değerlendirme aşamasında iki önemli husus vardır. Birincisi, sonuçların bir araya getirilmesidir. Çıkarılan ham verilerin analizi çok karmaşıktır ve iş hedeflerine göre işlenip bir araya getirilmesi gerekmektedir. Bu işlemin gerçekleştirilmesi önceki aşamaların (iş hedeflerini belirleme, veriyi anlama, veriyi hazırlama, modelleme) sonuçlarının doğru bir şekilde birleştirilmesine bağlıdır. Diğeri ise, veri madenciliği sonuçlarını göstermek için kullanılacak görselleştirme aracını seçmektir (Çetin, 2011).

Bütün modelin projenin iş hedeflerine bakılarak değerlendirilmesi çok önemlidir. Modeli değerlendirmenin yollarından birisi gerçek dünyada test etmektir. Tüm veri kümesinden bir veri örneği seçilerek modelin doğruluğu test edilmeli ve sonuçların tahmin edilen sonuçlara yakınlığı gözlemlenmelidir (Çetin, 2011).

Bu aşamanın sonunda genel olarak aşağıda yer alan başlıklar tamamlanır (Larose, 2005).

- Uygulama aşamasından önce etkinlik ve kalite için modelleme aşamasında geliştirilen bir veya daha fazla modeli değerlendirmek,
- Modelin ilk aşamada belirlenen hedeflere ulaşıp ulaşmadığını belirlemek,
- Araştırma sorununun önemli bir yönünün yeterince açıklanıp açıklanmadığını belirlemek,
- Veri madenciliği sonuçlarının kullanımı ile ilgili karar vermek.

## **Uygulama/ Dağıtım (Deployment)**

Veri madenciliği sürecinin son aşaması olan Uygulama/Dağıtım, önerilen çözümün gerçek hayatta çalışan işletim ortamlarına yerleştirilmesi ile sunulmasıdır (Antonio ve ark., 2017). Dağıtım, iyileştirmeler yapmak için yeni bilgileri kullanma sürecidir. Uygulama/ Dağıtım, Planların Dağıtılması, Planların İzlenmesi ve Sürdürülmesi, Final

Raporunun Hazırlanması, Final Raporunun Sunulması ve Projenin Gözden Geçirilmesi aşamalarından oluşur (Chapman ve ark., 2000).

Bu aşamanın sonunda genel olarak aşağıda yer alan başlıklar tamamlanır (Larose, 2005).

- Oluşturulan modelleri kullanmak,
- Rapor gibi basit bir dağıtım örneği oluşturmak,
- Final raporu gibi kompleks bir dağıtım örneğini oluşturmak,
- Başka departmanlar için paralel bir veri madenciliği sürecini uygulamak

### **2.3.2. Veri Madenciliği Modelleri**

Veri madenciliği çalışmalarının başarılı olabilmesi için en önemli şeylerden bir tanesi de doğru algoritmanın seçilmesidir. Algoritma seçiminde verinin türünün veya yapısının ne olduğu büyük önem taşımaktadır. Birçok çalışmada veri madenciliği model ve algoritmaları farklı başlıklar altında ele alınmıştır. En sık gösterim olarak kullanılan başlıklara göre veri madenciliği modelleri işlevlerine göre 3'e ayrılmaktadır. Bu başlıklar;

- Sınıflandırma,
- Kümeleme,
- Birliktelik Kuralları'dır.

#### **Sınıflandırma ve Kestirim**

Sınıflandırma ve kestirim veri madenciliği modellerinden birisidir fakat birbirlerinden ayrılan yönleri olduğundan uygulamada farklılık göstermektedirler. Sınıflandırma ve kestirim algoritmaları da istatistik, uzman sistem ve makine öğrenmesi yöntemlerine dayanmaktadır (Gündoğdu, 2007).

Sınıflandırma, esas olarak belirli bir veri kümesini analiz etmek için kullanılan ve onun her örneğini alarak, bu örneği belirli bir sınıfa, en az hata olacak şekilde atayan veri madenciliği modellerinden birisidir. Veri kümesindeki önemli veri sınıflarını doğru bir şekilde tanımlayan modeller çıkarmak için kullanılır. Sınıflandırma iki adımlı bir işlemdir. İlk adımda, veri seti üzerinde sınıflandırma algoritmasını uygulayarak model oluşturulur. İkinci adımda ise elde edilen model, önceden eğitilmiş performans ve doğruluğu ölçmek için önceden tanımlanmış bir test veri setinde test edilir (Nikam, 2015).

Kestirim ile sınıflandırma arasındaki fark kestirim algoritmalarının farklı prensip ile çalışmasıdır. Kestirim veya tahmin yürütme tahmin edilen sınıfa göre yapılır ve dolayısıyla doğruluğu “bekle-gör” prensibini kullanmaktadır. Sınıflandırma ve kestirim yöntemleri Doğruluk kestirimi, Hız, Sağlamlık, Ölçülebilirlik ve Açıklanabilirlik kriterlerine göre karşılaştırılır (Gündoğdu, 2007).

Sınıflandırma ve Kestirim modellerinde başlıca kullanılan teknikler (Diler, 2016);

- Karar Ağaçları (Decision Tree)
- Yapay Sinir Ağları (Artificial Neural Networks)
- Genetik Algoritmalar (Genetic Algorithms)
- K-En Yakın Komşu (K-Nearest Neighbor)
- Regresyon (Regression)
- Bayesyen Sınıflandırma (Bayesian Classification)
- Vaka Tabanlı Akıl Yürütme (Case Based Reasoning)
- Kaba Küme Yaklaşımı (Rough Set Approach)
- Bulanık Küme Yaklaşımı (Fuzzy Set Approaches)

### **Kümeleme**

Kümeleme işlemi sınıflandırma modelinde olduğu gibi verileri gruptandırmaya çalışmaktadır. Aralarındaki fark sınıflandırma algoritmalarında verilerin sınıfları önceden bilinmekte fakat kümeleme algoritmalarında veri sınıfları önceden bilinmemektedir (Çataloluk, 2012). Küme benzer nesnelerin oluşturduğu bir gruptur ve kümelemenin temel prensibi küme üyeleri arasındaki uzaklığın minimum, kümeler arası uzaklığın maksimum olmasıdır. Kümeleme algoritmasının kalitesi bu prensibin çalışması ile doğru orantılıdır (Han ve Kamber, 2006).

Veri madenciliğinde kümeleme modelinin kullanılması için bazı gereksinimlerin sağlanmış olması gerekmektedir. Bunlar (Han ve Kamber, 2006);

1. Ölçeklendirebilme: Çok performans gerektiren büyük verilerin daha küçük boyutlu verilere ölçeklendirilmesi.

2. Değişik Nesne Tiplerine Göre Çalışabilme: Algoritmanın sayısal, sayısal olmayan ve ikili veri tiplerine uygun olarak çalışabilmesi.
3. Farklı Tipteki Nesneleri Ayırabilme: Farklı tip ve farklı yoğunluktaki nesneleri ayırt edebilecek algoritmalara uygun olarak çalışılabilmesi.
4. En Az Miktarda Alan Bilgisi Gerektirmesi: Algoritmanın parametrelere karşı hassaslığının ve bağımlılığının az olması.
5. Çöp Veri Ayıklayabilme: Algoritmanın kötü sonuçlar vermemesi için veri üzerindeki eksik, ayırık, tanımlanmamış, yanlış verilerin ayıklanması.
6. Algoritma, Verilen Parametrelerin Sırasına Duyarsız Olması: Algoritmaya girilen verideki sıra değişikliklerinden algoritmanın ve algoritma sonucunun etkilenmemesi.
7. Yüksek Boyutluluk: Algoritmanın 2, 3 veya daha fazla boyut üzerinde çalışabilmesi.
8. Kısıtlama Bazlı Kümeleme: Algoritma sonucuna etki edecek verilerin filtrelenebilir veya kısıtlanabilir olması.

Kümeleme modeli genel olarak ikiye ayrılır. Bunlar hiyerarşik ve hiyerarşik olmayan kümeleme algoritmalarıdır (Silahtar, 2013).

**Hiyerarşik Yöntemler:** Hiyerarşik kümeleme teknikleri bir kümeyi bir başkası ile ardı ardına birleştirme sürecidir ve bir küme farklı bir küme ile birleştikten sonra ilerleyen adımlarda kesinlikle ayrılamaz, üzerine eklenerek devam eder (Ümit, 1997). Hiyerarşik kümeleme algoritmalarının sonuçlarının ağaç diyagramı ile gösterilmesine dendogram denir (Lorr, 1983).

- Toplaşım Kümeleme Algoritmaları,
- Bölünür Kümeleme Algoritmaları.

**Hiyerarşik Olmayan Yöntemler:** Hiyerarşik Yöntemlere göre çok daha büyük veri setlerine uygulanan bu yöntemde önceden belirlenmiş veya araştırmacının anlamlı şekilde karar verdiği küme sayıları vardır (Çelik, 2013). Bu yöntemde birimlerin kümelere ayrılması rastgele olabilir ve belirlenen küme sayılarına ve küme kriterlerine göre küme üyelerinin atamaları yapılır (Ada ve ark., 2017).

- Yer Değiştirmeli Algoritmalar,
- Olasılıksal Algoritmalar,
- K-Medoid Yöntemi,
- K-Means Yöntemi,
- Yoğunluğa Dayalı Algoritmalar,
- Yoğunluğa Dayalı Bağlantılı Algoritmalar,
- Yoğunluk Fonksiyonlu Algoritmalar.

### **Birliktelik Kuralları**

Birliktelik modeli, kategorik değişkenler arasındaki birliktelikleri veya ilişkileri tespit etmek amacı ile kullanılır. Birliktelik kuralları veri seti üzerindeki gizli kalan ve ilk aşamada tespit edilemeyen örüntülerin tespit edilmesine izin vermektedir. Aynı zamanda birliktelik kuralları farklı veri türlerinde ve özelliklerinde herhangi bir bilgiye ihtiyaç duymaksızın oluşturabilmektedir (Nisbet ve ark., 2009).

Birliktelik kuralları kullanıldığında küçük bir veri setinden çok fazla sayıda kural çıkarılabilir. Bu nedenle kural sayısını aza indirmek, geçerli ve güvenilir ilişkiler üzerinde yoğunlaşmak için iki kavram ortaya çıkmıştır. Bunlar Destek (Support) Sınırı ve Güven (Confidence) Sınırı kavramlarıdır. Destek Sınırı veri seti içindeki veri kombinasyonun oranını, Güven Sınırı ise bu kombinasyon ile birlikte diğer verilen birlikte geçme olasılığını göstermektedir (Hatipoğlu, 2013).

Birliktelik kuralları temel olarak iki aşamadan oluşmaktadır (Çataloluk, 2012).

1. Sık Geçen Nesne Kümelerinin Belirlenmesi; Minimum destek şartını sağlayan nesne kümelerinin tespit edilmesi.

2. Sık Geçen Nesne Kümelerinden Güçlü İlişki Kurallarının Oluşturulması; Sık geçen nesne kümelerinin yukarıda bahsedilen destek ve güven sınırını sağlaması ve bunun sonucu olarak güçlü birliktelik kurallarının oluşturulması.

### **2.3.3. Veri Madenciliğinin Uygulama ve Kullanım Alanları**

Veri madenciliği için ihtiyaç duyulan tek şey büyük miktarda verinin var olmasıdır. Bu nedenle uygulama ve kullanım alanları bakımından veri madenciliği oldukça geniş bir alana yayılmış durumdadır.

“Birçok alana hitap eden bu disiplinin başlıca kullanım alanları (Telcioğlu, 2007; Çataloluk, 2012; Köktürk ve ark., 2009);

#### **1. Pazarlama**

- Müşteri segmentasyonu,
- Müşterilerin demografik özellikleri arasındaki bağlantıların kurulması,
- Çeşitli pazarlama kampanyaları,
- Mevcut müşterilerin elde tutulması için geliştirilecek pazarlama stratejilerinin oluşturulması,
- Pazar sepeti analizi,
- Çapraz satış analizleri,
- Müşteri değerlemesi,
- Müşteri ilişkileri yönetimi,
- Çeşitli müşteri analizleri,
- Satış tahminleri.

#### **2. Bankacılık**

- Farklı finansal göstergeler arasındaki gizli korelasyonların bulunması,
- Kredi kartı dolandırıcılıklarının tespiti,
- Müşteri segmentasyonu,
- Kredi taleplerinin değerlendirilmesi,



- Usulsüzlük tespiti,
- Risk analizleri,
- Müşteri satın alma sıklığı,
- Risk yönetimi.

### 3. Sigortacılık

- Yeni poliçe talep edecek müşterilerin tahmin edilmesi,
- Sigorta dolandırıcılıklarının tespiti,
- Riskli müşteri tipinin belirlenmesi.

### 4. Perakendecilik

- Satış noktası veri analizleri,
- Alış-veriş sepeti analizleri,
- Tedarik ve mağaza yerleşim optimizasyonu.

### 5. Borsa

- Hisse senedi fiyat tahmini,
- Genel piyasa analizleri,
- Alım-satım stratejilerinin optimizasyonu.

### 6. Telekomünikasyon

- Kalite ve iyileştirme analizleri,
- Hisse tespitleri,
- Hatların yoğunluk tahminleri.

### 7. Endüstri

- Kalite kontrol analizleri,
- Lojistik,
- Üretim süreçlerinin optimizasyonu.

## 8. Bilim ve Mühendislik

- Ampirik veriler üzerinde modeller kurarak bilimsel ve teknik problemlerin çözümlenmesi.

## 10. Sağlık ve İlaç

- Test sonuçlarının tahmini,
- Ürün geliştirme,
- Tıbbi teşhis,
- Tedavi sürecinin belirlenmesi ”.

### 2.3.4. Tıp ve Sağlık Hizmetlerinde Veri Madenciliği Çalışmaları

Veri madenciliği çalışmalarında istatistiksel yaklaşımlar kullanılarak büyük veri kümelerinde gizli bilgiler keşfedilebilir. Bu nedenle klasik yöntemlere kıyasla birçok avantajı vardır. Bu avantajlar birçok sektörde olduğu gibi tıp ve sağlık alanında da fazlasıyla kullanılmaktadır. Sağlık hizmetlerinde veri madenciliği uygulamaları değerlendirildiğinde bu çalışmalar geniş kategoriler halinde gruplandırılabilir (Durairaj ve Ranjani, 2013).

**Tedavi etkinliği:** Veri madenciliği uygulamaları, tıbbi tedavilerin etkinliğini değerlendirmek için geliştirilebilir. Veri madenciliği neden, semptom ve tedavilerin seyrini karşılaştırarak eylemin hangi yolda etkili olduğunu kanıtlayan bir analiz sunabilir.

**Sağlık Yönetimi:** Veri madenciliği uygulamaları, kronik hastalık durumlarını ve yüksek riskli hastaları daha iyi tanımlamak ve izlemek, uygun müdahaleler tasarlamak ve sağlık yönetimine yardımcı olmak ve hastaneye kabul edilen kişi sayılarını azaltmak için geliştirilebilir.

**Dolandırıcılık ve İstismar:** Tıpta veri madenciliği uygulamaları ile dolandırıcılık ve istismar uygulamaları, uygun olmayan reçeteler, tavsiyeler, hileli sigorta ve tıbbi talepler gibi konularda farklılıkların ortaya konulması ve risklerin azaltılması konusunda çalışma yapılabilir.

**İlaç endüstrisi:** Veri madenciliği teknolojisi, ilaç firmalarının stoklarını yönetmesine yardımcı olmak ve yeni ürün ve hizmetler geliştirmek için kullanılmaktadır. İlaç verilerinde gizlenmiş bilginin derin bir anlayışı, bir firmanın rekabetçi konumu ve örgütsel karar verme mekanizması için veri madenciliği büyük önem taşır.

**Hastane Yönetimi:** Modern hastaneler dâhil kuruluşlar büyük miktarda veri üretebilir ve toplayabilir. Veri madenciliğinin küresel hastane faaliyetlerinin zamansal davranışının görselleştirildiği bir hastane bilgi sisteminde saklanan verilere uygulanması yaygın olarak kullanılabilir. Bu noktada hastane yönetiminin 3 katmanı vardır.

- Hastane yönetimi hizmetleri
- Tıbbi personel için hizmetler
- Hastalar için hizmetler

**Sistem Biyolojisi:** Biyolojik veri tabanları, genellikle zengin ilişkisel yapıya sahip çok çeşitli veri türleri içerir. Bununla birlikte, çok yönlü veri madenciliği teknikleri biyolojik verilere sıklıkla uygulanmaktadır. Ayrıca sistem biyolojisi açısından, veri madenciliği teknikleri ile genetik değişiklikler konusunda çalışma yapılabilir.

Bu çalışma alanlarının yanı sıra tıbbi veriler diğer veri türleri ile farklılık göstermekte ve diğer verilerde olmayan bir takım özellikler taşımaktadır. Tıp alanındaki verilerin standart olmayışı, tıbbi alanda kullanılan terimlerin karmaşık olması gibi nedenlerden dolayı veri tabanlarının oluşturulmasında yaşanan zorluklar, tıbbi verideki etik ve yasal sorunlar ve hasta mahremiyeti bu alanda yapılan çalışmalarını zorlaştırmaktadır (Köktürk ve ark., 2009). Tüm bu zorluklara rağmen tıbbi veri tabanlarında veri madenciliği çalışmaları giderek artmaktadır. Bu anlamda veri madenciliği alanında yapılan çalışmalardan bazıları aşağıda verilmiştir:

Veri madenciliği çalışmaları hastalık için tahminleme, sınıflandırma, karar destek sistemleri gibi alanların yanı sıra sağlık kalitesinin iyileştirilmesinde de yaygın olarak kullanılmaktadır. Chae ve arkadaşlarının 2003 yılında yaptığı çalışma bu konu ile ilgili örnek olarak verilebilir. Kalite geliştirme stratejilerinin geliştirilmesi için veri madenciliği kullanarak sağlık kalitesi göstergelerinin bir analizini sunan çalışma 1 Aralık 2000 ve 31

Ocak 2001 döneminde hastaneden taburcu edilen 8405 hasta için veri madenciliği tekniklerinden karar ağacı yöntemi kullanılarak yatan hastalarda mortaliteyi etkileyen önemli faktörlerin tespiti yapılmıştır. Bu sistemi analiz etmek ve izlemek için bir karar destek sistemi (KDS) geliştirilmiştir. Çalışma sonucu, karar ağaçlarının ve veri madenciliğinin kullanılması ile hastanede yatan ölüm vakalarını bulmada ve kaliteyi etkileyen faktörleri tespit etmede oldukça başarılı olduklarını göstermiştir (Chae ve ark., 2003).

Obenshain 2004 yılındaki Veri Madenciliği Tekniklerinin Sağlık Verilerinde Uygulanması isimli makalesinde sağlık alanında uygulanmış olan iki başarılı örnek vermiştir. Bunlardan ilki Hastane Enfeksiyon Kontrolü çalışmasıdır. Her yıl Amerika Birleşik Devletlerinde 2 milyona ulaşan enfeksiyon sayısı nedeni ile salgınların erken tanınması ve ortaya çıkan dirençlerin gözlenmesi gerekliliğinden ortaya çıkmıştır. Yapılan çalışma yüksek riskli hastaları, uzman sistemleri ve muhtemel vakaları belirleme ve önceden tanımlanmış olayların ortaya çıkmasındaki sapmaları tespit etmeye odaklanmıştır. Enfeksiyon kontrol verilerindeki yeni ve ilginç örüntüleri tanımlamak için veri madenciliği tekniklerini kullanan bu sistem Alabama Üniversitesi'nde uygulanmıştır. Laboratuvar bilgi yönetim sistemlerinden elde edilen kültür ve hasta bakım verileri ilişki belirleme yöntemlerini kullanan 15 sistem ile denenmiştir ve sonuçlar uzmanlar tarafından gözden geçirilmiştir. Sonuç olarak sistem geliştiriciler tarafından, veri madenciliği sistemi ile enfeksiyon denetimini geliştirme sonucu geleneksel enfeksiyon kontrol gözetiminden daha hassas bulunmuştur (Obenshain, 2004).

American Healthways, diyabet hastalarının yaşam kalitesini arttırmak ve tedavi maliyetini düşürmek için hastanelere ve sağlık planlarına şeker hastalığı yönetim hizmetleri sistemi sunmuştur. Bu sistem için tahmini modelleme teknolojisi kullanılmıştır. Sistem kısa süreli sağlık sorunları olasılığını öngörmek için kapsamlı hasta bilgileri birleştirmekte ve araştırmaktadır. Sonuçlara bakıldığında ise, oluşturulan güçlü bir veri madenciliği ve model oluşturma çözümü, yüksek riskli bir duruma doğru ilerleyen hastaları tanımlamıştır. Bu bilgi ile hemşire bakım koordinatörlerine yüksek riskli hastaların belirlenmesinde ve böylece hastaların sağlık hizmetlerinin kalitesini iyileştirmek ve

gelecekte sađlık problemlerini önlemek için ilk adımlarının atılabileceđi belirtilmiřtir (Obenshain, 2004).

Dolce ve arkadaşlarının 2008 yılındaki erken prognoz hastalarına yönelik veri madenciliđi çalışmasında karar ağaçları kullanılmıştır. Nisan 1998-Mart 2006 yılları arasında S.Anna Enstitüsünün yoğun bakım ünitesine kabul edilen dört yüz elli hasta geriye dönük olarak değerlendirilmiştir. Şiddetli omurga kırıkları, spinal kord hasarı veya başka türlü yoğun tedavi veya ameliyat gerektiren beyin hasarından 4 hafta içinde bilinçlenmeyi iyileştiren hastalar yanlış tanıları en aza indirmek için çalışmadan hariç tutulmuştur ve 265 erkek ve 68 kadın, üç yüz otuz üç hasta verisinde çalışılmıştır. Hastaların kayıtlarından elde edilen mantıklı veriler, veri madenciliđi prosedürlerini uygulamadan önce veri ön işleme işlemleri gerçekleştirilmiştir. Veriler, WEKA programında Sınıflama ve Regresyon Ağacı yöntemi ile işlenmiştir. Çalışma sonucunda ise Sınıflama ve Regresyon modelinin vejetatif haldeki hastaların sonuçlarının tahmini için yeterli olduđu kanıtlanmıştır (Giuliano ve ark., 2009).

Veri madenciliđi çalışmalarında mevcut yazılım ve algoritmaların kullanımına ek olarak bu alanda yazılım geliştirilmesine yönelik çalışmalar da mevcuttur. Evangelista ve arkadaşları tarafından 2010 yılında bu amaçla bir yazılım geliştirilmiştir. Yapılan çalışmada Diyabet veya kanser gibi ciddi hastalıklara yol açan kronik enflamasyon ilerlemesinin ana oyuncularını olan sitokinlerin sağlıklı kişiler ve hastalar üzerinde profillerinin, klinik ve laboratuvar verileriyle birlikte kullanılarak, hasta profillerini istatistiksel araçlarla karşılařtırmak ve hastalık ilerlemelerini değerlendirmek için bir Klinik Veri Madenciliđi Yazılımını geliştirilmiştir (Evangelista ve ark., 2010).

2010 yılında Koyuncugil ve Özgülbaş tarafından tıbbi yönünün de büyük önem taşıdığı veri madenciliđine katkı sağlayacak bir çalışma gerçekleştirilmiştir. Çalışma organ transplantasyonunda veri madenciliđine dayalı bir donör araştırma ve eşleştirme sistemidir. Çalışma organ transplantasyonunda donör araması ve eşleştirme sistemi için bir sistem geliştirilmesi yoluyla hayat kurtarmak için bir katkı sunmayı amaçlamaktadır. Sistemde veri madenciliđi yöntemi olarak A-Priori algoritması kullanılmıştır. Çalışma sonucunda gerçek veriler kullanılarak sistemde iyileştirmeler yapılabileceđiyle birlikte veri madenciliđinin, klinik arařtırmalarda alternatifleri değerlendirmek için teşhis için

yardımcı bir araç olarak kullanılabileceğini ayrıca veri madenciliğine dayalı erken uyarı ve gözetim sistemlerinin, iletişimsel hastalıkların tespiti ve kontrolü için de geliştirilebileceği belirtilmiştir (Koyuncugil ve Özgülbaş, 2010).

Bir başka çalışma ise Jackson ve arkadaşları tarafından yapılmıştır. Jackson ve arkadaşları 2011 yılında üriner sistem/idrar yolları enfeksiyonu hastaları için ampirik tedaviyi geliştirici veri madenciliği çalışması yapmışlardır. Çalışmada Northwestern Üniversitesi'nin Feinberg Tıp Fakültesi Kuzeybatı poliklinik ve Üroloji ve Dahiliye klinikleri ve Acil Servis Departmanında 2005-2009 yılları arasında olumlu bir idrar kültürü ile ilişkili 3.308 ziyaretten alınan elektronik tıbbi kayıtlar sorgulanmıştır. Veri madenciliği tekniklerini kullanarak idrar yolu enfeksiyonlarının ampirik tedavisi için algoritmalar oluşturulmuştur ve veri madenciliği ile elde edilen teorik sonuçlar geleneksel terapiden elde edilen sonuçlar ile karşılaştırılmıştır. Tedavi zamana, ilaca ve bölgeye bağlı olsa da özellikle kadın hastalar için geleneksel terapiye kıyasla teorik sonuçları önemli ölçüde geliştirdiği beyan edilmiştir. Çalışma sonuçlarına göre önerilen algoritma kullanıldığında, idrar yolları enfeksiyonlu hastalar, geleneksel terapi uygulamasına kıyasla % 11.3'lük genel bir iyileşme göstermişlerdir (Jackson ve ark., 2011).

Serkan Güllüoğlu tarafından yapılan başka bir çalışmada ise kanser teşhisine yönelik bir ön çalışma yapılmıştır. 2011 yılında gerçekleşen bu çalışmada en sık rastlanan kanser türlerinden olan bağırsak, meme ve akciğer kanseri seçilmiştir. Analizler Matlab Programı kullanılarak yapılmıştır. İstanbul Arel Üniversitesi Hastanesi'nin Radyoloji Onkoloji bölümü arşivinden elde edilen veriler sonuçlarla ilişkilendirilerek yapay sinir ağının eğitiminde ve test setlerinde kullanılmıştır. Ağın başarı oranı istenen seviyeye ulaşıncaya kadar eğitim devam ettirilmiştir. Elde edilen sonuçlara göre veri madenciliğinin sağlık sektöründe konunun uzmanı kişilerce kullanılmasının hizmetlerin etkin sunumu, bilgiye erişim ve kaynakların daha verimli kullanımı açısından önerilen bir teknik olarak sunulmuştur (Güllüoğlu, 2011).

Tang ve arkadaşları tarafından 2012 yılında yayınlanan çalışmada Veri madenciliği teknikleriyle yaygın anksiyete bozukluğunun sendrom analizi üzerine çalışılmıştır. 1 Ağustos 2009'dan 31 Temmuz 2010'a kadar, Pekin'in 10 hastanesinde anksiyete bozukluğu olan 705 hasta bir yılda araştırılmıştır. Anksiyete bozukluğu olan hastaların

sendromunu analiz etmek için Bayes algoritmaları ve küme analizi gibi veri madenciliği teknikleri kullanılmıştır. Yinelenen sendromlar için tarama yaptıktan ve uzmanların tecrübeleriyle geleneksel Çin tıbbı teorisini birleştirdikten sonra anksiyete bozukluğunun altı sendromu tanımlanmıştır. Yapılan bu çalışmada elde edilen sonuçlara göre Bayes algoritmalarının ve küme analizi gibi veri madenciliği tekniklerinin sendrom modelleri oluşturma ve sendromları analiz etme potansiyeline sahip olduğu ve bu nedenle sendrom farklılaşmasının araştırılması için uygun yöntemler olduğu tespit edilmiştir ( Tang ve ark., 2012).

Bilge ve arkadaşlarının 2013 yılında yapmış olduğu çalışma da özellikle veri madenciliği yöntemlerinin kullanılarak hastalığın tespit edilmesine yönelik kullanılacak kurallar geliştirilmesine ve ilişkilerin araştırılmasına katkı sunmaktadır. Yapılan çalışmada inme için önemli bir faktör olan karotis (şah damarı) darlığına etki eden çeşitli faktörlerin incelenmesi üzerine çalışılmıştır. Çalışmada genetik algoritma, lojistik regresyon ve ki kare testleri uygulanmış ve test sonuçları karşılaştırılmıştır. Sonuçta ise koroner arter hastalığının ve hipertansiyonun, koroner karotis darlığının öngörülmesinde önemli bir gösterge olduğu tespit edilmiştir (Bilge ve ark., 2013).

Veri madenciliği alanında geliştirilen başka bir yazılım da Cheng ve arkadaşlarına aittir. Yazılım bugüne kadar 20 kanser türü, 33 tümör tipi ve 10.000 vaka'nın veri kayıtlarının bulunduğu Kanser Genomu Atlası (TCGA) verilerini analiz etmek amacı ile geliştirilmiştir. Genom verilerini görselleştirmek, analiz etmek ve yorumlamak için biyoenformatik eğitim almayan temel araştırmacılara ve klinisyenlere yardımcı olacak bir araç olan bu yazılım, kanser genomu veri madenciliği için son derece güçlü bir kaynak oluşturmuştur (Cheng ve ark., 2015).

Arslan, Çolak ve Sarihan tarafından 2016 yılında, dünyadaki yüksek mortalite ile ilişkili ve en önemli halk sağlığı sorunları arasında yer alan İskemik İnmenin tahminine yönelik farklı veri madenciliği yaklaşımları değerlendirilmiştir. Malatya, İnönü Üniversitesi Tıp Fakültesi, Acil Tıp Anabilim Dalı, Turgut Özal Tıp Merkezi'nde, 80 iskemik inmeli hastayı (grup I) ve 112 sağlıklı bireyi (grup II) kapsayan çalışma yürütülmüştür. Destek Vektör Makinesi, Stokastik Gradyen Artırma ve Lojistik Regresyon Veri Madenciliği yöntemleri kullanılmıştır. İskemik İnmenin sınıflandırılması için çalışılan tüm modellerin

performans deęerleri olduka yksektir. Yapılan alıřma ve daha nceki alıřmalar neticesinde bilgisayar destekli yaklařımların iskemik inmenin ngrlmesinde etkili bir ara olduğunu ve veri kmelerindeki gizli iliřkileri arařtırdığı ortaya koyulmuřtur (Arslan ve ark., 2016).

Zhang ve arkadařları tarafından 2016 yılında yazılan bir dięer makalede klinik tıpta byk veri madencilięinin uygulamaları, teorileri ve teknolojileri gzden geirilmiřtir. alıřmada, PubMed aracılıęıyla 1975'ten 2015'e kadar byk veri madencilięi teorileri ve teknolojilerinin klinik tıpta somut uygulamalarıyla ilgili olarak İngilizce veya ince yayınlanan literatr taranmıřtır. Yapılan alıřma sonucunda bulanık teori, kaba kme teorisi, bulut teorisi, Dempster-Shafer teorisi, yapay sinir aęı, genetik algoritma, veri madencilięi teorisi, bayes aęları, karar aęı, desen tanıma, performans hesaplamaları ve istatistiksel analiz tekniklerinin yoęun olarak hastalık riski deęerlendirmesi, klinik karar verme desteęi, hastalık geliřiminin tahmini, rasyonel ila kullanım kılavuzluęu, tıbbi ynetim ve kanıta dayalı tıp alanlarında kullanıldığı tespit edilmiřtir (Zhang ve ark., 2016).

Senio ve arkadařları tarafından 2018 yılında yapılan alıřmada depresyonun astımın řiddetini etkileyip etkilemedięinin belirlenmesi iin kmeleme analizi kullanılmıřtır. 2.273 hastanın 128'i depresif belirtileri pozitif olarak deęerlendirilen hasta veri setinde yař, hastalık sresi, vcut kitle indeksi (VKİ) ve kullanılan ila bilgileri de yer almaktadır. Yapılan kmeleme analizi sonucunda depresif belirtisi pozitif olan hastaların yaklařık yarısı ciddi astım hastaları kmesini oluřturmuřtur. (Seino ve ark., 2018)

Phan ve arkadařlarının Hipoksik-İskemik Beyin Hasarında Sonuların Sınıflandırılmasında Karar Aęacı Analizinin Aıklayıcı Kullanımı isimli makalesinde alıřmamıza benzer řekilde karar aęacı yntemi kullanılmıřtır. Karar aęacının uygulanması iin CART algoritması ile alıřma yapılmıřtır. 41 kadın hastaya ait veri seti zerinde yapılan alıřmada, karar aęacı analizinin koma sonucu sınıflandırma yapmak iin yardımcı bir yntem olduęu tespit edilmiřtir. (Phan ve ark., 2018)



### 3. GEREÇ VE YÖNTEM

#### 3.1. Araştırmanın Amacı

Yapılan tez çalışması ile: Amiyotrofik Lateral Skleroz (ALS) tanılı olgulardan elde edilmiş olan genetik, klinik ve demografik verilerin, hiyerarşik kümeleme ve karar ağaçları veri madenciliği yöntemleri kullanılarak ilişkilendirilmesi ve bu yöntemler ile klinik ve demografik özelliklerden yararlanılarak genetik mutasyonların gruplandırılması amaçlanmıştır.

#### 3.2. Araştırma Soruları

ALS hastalığında klinik veriler ile gen verilerini ilişkilendirmek mümkün müdür?

ALS hastalığında klinik veriler vasıtası ile gen farklılıklarını tahmin etmek mümkün müdür?

ALS hastalığında genetik, klinik ve demografik verilerden anlamlı ilişkiler tespit etmek amacı ile yeni bir metot geliştirilebilir mi?

#### 3.3. Araştırmanın Evren ve Örneklemi

Araştırmanın evrenini, Türkiye'deki Amiyotrofik Lateral Skleroz teşhisi konulmuş hastalar oluşturmaktadır.

Araştırmanın örnekleme, Akdeniz Üniversitesi Hastanesi Nöroloji Polikliniği'ne çeşitli şikâyetler ile başvuru yaparak muayene sonrası Amiyotrofik Lateral Skleroz teşhisi konulan veya daha önceden Amiyotrofik Lateral Skleroz teşhisi konulmuş ve tedavi için başvuru yapan hastalardan oluşmaktadır.

#### 3.4. Araştırmanın Etik Yönü

Tez çalışması doğrultusunda 2016 yılında çalışmasına başlanmış “Amiyotrofik Lateral Sklerozu anlamak için ontolojik temelli ve Web'e dayalı veritabanı” ( “Ontology-based Web Database for Understanding Amyotrophic Lateral Sclerosis” ) projesi kapsamında Tıp Fakültesi Nöroloji Anabilim Dalı'nda devam eden çalışmanın klinik ve demografik verileri kullanılmıştır. Devam eden proje için Akdeniz Üniversitesi Tıp Fakültesi

Girişimsel Bilimsel Çalışmalar Etik Kuruluna başvurularak 04.05.2015 tarih ve 248 sayılı karar ile onay alınmıştır.

Bunun ile birlikte proje kapsamındaki verilerin tez kapsamında kullanılması için proje yürütücüsünden onay alınmıştır.

### 3.5. Veri Toplama Araçları

Hastalara ait klinik ve demografik özellikleri belirlemeye yönelik olarak 156 ana soru sorulmuştur. Bu soruların alt soruları da düşünüldüğünde 545 soruya cevap verilmiştir. Soruların cevapları Akdeniz Üniversitesi Hastanesi Nöroloji Polikliniği'nde görev yapan ve uluslararası projeye dâhil olan doktorlar tarafından hastalar ile birebir iletişim kurularak alınmıştır. Sorulara verilen cevaplar çeşitli kurallar ve filtreler kullanılarak Excel tablosunda tutulmuştur.

### 3.6. Veri Seti Genel Bilgileri

Excel tablolarına kaydedilen bilgiler doğrultusunda verileri eksiksiz olan ve tez kapsamında analize alınan 65 hasta ile ilgili tanımlayıcı bilgiler aşağıda yer almaktadır. 65 hastanın yaş aralığına göre dağılımı Şekil 3.1'de verilmiştir. Literatür özetinde de belirtildiği üzere hastalığın ortaya çıktığı 55-75 yaş aralığındaki hasta sayısı 39'dur ve toplam hastaların %60'ına karşılık gelmektedir.

**age**

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 30-40	9	13,8	13,8	13,8
41-50	16	24,6	24,6	38,5
51-60	20	30,8	30,8	69,2
61-70	9	13,8	13,8	83,1
71-80	10	15,4	15,4	98,5
81-90	1	1,5	1,5	100,0
Total	65	100,0	100,0	

Şekil 3.1. 65 hastanın yaş aralığına göre dağılımı

65 hastanın cinsiyet dağılımı Şekil 3.2’de verilmiştir. Erkek hastalar toplam hastaların %67’sini oluşturmaktadır.

		sex			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Kadın	21	32,3	32,3	32,3
	Erkek	44	67,7	67,7	100,0
	Total	65	100,0	100,0	

Şekil 3.2. 65 hastanın cinsiyete göre dağılımı

65 kişiden oluşan hasta grubunun doğum yeri büyüklüğüne göre yapılan sınıflandırma Şekil 3.3’de verilmiştir. Hastaların yarısından fazlası (%53) köy gibi küçük yaşam alanını oluşturmuştur.

		plbirth			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Village	21	32,3	32,3	32,3
	Country	14	21,5	21,5	53,8
	Small Town	7	10,8	10,8	64,6
	Middle Town	9	13,8	13,8	78,5
	Large Town	14	21,5	21,5	100,0
	Total	65	100,0	100,0	

Şekil 3.3. 65 hastanın doğum yeri büyüklüklerine göre sınıflandırılması

Hastaların doğum yeri özellikleri kırsal alan ve kentsel alan olarak gruplandırılmıştır. Yapılan bu gruplandırmaya göre hastaların %66’sı kırsal alanda doğmuştur. Bu hastalara ait bilgiler Şekil 3.4’de verilmiştir.

		plbirth2			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Kırsal	40	61,5	61,5	61,5
	Kentsel	25	38,5	38,5	100,0
	Total	65	100,0	100,0	

Şekil 3.4. 65 hastanın doğum yeri özelliği

Hastaların ilk semptomun görülmesinin ardından kaç ay sonra hayatını kaybettiği veya hayatta olduğu ile ilgili bilgiler Şekil 3.5’de verilmiştir. Veri setinde 56 hasta hala hayattadır ve toplam hasta grubunun %86’sını oluşturmuştur.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0-12 ay	1	1,5	1,5	1,5
	12-24 ay	2	3,1	3,1	4,6
	36-48 ay	2	3,1	3,1	7,7
	60-72 ay	2	3,1	3,1	10,8
	72-84 ay	1	1,5	1,5	12,3
	Hayatta	56	86,2	86,2	98,5
	96 aydan fazla	1	1,5	1,5	100,0
Total		65	100,0	100,0	

Şekil 3.5. 65 hastanın hayatını kaybettiği dönem aralığı

65 hastanın sigara kullanımı ile ilgili bilgiler Şekil 3.6’da verilmiştir. Hasta grubu ağırlıklı olarak sigara kullanmamakta ve toplam hasta grubu içerisinde %72’lik bir oranı oluşturmaktadır.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	18	27,7	27,7	27,7
	No	47	72,3	72,3	100,0
Total		65	100,0	100,0	

Şekil 3.6. 65 hastanın sigara kullanım durumu

65 hasta içerisinde mutasyonu tanımlı 27 hasta bulunmaktadır. Mutasyonu tanımlanmış hastalarda görülen mutasyon tiplerine yönelik bilgiler Şekil 3.7’de verilmiştir. Mutasyonu tanımlanmış grup içerisindeki hastaların %40’ı sadece SOD1 mutasyonunu göstermiştir.

mutdur

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Tanımlanmamış	38	58,5	58,5	58,5
FUS	4	6,2	6,2	64,6
SOD1+FUS	3	4,6	4,6	69,2
FUS+TARDBP	3	4,6	4,6	73,8
SOD1+FUS+TARDBP	2	3,1	3,1	76,9
SOD1+FUS+TARDBP+C9ORF72	1	1,5	1,5	78,5
FUS+C9ORF72	1	1,5	1,5	80,0
C9ORF72	1	1,5	1,5	81,5
C9ORF72	1	1,5	1,5	83,1
SOD1	11	16,9	16,9	100,0
Total	65	100,0	100,0	

Şekil 3.7. 65 hastanın mutasyon bilgileri

65 hastanın fiziksel egzersiz yapıp yapmama durumuna göre alınan bilgiler Şekil 3.8’de verilmiştir. Hastaların %72’sinin düzenli egzersiz yapıp yapmadığı bilinmemekle birlikte meslek gruplarına bakıldığında beden ağırlıklı işler yaptığı görülmektedir.

mainoccup

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Unknown	47	72,3	72,3	72,3
Yes	14	21,5	21,5	93,8
No	4	6,2	6,2	100,0
Total	65	100,0	100,0	

Şekil 3.8. 65 hastanın düzenli fiziksel egzersiz yapma durumu

Hastaların mesleklerine yönelik alınan bilgiler Şekil 3.9’da verilmiştir. Bu bilgilere göre hastaların çoğunluğu nitelik gerektirmeyen işlerde ve nitelikli tarım, ormancılık ve su ürünleri alanında çalışanlardan oluşmaktadır.

		work			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Silahlı Kuvvetler Çalışanı	8	12,3	12,3	12,3
	Yöneticiler	2	3,1	3,1	15,4
	Profesyonel Meslek Grupları	4	6,2	6,2	21,5
	Teknisyenler, Teknikerler ve Yardımcı Profesyonel Meslek Grupları	1	1,5	1,5	23,1
	Büro Hizmetlerinde Çalışan Elemanlar	2	3,1	3,1	26,2
	Hizmet ve Satış Elemanları	3	4,6	4,6	30,8
	Nitelikli Tarım, Ormancılık ve Su Ürünleri Çalışanları	17	26,2	26,2	56,9
	Sanatkarlar ve İlgili İşlerde Çalışanlar	2	3,1	3,1	60,0
	Tesis ve Makine Operatörleri ve Montajcılar	1	1,5	1,5	61,5
	Nitelik Gerektirmeyen İşlerde Çalışanlar	23	35,4	35,4	96,9
	Unknown	2	3,1	3,1	100,0
	Total	65	100,0	100,0	

Şekil 3.9. 65 hastanın meslek bilgileri

### 3.7. Verilerin Analizi

Çalışma kapsamında genetik mutasyonları tanımlanmış 27 hastanın klinik, demografik ve genetik verilerinin dâhil olduğu toplam 65 hastaya ait klinik ve demografik bilgiler kullanılmıştır. Tez çalışmasının analiz süreçlerinde, mevcut bir yazılım olan ve hazır paketlerin kullanıldığı R Programı ile birlikte uygulama geliştirilmesi için Java programlama dili kullanılmıştır.

R programının kullanılma sebebi; doğrusal ve doğrusal olmayan modelleme, istatistik teknikleri, kümeleme, sınıflandırma gibi geniş istatistik ve grafik tekniklerini içeren, açık kaynak kodlu olması nedeni ile geliştirmeye açık bir program olmasıdır. Bunun yanı sıra ücretsiz olması, araştırmacıların geliştirdikleri algoritmaları R ortamında devamlı paylaşıyor olmaları, işletim sisteminden bağımsız olarak çalışabilmesi ve diğer diller ve programlar ile bağlantı desteği sunması R'in tercih edilme nedenleri arasındadır (Baydoğan ve Orbay, 2014).

Çeşitli dillerde yazılmış kütüphaneleri rahatlıkla kullanabilmesi, kolay bir şekilde diğer platformlara taşınabilmesi ve temel programlama dillerinden birisi olması sebebi ile de Java programlama dili bu çalışma içerisinde kullanılmıştır (Kuşçu, 2009).

Çalışmada; nesnelerin veya bireylerin temel özelliklerini dikkate alarak benzer özelliklerine göre onları gruplandırma ve özet bilgiler ile sonuçları sunma olanağı sağlayan hiyerarşik kümeleme yöntemi kullanılmıştır. Bu yöntem kümelerin homojen ve heterojen gruplar oluşturmasını sağlayarak hiyerarşik bir yapı meydana getirir (Yaz, 2014). Veri setinin özellikleri de göz önünde bulundurulduğunda çalışmada birleştirici hiyerarşik küme yönteminin kullanılmıştır. Bu kümeleme yönteminde başlangıç aşamasında her bir hasta farklı bir küme olarak değerlendirilmiş ve benzer özellikler ortaya çıktıkça diğer bir küme ile birleşmiştir ve nihayetinde tüm ilişkileri gösteren dendogram grafikleri oluşturulmuştur. Bu yöntem her bir hastanın özelliklerinin diğer bir hastanın özellikleri ile olan benzerliğini karşılaştırarak çıktı üretmesi nedeni ile tercih edilmiştir. Hiyerarşik kümeleme yönteminde, küme ayrıştırması için toplam küme içerisindeki varyansı minimize eden ve her adımda kümeler arası minimum mesafe olan diğer küme ile birleşen Vard Minimum Varyans Yöntemi ile Öklit Uzaklık Metriği kullanılmıştır.

Kullanılan diğer bir yöntem karar ağaçları yöntemidir. Büyük miktardaki verilerin daha küçük kayıtlara bölünmesi ile oluşur. Örnek eğitim veri setinden tümevarım yönteminin kullanılması ile ağaç oluşur. Karar ağaçları tahmin etmeye ve tanımlamaya yöneliktir ve büyük veri tabanlarının sınıflandırılmasında, problem çözümünde yaygın olarak kullanılmaktadır. Karar ağaçlarının yorumlanmasının kolay olması, güvenilirliğinin daha yüksek olması ve yaygın kullanılan bir yöntem olması (Albayrak ve Yılmaz, 2009) nedeni ile tez çalışmasında bu yöntem tercih edilmiştir. Karar ağacı için tüm niteliklerin birbirinden bağımsız ve aynı derecede önemli olduğu ve kategorik veya sürekli verilerde kullanılabilen Naive Bayes Algoritması kullanılmıştır.

Bu yöntemler doğrultusunda verilerin analizi için **R 3.4.3** programı kullanılmıştır. Analiz sonuçları, bulgular bölümünde açıklamalı olarak verilmiştir. R programında yapılan analizlerin çıktıları temelde üç temel görsel tipi aracılığıyla sunulmuştur. Bu görseller;

**Kümeleme Dendogram Görüntüleri:** Dendogramda aşağıdan yukarıya doğru çıktıkça birbirine benzer olan hastalar dallar halinde birleşmeye başlamaktadır. Otomatik olarak R Programının vermiş olduğu bu görselde yakın gruplar öncelikle birleşmektedir. Daha sonra bu gruplar diğer bir yakın grup ile birleşerek bir ağaç oluşmaktadır. Dikey ekseninde gözlenen birleşim yüksekliği hastalar arasındaki benzemezliğin göstergesidir. Yani daha uzun bir dal daha az benzeyen hastaya karşı gelir. Alt grupları belirlemek ve kümeleme grafiklerini oluşturmak için, hasta sayısı da göz önünde bulundurularak dendogramlar belirli yükseklikten 2 grup olarak kesilmiş ve incelenmiştir.

**Heatmap (Isı Haritası) Görüntüleri:** Isı haritalarında oluşan grafikler hastalar ve özelliklerinin dikey ve yatay ekseninde gösterilmesi ile oluşmuştur. Isı haritası her sütun ve satır arasındaki mesafeyi hesaplar ve benzerlik temelinde düzenler. Isı haritası aslında sayıların renkli hücrelere dönüştürülmesidir ve bu renkler değerler arasındaki ilişkinin yoğunluğunu temsil etmektedir. Benzerlik ölçüsü oluşturmak için, veriler arası uzaklık parametresi kullanılmaktadır. Böylece yakın şeylerin birbirine benzediği ve uzak şeylerin birbirine benzemediği mesafelere çevirim yapılmaktadır. Burada kırmızı, o özelliğe ait ortalamaya kıyasla bir artışı gösterir iken daha açık renkler ise özellikteki ortalamaya kıyasla bir düşüş olduğunu gösterir.

**Kümeleme Grafiği Görüntüleri:** Kümeleme grafikleri temel olarak dendogramlarda kesilmiş olan 2 hasta grubunun benzerliğini 2 boyutlu ortamda şematize etmektedir. Hastaları kümelerken, aynı gruptaki hastaların benzer olması ve farklı gruplardaki hastaların birbirinden farklı olması grupları ayırt etmek için önemlidir. Bir takım özelliklerin ortak olması ve her iki grupta da yer alan hastalarda bu özelliklerin gözlemlenmesi kesişim kümesinin oluşmasına sebep olmaktadır. Fakat kesişim grupları hasta gruplarının net bir şekilde oluşmadığını gösterdiğinden kesişim alanlarının oluşması tercih edilmemektedir.

Bu görsellere ek olarak analizlerde iki farklı yöntem daha kullanılmıştır. Bunlar değişken matrisi ve hasta matrisi olarak adlandırılan değişkenler arası benzerlikleri ve hastalar arasındaki benzerlikleri gösteren ısı haritalarıdır. Ayrıca istatistiksel analiz olarak çoğunluğu kategorik değişkenlerden oluşan veri setinde Kikare analizi yapılmıştır.



## 4. BULGULAR

Hastalara ait klinik özellikleri belirlemeye yönelik olarak cevaplandırılan 156 soru sayısı uzman desteği ile azaltılmıştır. 79 farklı soruya düşürülen özellikler neticesinde mutasyonu tanımsız 38 hasta, mutasyonu tanımlı 27 hasta ve tüm hastaların oluşturduğu 65 kişilik hasta grubu için analizler gerçekleştirilmiştir.

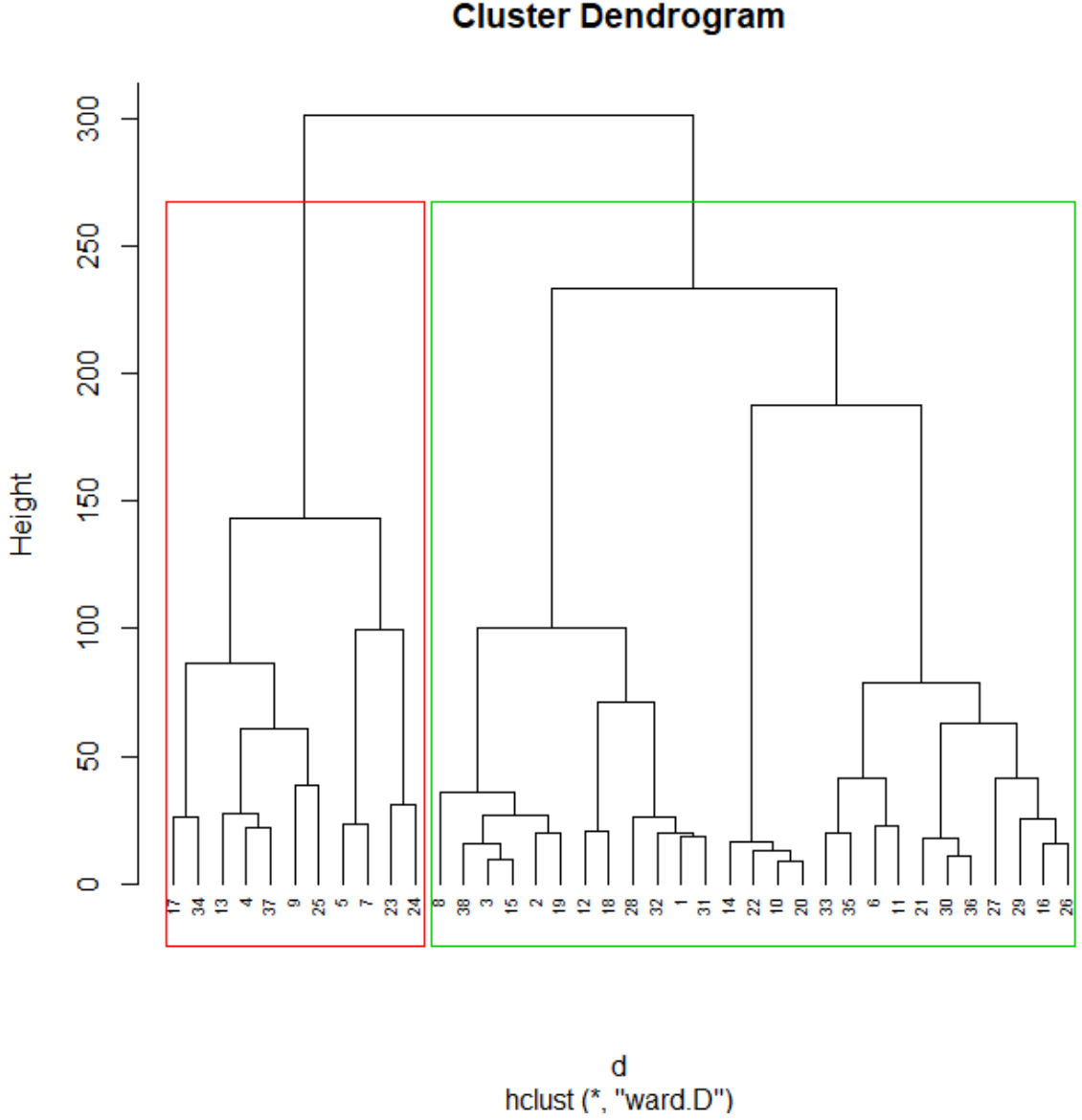
### 4.1. Mutasyonu Tanımlı Olmayan Hasta Grubu

İlk olarak mutasyonu tespit edilmemiş 38 hasta için kümeleme analizi uygulanmıştır. Hastaların mutasyon tipleri belirli olmadığı için kümeleme demografik ve klinik verileri bakımından kümelenmiştir. Bu başlık altında öncelikli olarak genel kümeleme yapısını görmek için 38 hastanın tüm verilerinin olduğu sonuçlar sonrasında da demografik ve klinik özelliklere göre oluşan gruplar verilmiştir.

R programında yapılan kümeleme analizi sonucu oluşan dendogram Şekil 4.1'de verilmiştir. Şekilde, 11 hastadan oluşan 1. Grup ve 27 hastadan oluşan 2. Grup bulunmaktadır. Hastaların tüm verilerinin kümelenmesi sonucu oluşan dendogram grafiği incelendiğinde 1.grupta bulunan hastaların %81'i erkektir ve hastaların 2'si hayatını kaybetmiştir. Hastaların %81'i kol ve bacak başlangıçlıdır ve başlangıçta duygusal kararsızlığı yoktur. Hastaların hepsinde baskın taraf distal'dir ve %90'ında başlangıçta seğirme vardır. Bu grupta bulunan hastaların %90'ında başlangıçta bilişsel belirti olmamakla birlikte sadece 1 hastada dil spastisitesi vardır. Bir hasta hariç hiçbir hastada alt ve üst ekstremitede istirahet tremoru ve alt ekstremitede hareket yavaşlığı yoktur. Ayrıca hastaların hiç birisinde üst ekstremitede hareket yavaşlığı yoktur. Bu grupta bulunan hastalarda maksimum yayılan bölge sayısı 9'dur ve El Escorial kriterlerine göre gruplandırmada tanımlanmış (Define) ALS'dir. Bu grupta bulunan hastaların %81'i sigara kullanmamaktadır.

2.Grupta ise hastaların %70'i erkek ve hastaların %85'i kol ve bacak başlangıçlıdır ve hastaların hiç birisinde alt ekstremitede istirahet tremoru yoktur. 1. Grupta hastalığın yayılan maksimum bölge sayısı 9 iken 2. Grupta yer alan hastaların maksimum bölgeye yayılma sayısı 7'dir ve hastaların %92'sinde hipertansiyon, tip I/II diyabet,

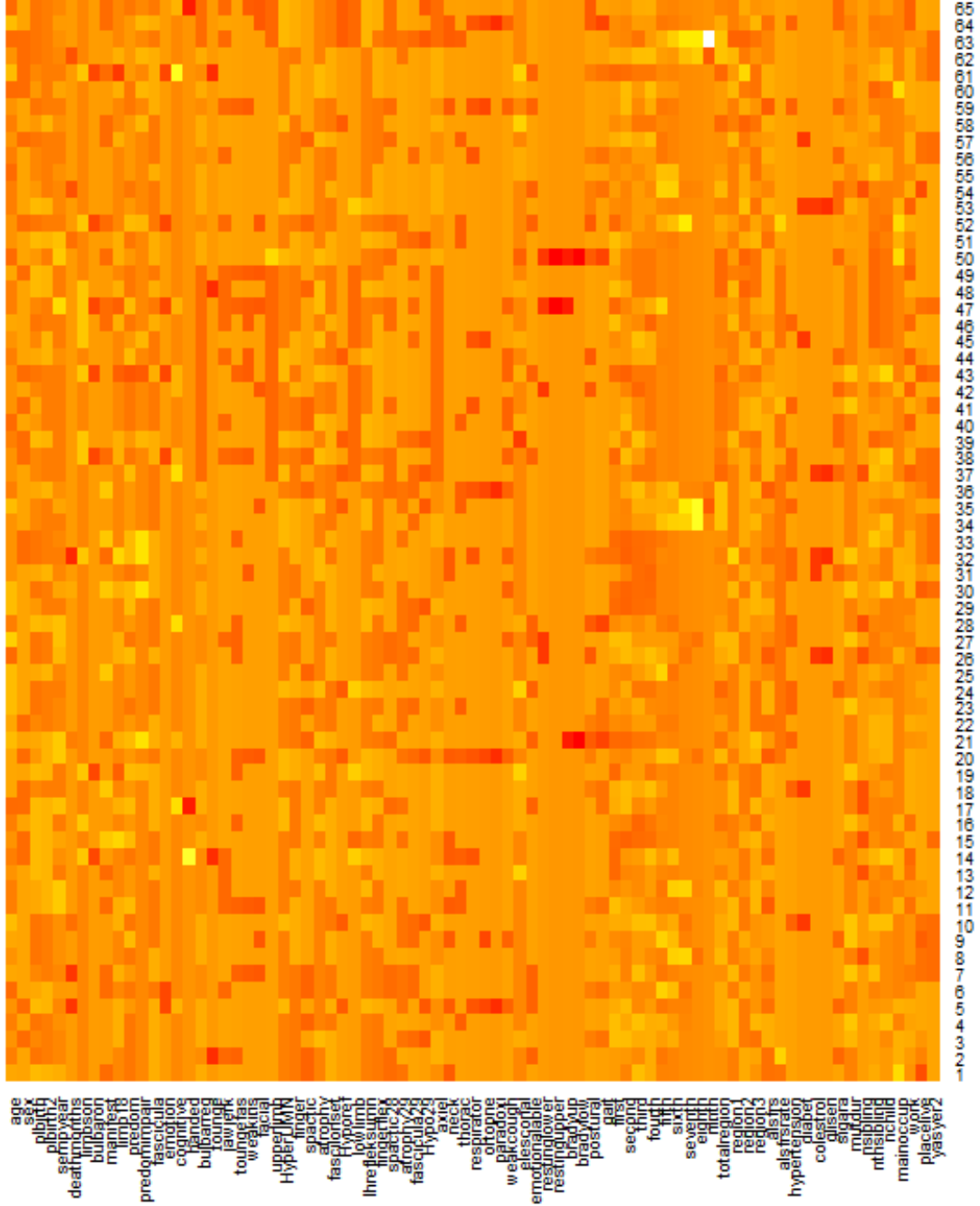
hiperkolesterolemi ve trigliset yüksekliđi yoktur. Ayrıca hastaların %77'si kentsel alanda yařamakta ve %62'si sigara kullanmamaktadır.



**řekil 4.1.** 38 hastanın tüm verilerinin dendrogram görüntüsü

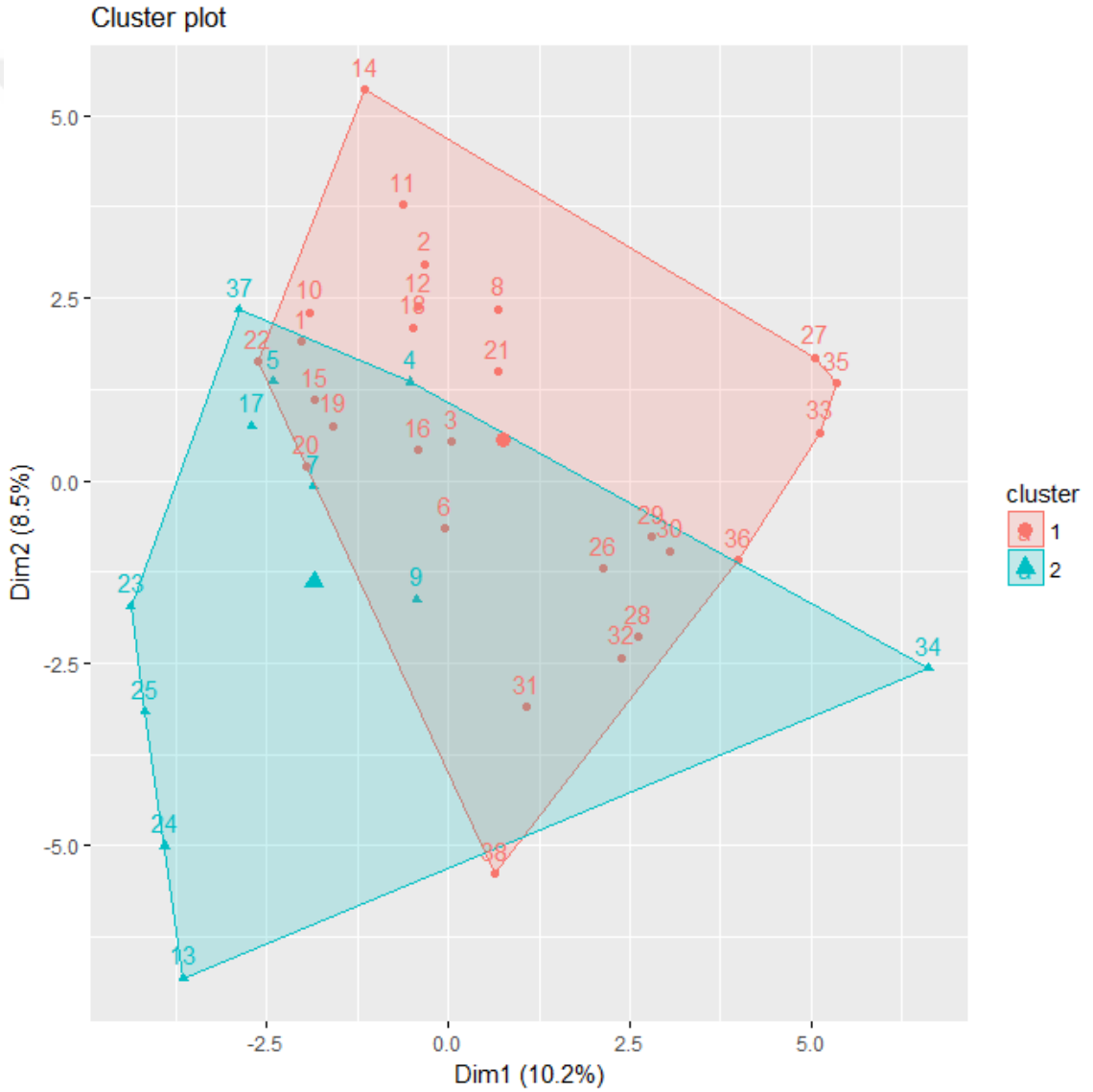
Isı haritası bir matristeki deđerlerin tek tek renk olarak temsil edildiđi verileri grafiksel olarak gösterir. Bu nedenle ısı haritası sayısal verilerin genel bir görünümünü oluşturmak için oldukça önemlidir. Daha önce bulgular başlıđında verildiđi üzere ısı haritasındaki özelliđe ait ortalamaya kıyasla bir artış renklerin koyulařmasını sağlamaktadır. Fakat

Şekil 4.2’de de görüldüğü gibi genetik verisi bulunmayan 38 hasta grubunda kırmızı renk neredeyse yok denecek kadar azdır. Bu da ortalamanın üzerinde bir benzerlik olmadığını göstermektedir.



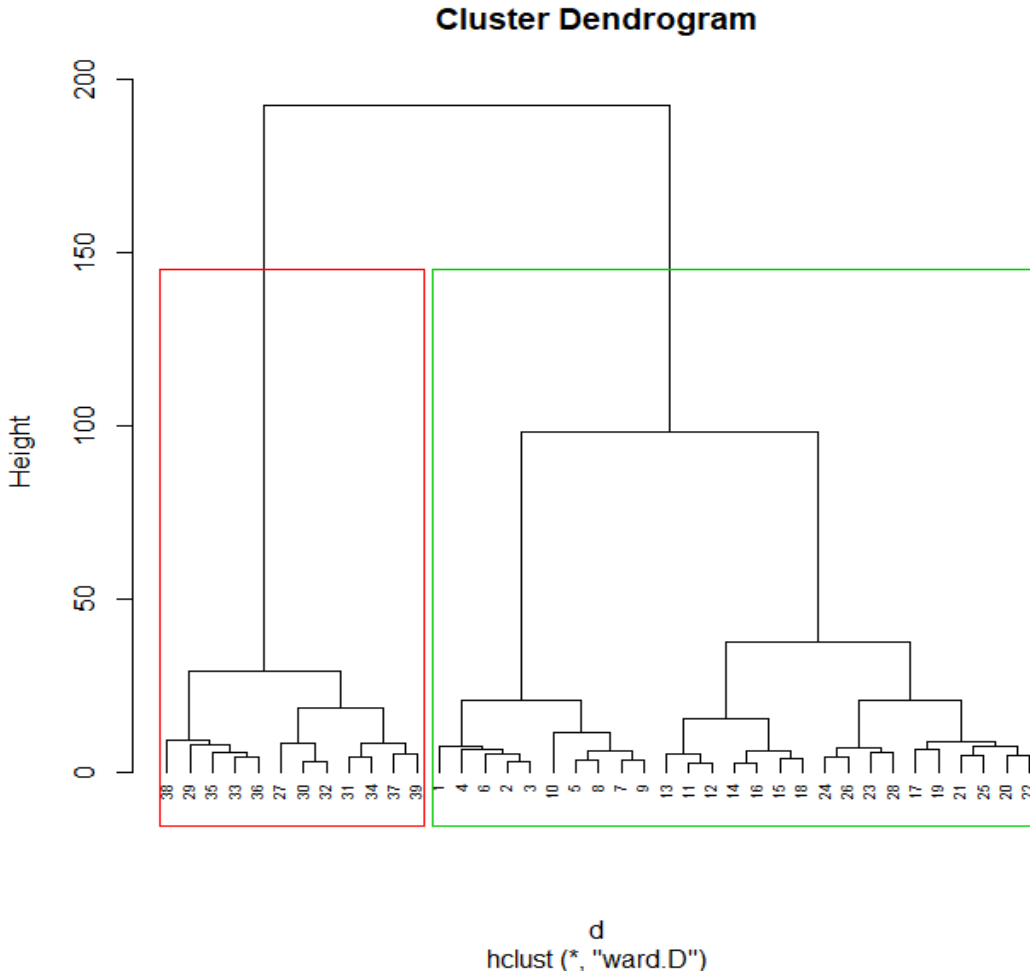
Şekil 4.2. 38 hastanın tüm verilerinin ısı haritası (heatmap) görüntüsü

Şekil 4.3’de verilen hiyerarşik kümeleme grafiği, dendogramda oluşan 2 grubun üyelerini görselleştirmektedir. Dendogramda oluşan 2 farklı gruptaki hastaların çok büyük bir çoğunluğu kümeleme grafiğinde iç içe geçmiş durumdadır. Bu görünüm özellikler bakımından hasta gruplarının ve grup üyelerinin ayrılmadığını göstermektedir. Kesişim kümesi ayrıntılı olarak incelendiğinde sadece hastalarda alt ve üst ekstremitelerde hareket yavaşlığı olmaması, alt ekstremitelerde istirahat tremoru olmaması ortak özellik olarak söylenebilir.



Şekil 4.3. 38 hastanın tüm verilerinin kümeleme grafiği görüntüsü

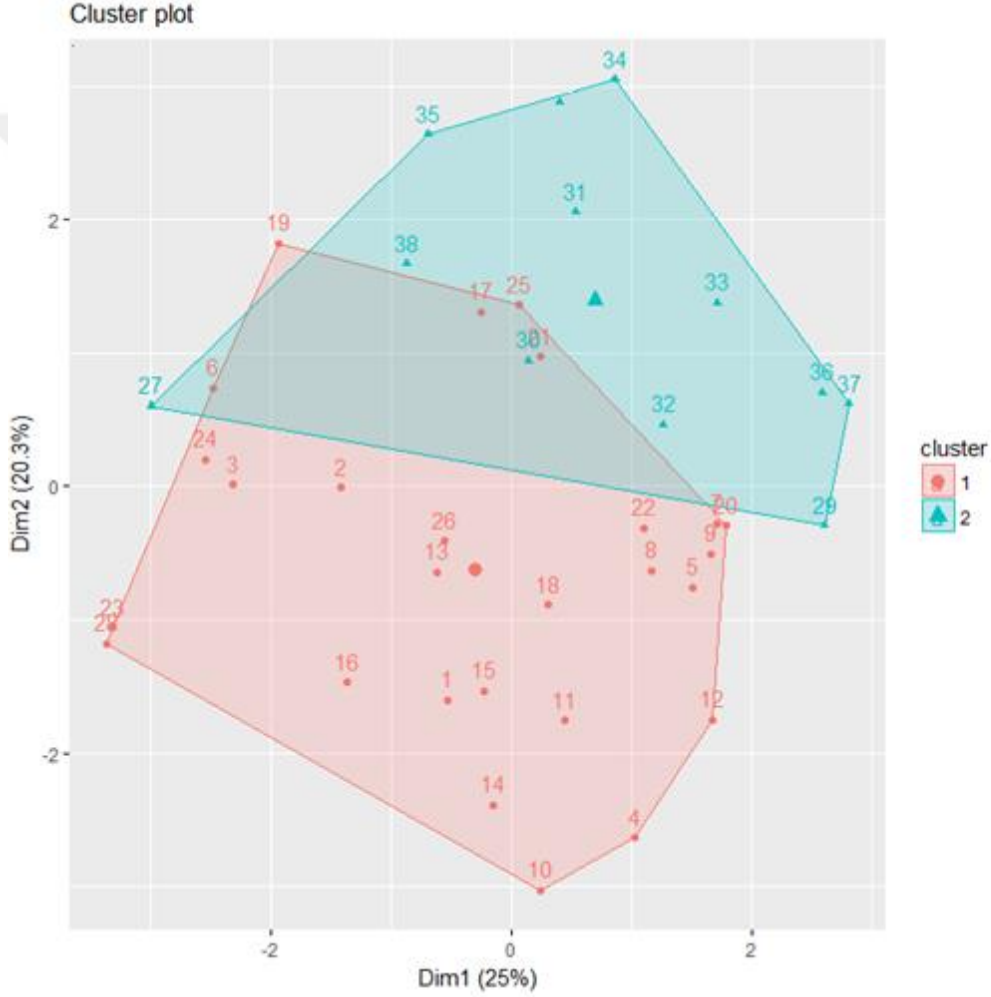
Mutasyonu tanımlanmamış olan hasta grubu için incelenen bir diğer unsur demografik verilerin ilişkisidir. Demografik açıdan kümelenen hasta grubu incelendiğinde 2 ana küme oluşmuştur. Şekil 4.4.'de şematize edildiği gibi, 1. grup 12 hastadan oluşmaktadır ve bu grup içerisinde sadece 29 ve 37 nolu hastalar kadındır. Dolayısı ile bu grubun %84'ü erkek hastalardan oluşmaktadır. Ayrıca bu gruptaki hastaların %84'ü kentsel alanda yaşamaktadır. Klinik özelliklerine bakıldığında ise bu gruptaki hastaların %92'si kol ve bacak başlangıçlıdır. Yine hastaların %92'sinde dil spastisitesi, hiporefleksi, boyun zayıflığı, torasik kaslarda seyirme, duygusal kararsızlık, tip I/II diyabet, hiperkolesterolemi, trigliset yükseklği yoktur ve bu hastalardan sadece bir tanesinin egzersiz yaptığı bilinmektedir. 27 hastadan oluşan 2. Grup'ta bulunan hastaların %93'ünde baskın taraf distal ve %96'sı üst ekstremitate klinik özelliklerini göstermektedir.



**Şekil 4.4.** 38 hastanın demografik verilerinin dendrogram görüntüsü



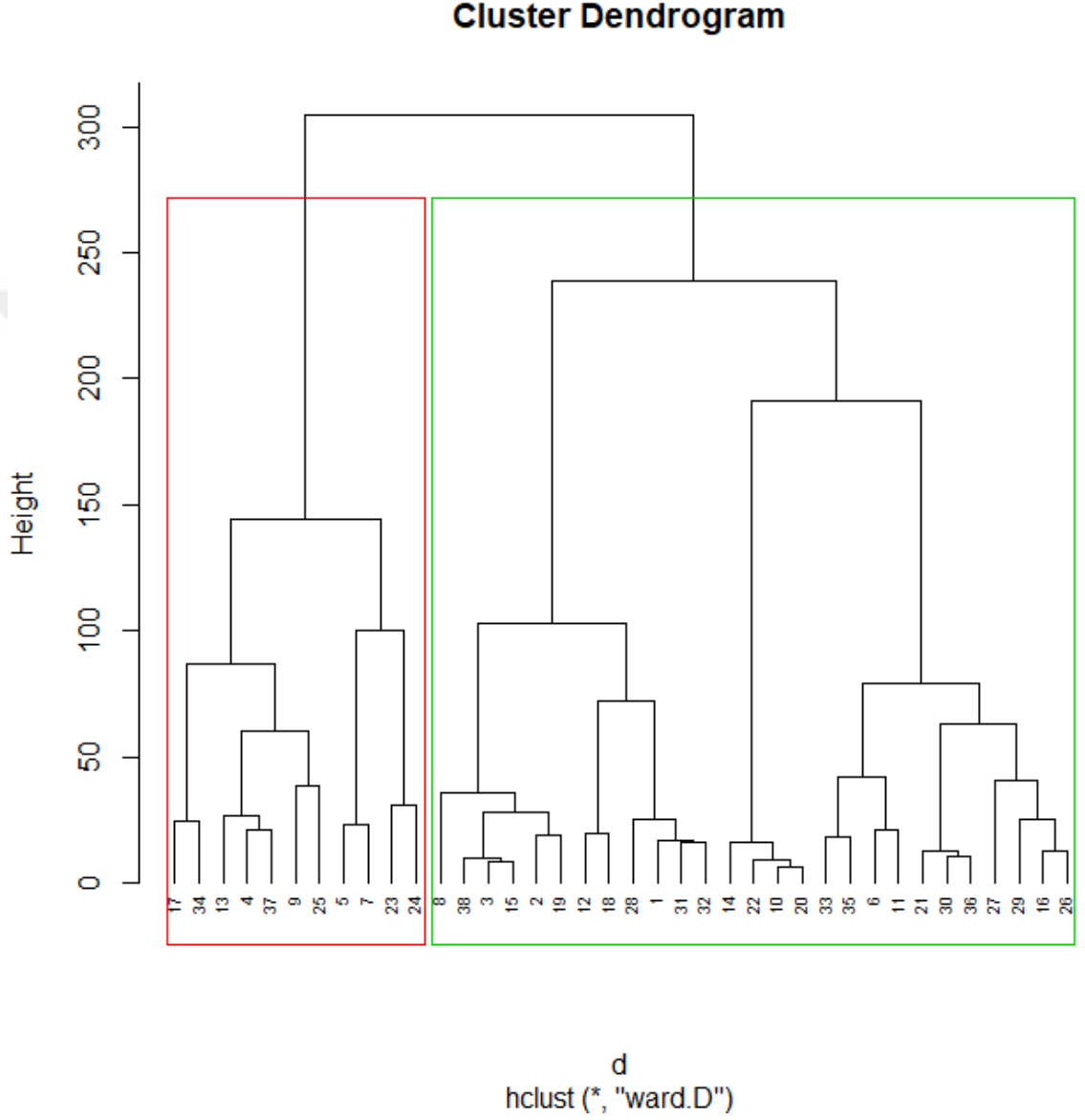
Şekil 4.9.'da verilen hiyerarşik kümeleme grafiği, dendogramlarda kesilmiş olan 2 grubu şematize etmektedir. Bu gruplar arasındaki bazı hastalar tüm veri seti içerisinde benzerlik göstermesi sebebi ile kesişim kümesini oluşturmaktadır. Mutasyonu henüz tespit edilmemiş 38 hastaya ait demografik verilerin kesişim kümesi, bu gruptaki hastaların hepsinin erkek olduğunu ve 1 hasta hariç hepsinin kentsel alanda ve büyük şehirde yaşadığı tespit edilmiştir.



Şekil 4.6. 38 hastanın demografik verilerinin kümeleme grafiği görüntüsü

İncelenen diğer bir unsur 38 hastaya ait klinik verilerin oluşturduğu kümelerdir. Şekil 4.7.'de sunulan dendogram incelendiğinde temelde 2 küme oluşmuştur ve 1. Grupta 11 hasta, 2. Grupta 27 hasta bulunmaktadır. Klinik özellikler bakımından gruplanan hastalar açısından incelendiğinde oluşan gruplar Şekil 4.1 ile aynı olduğunu görmek mümkündür.

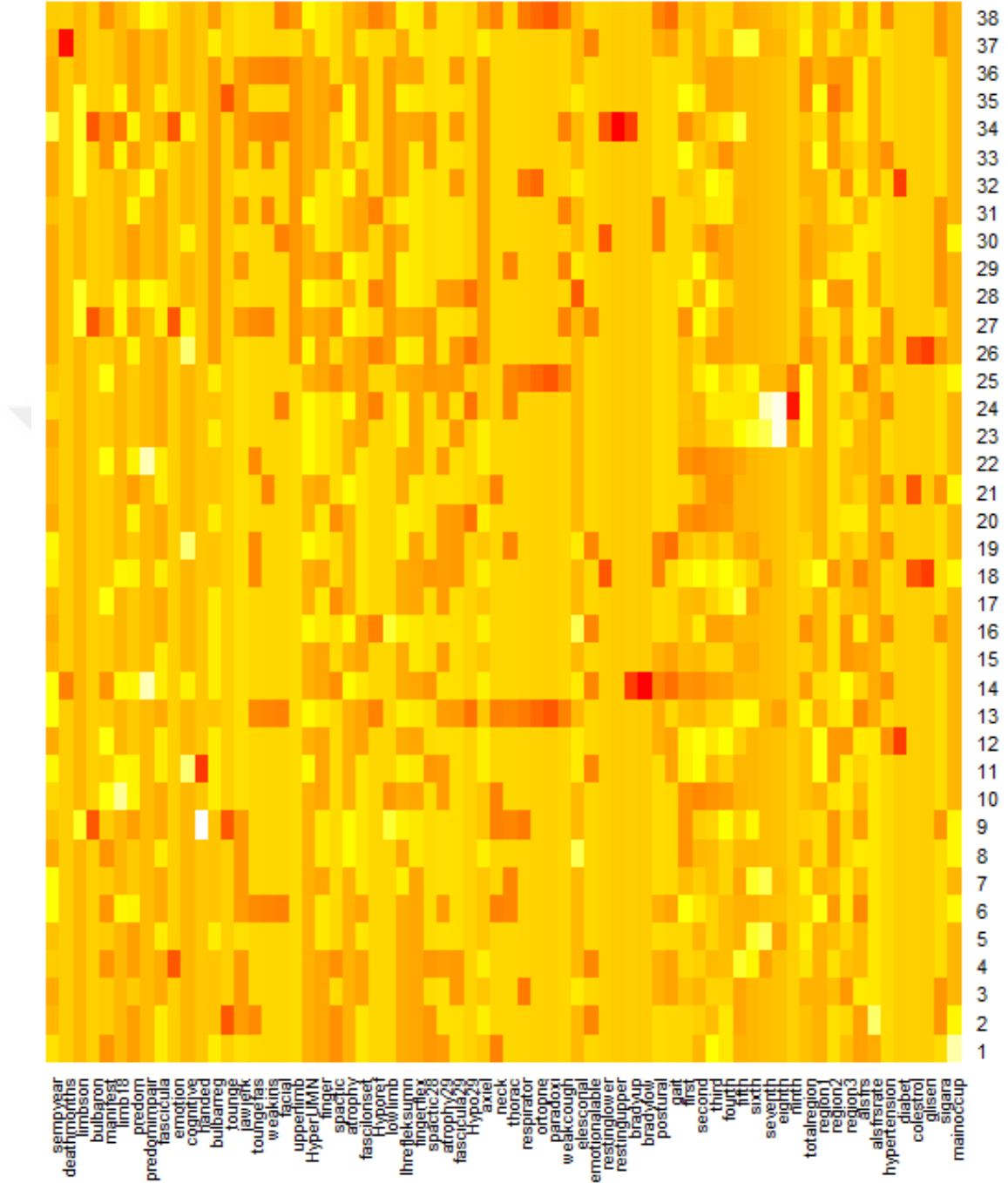
Bu durum, klinik özelliklerin tüm veriler içinde belirleyici ve ayırt edici bir faktör olduğu yönünde yorumlanabilir.



**Şekil 4.7.** 38 hastanın klinik verilerinin dendrogram görüntüsü

Mutasyonu tanımlı olmayan 38 hastanın klinik verilerine ait Şekil 4.8’de verilen ısı haritası görüntüsü özelliklere ait ortalamaya kıyasla bir artış olmadığını göstermektedir. Renklerden de anlaşılacağı üzere klinik özellikler ortalamanın altında bir eğilim göstermiştir. 38 hastanın tüm verilerinin değerlendirildiği ısı haritası ile kıyaslandığında klinik veri setinin kümeler üzerine etkisinin fazla olduğunu söylemek mümkündür.





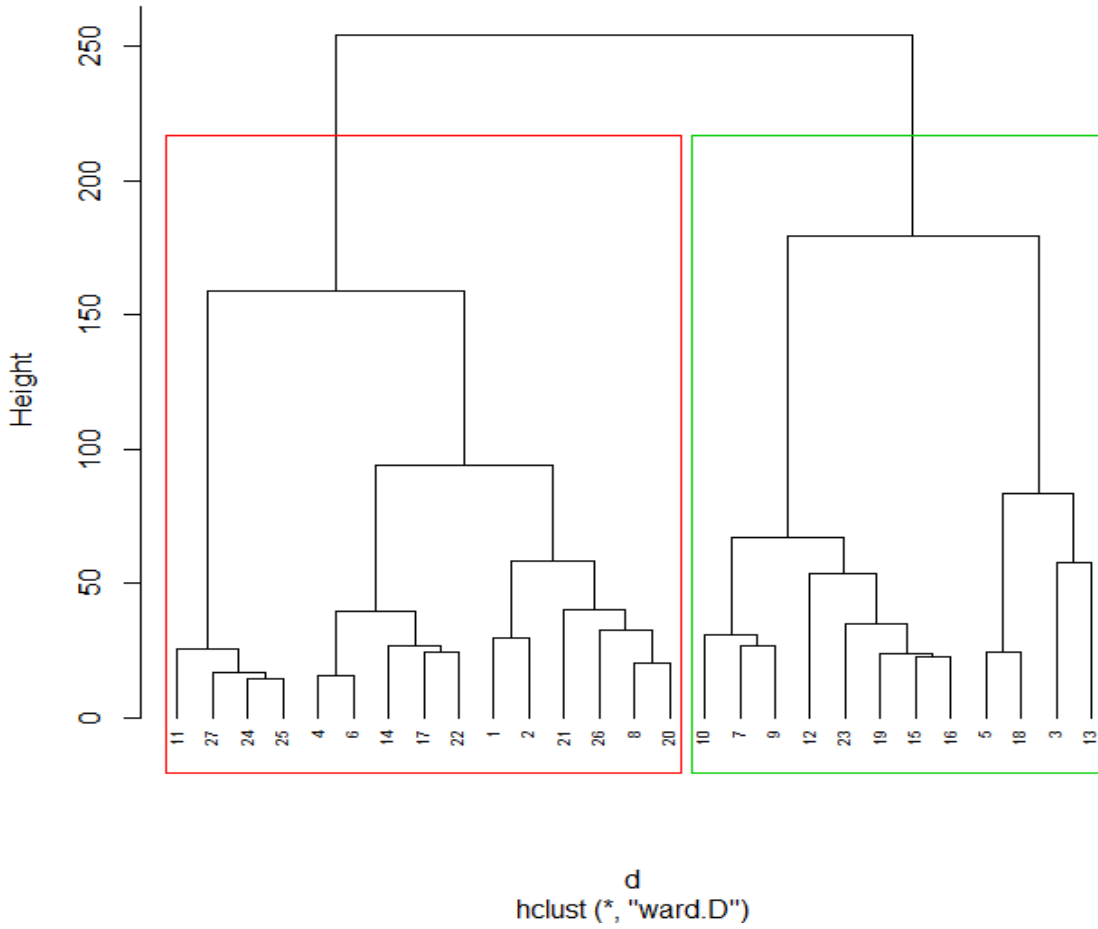
Şekil 4.8. 38 hastanın klinik verilerinin ısı haritası (heatmap) görüntüsü

Şekil 4.9’da verilen 38 hastaya ait klinik verilerin analiz edilmesi sonucu oluşan dendogramlar neticesindeki kümeleme grafiği ve gruplar arasındaki ilişki incelendiğinde, hastaların birbirinden önemli ölçüde ayrışmadığı ve hastaların sayıca çoğunluğunun bir alanda toplandığı görülmektedir. Kesişim alanı incelendiğinde ise, maksimum yayılan



Mutasyonu tanımlı 27 kişilik hasta grubuna ait Şekil 4.10.'da belirtilen dendrogram incelendiğinde, 1. Grupta yar alan 15 kişilik hasta kümesinin hiç birisinde ortopne ve paradoksik solunum yoktur. Hastaların %93'ünde dil spastisitesi, boyun zayıflığı ve diyabet bulunmamaktadır. 1. Grubun genetik verileri incelendiğinde SOD1 mutasyonunun bulunduğu hasta sayısı fazla olsa da 2. Grup ile karşılaştırıldığında anlamlı bir değişiklik bulunmamaktadır. Ayrıca 2. Gruptaki hastaların hiç birisinde dil spastisitesi, başlangıçta bilişsel belirti, paradoksik solunum, trigliset yüksekliği ve hiperkolesterolemi bulunmamaktadır. Hastaların %91'inde atrofi ve güçsüzlük bulunmakla birlikte kol ve bacak başlangıcı, üst ve alt ekstremitte özelliğini göstermektedir ve hastaların %93'ünün ALSFRS-R düşme oranı 0,1-0,9 aralığındadır.

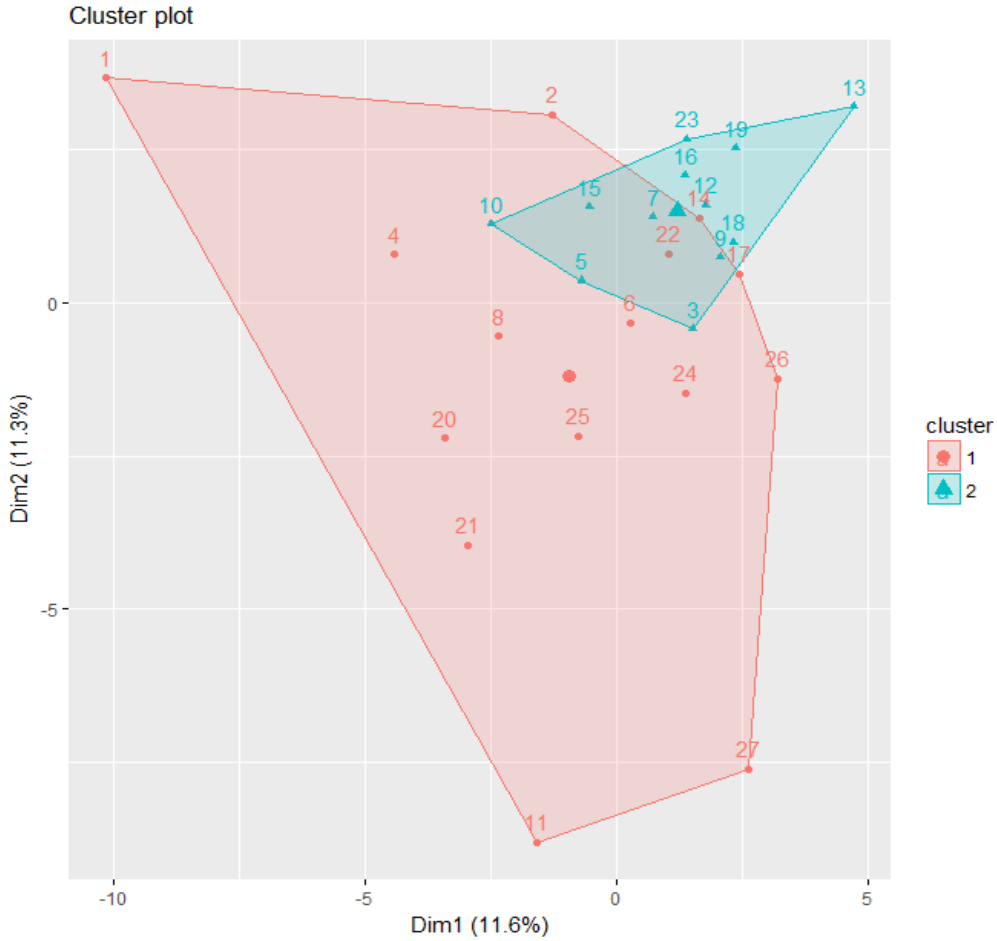
### Cluster Dendrogram



Şekil 4.10. 27 hastanın tüm verilerinin dendrogram görüntüsü



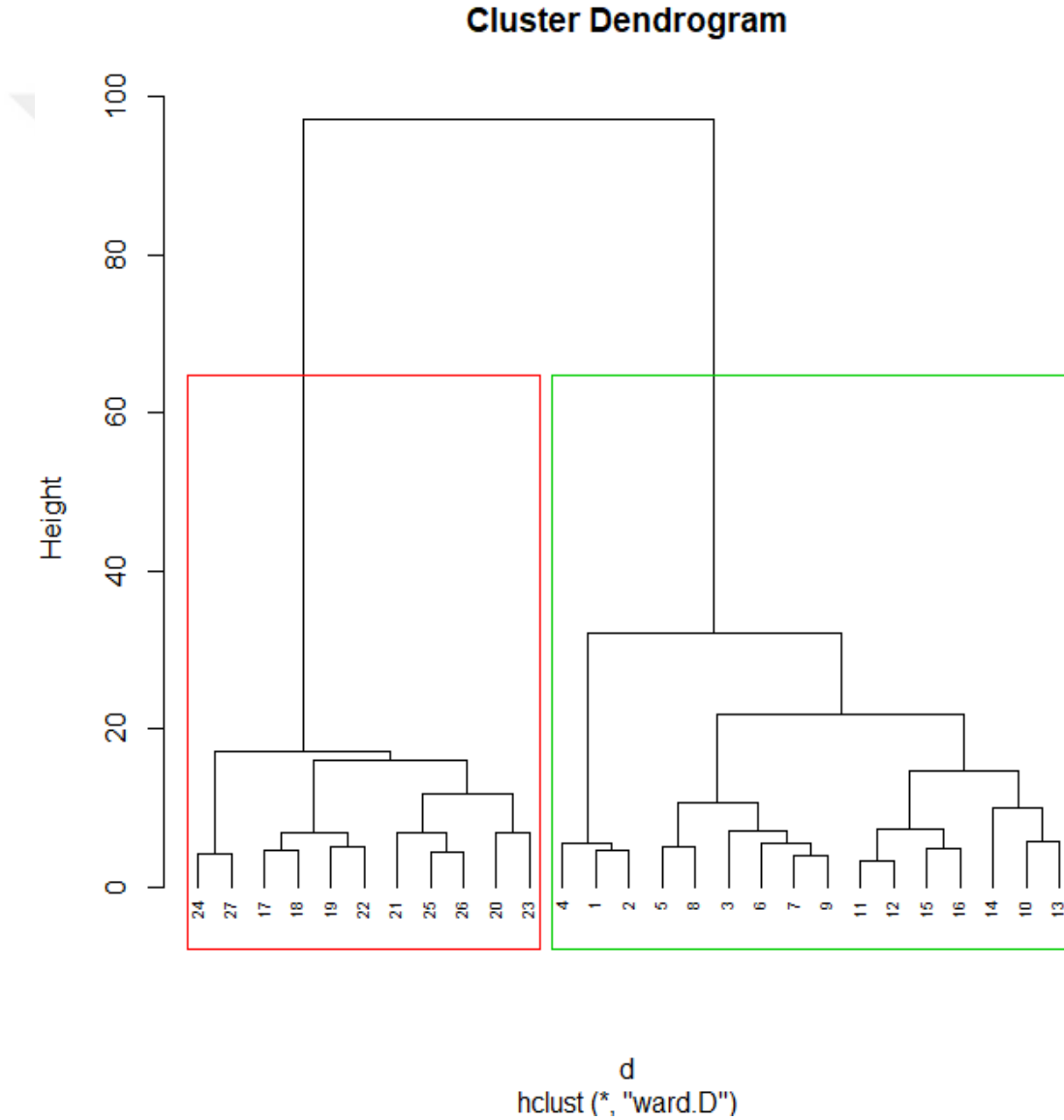
Şekil 4.12.'de verilen hiyerarşik kümeleme grafiği, dendogramlarda bulunan 2 grubu şematize etmektedir. Mutasyonu tanımlanmış olan 27 hastanın kümeleme grafiği, hastaların benzer özelliklerine göre bir alanda toplandığını göstermektedir. Kesişim kümesi incelendiğinde ise, gruptaki hastaların tamamının kentsel alanda yaşadığını, hiç birisinde başlangıçta bilişsel belirti, alt ve üst ekstremitelerde istirahat tremoru, alt ve üst ekstremitelerde hareket yavaşlığı olmadığını ve hepsinde üst motor nöronlarda hiperrefleksi olduğunu tespit edilmiştir.



Şekil 4.12. 27 hastanın tüm verilerinin kümeleme grafiği görüntüsü

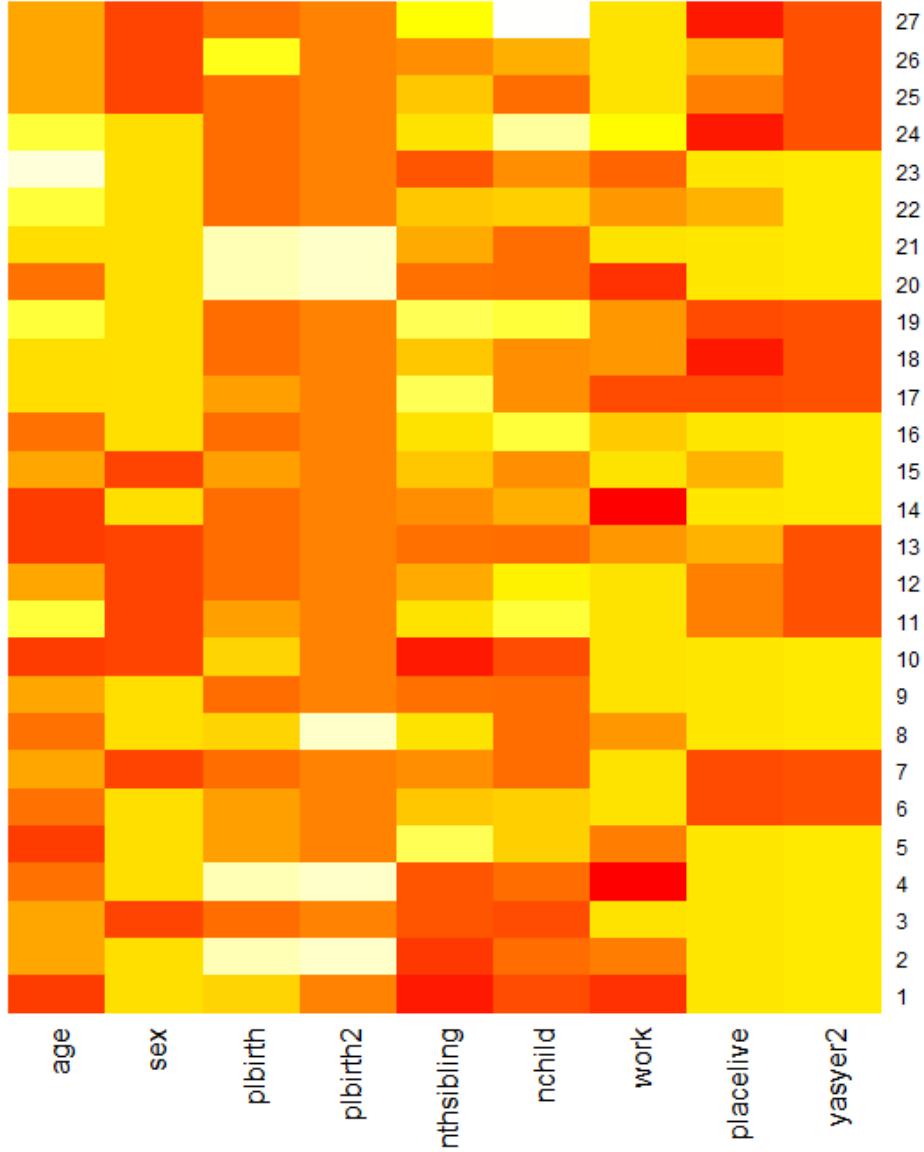
Mutasyonu belirli olan 27 hastanın demografik verilerine ilişkin yapılan analiz sonuçları Şekil 4.13.'de verilmiştir. Analizler sonucunda 2 ana grup oluşmuştur ve gruplar klinik ve genetik açıdan incelenmiştir. Genetik açıdan, 11 hastanın bulunduğu 1. Grubun tamamının SOD1 mutasyonunu gösterdiği tespit edilmiştir. Sadece 21 numaralı hasta SOD

mutasyonunun yanı sıra C9ORF72 mutasyonunu da göstermiştir. 2. Grupta ise SOD1 mutasyonunu gösteren hastalar olsa da 2, 3 hatta 4 mutasyonu da gösteren hastalar bulunmaktadır. Klinik açıdan incelendiğinde ise 1. Grupta, 1 hasta hariç hiçbir hasta sigara kullanmamaktadır ve sadece 2 hastanın egzersiz yaptığı bilinmektedir. 2. Grupta bulunan hastalardan sadece ikisinde başlangıçta bilişsel belirti vardır ve hastaların %88'inde ALSFRS-R düşme oranı 0,1 - 0,9 arasındadır.



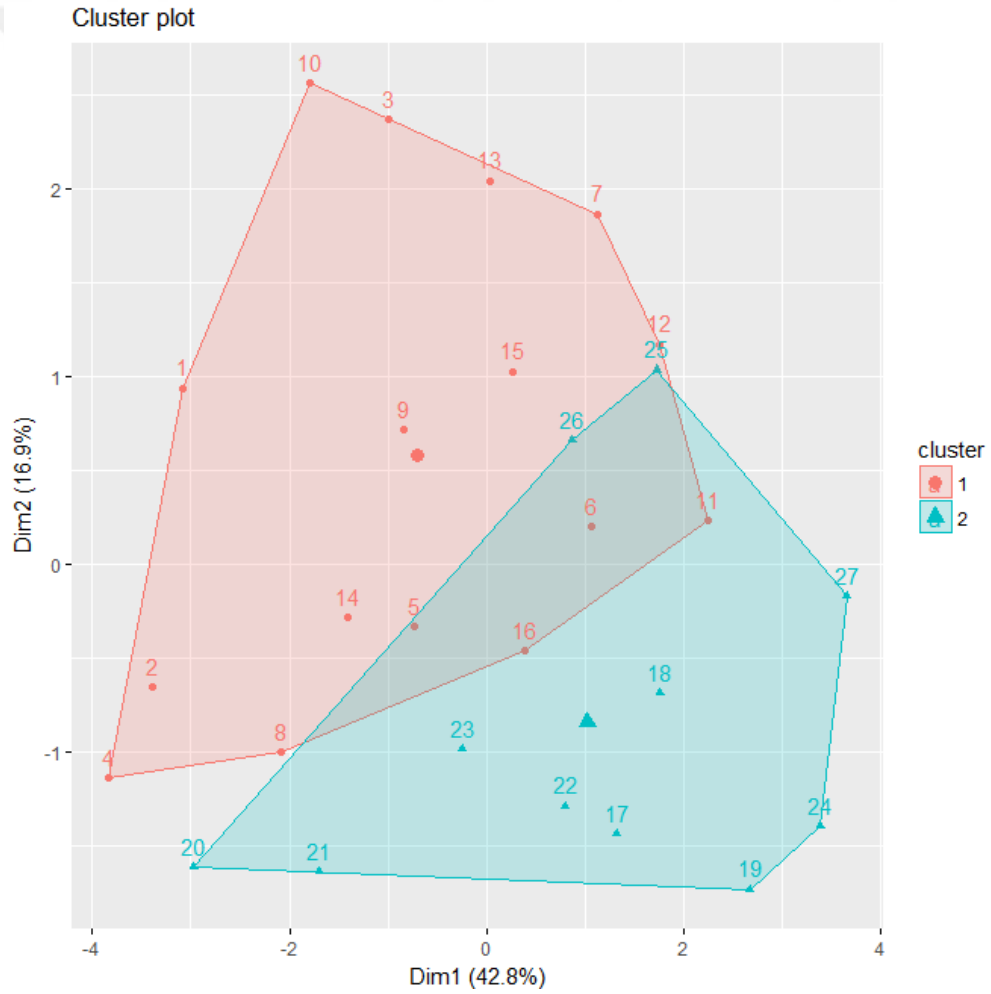
**Şekil 4.13.** 27 hastanın demografik verilerinin dendrogram görüntüsü

Mutasyonu belirsiz hasta grubuna ait demografik verilerin incelendiği ısı haritasında (Şekil 4.5.) olduğu gibi mutasyonu belirli olan hasta grubunun demografik verileri sonucu oluşan ısı haritasında diğerlerine kıyasla daha çok benzerlik gözlemlenmiştir. Bu gösterim Şekil 4.14.'de verilmiştir. Özellikle meslek özelliğinde 4 ve 14 numaralı hastalar, hastanın kaç kardeş olduğu özelliğinde 1 ve 10 numaralı hastalar ve yaşadığı yere göre 18,24,27 nolu hastalar ortalama orana kıyasla daha fazla benzerlik göstermektedir.



Şekil 4.14. 27 hastanın demografik verilerinin ısı haritası (heatmap) görüntüsü

Şekil 4.15.'de verilen hiyerarşik kümeleme grafiği, dendogram grafiğinde belirtilmiş olan 2 grubu şematize etmektedir. Bu gruplar arasındaki bazı hastalar tüm veri seti içerisinde benzerlik göstermesi sebebi ile kesişim kümesini oluşturmaktadır ve diğer grafikler ile karşılaştırıldığında daha az sayıda hasta bu alanda yer almaktadır. Mutasyonu tespit edilmemiş 27 hastaya ait demografik verilerin kesişim kümesinde yer alan hastaların hepsinin kırsal alanda yaşadığı, hepsinde alt ve üst motor nöronlarda hiporefleksi olduğu ve bu hastaların hiç birisinde dilde fasikülasyonu, çene refleksi, yüz kaslarında fasikülasyon, ortopne, atrofi ve güçsüzlük olmadığı tespit edilmiştir.

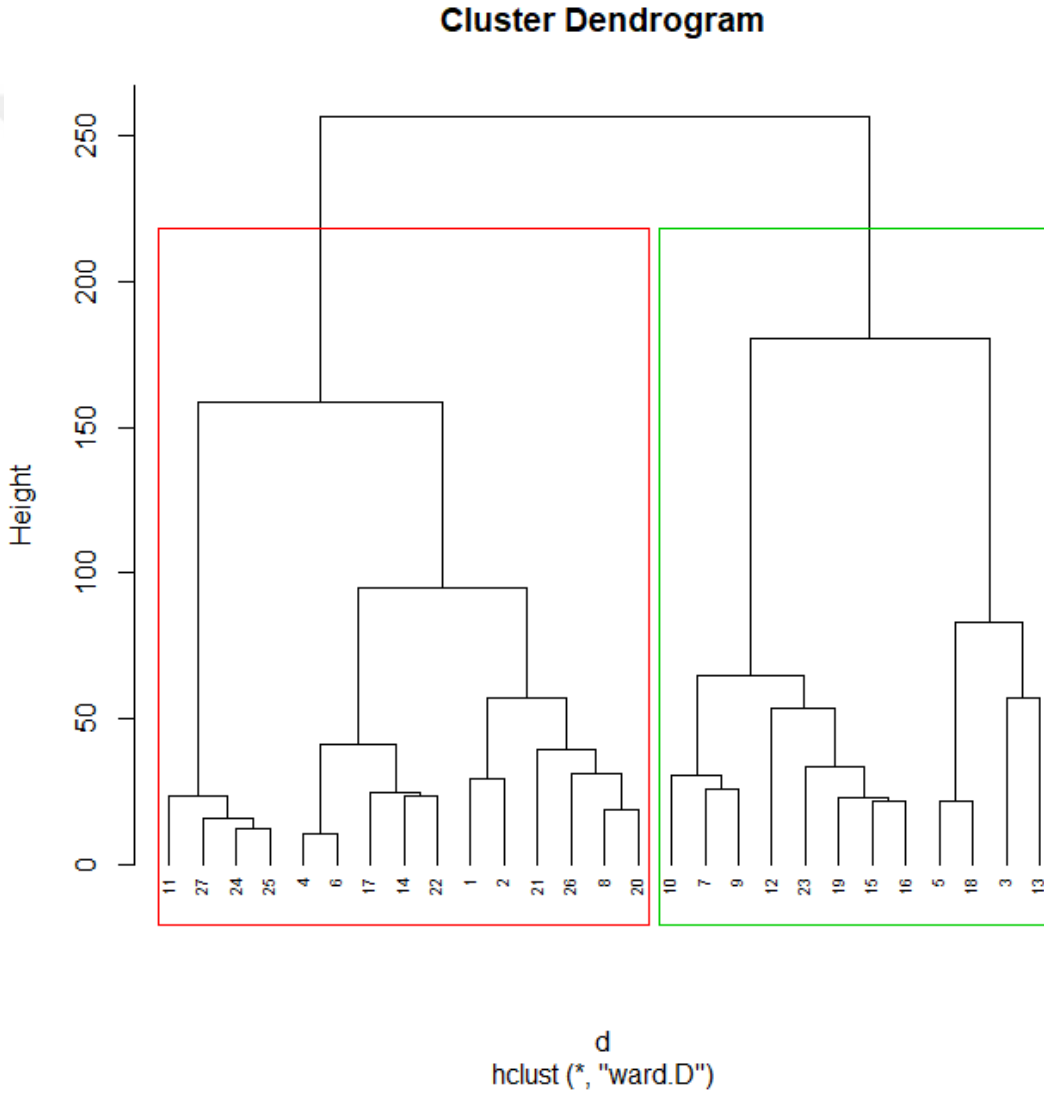


Şekil 4.15. 27 hastanın demografik verilerinin kümeleme grafiği görüntüsü

27 kişilik hasta grubunun klinik açıdan kümeleneceği incelendiğinde 2 ana küme oluşmuştur. Şekil 4.16.'da şematize edildiği gibi, 1. grup 15 hastadan 2. Grup ise 12 hastadan oluşmaktadır. Oluşan bu şemanın 27 hastaya ait tüm bilgilerinin incelenmesi



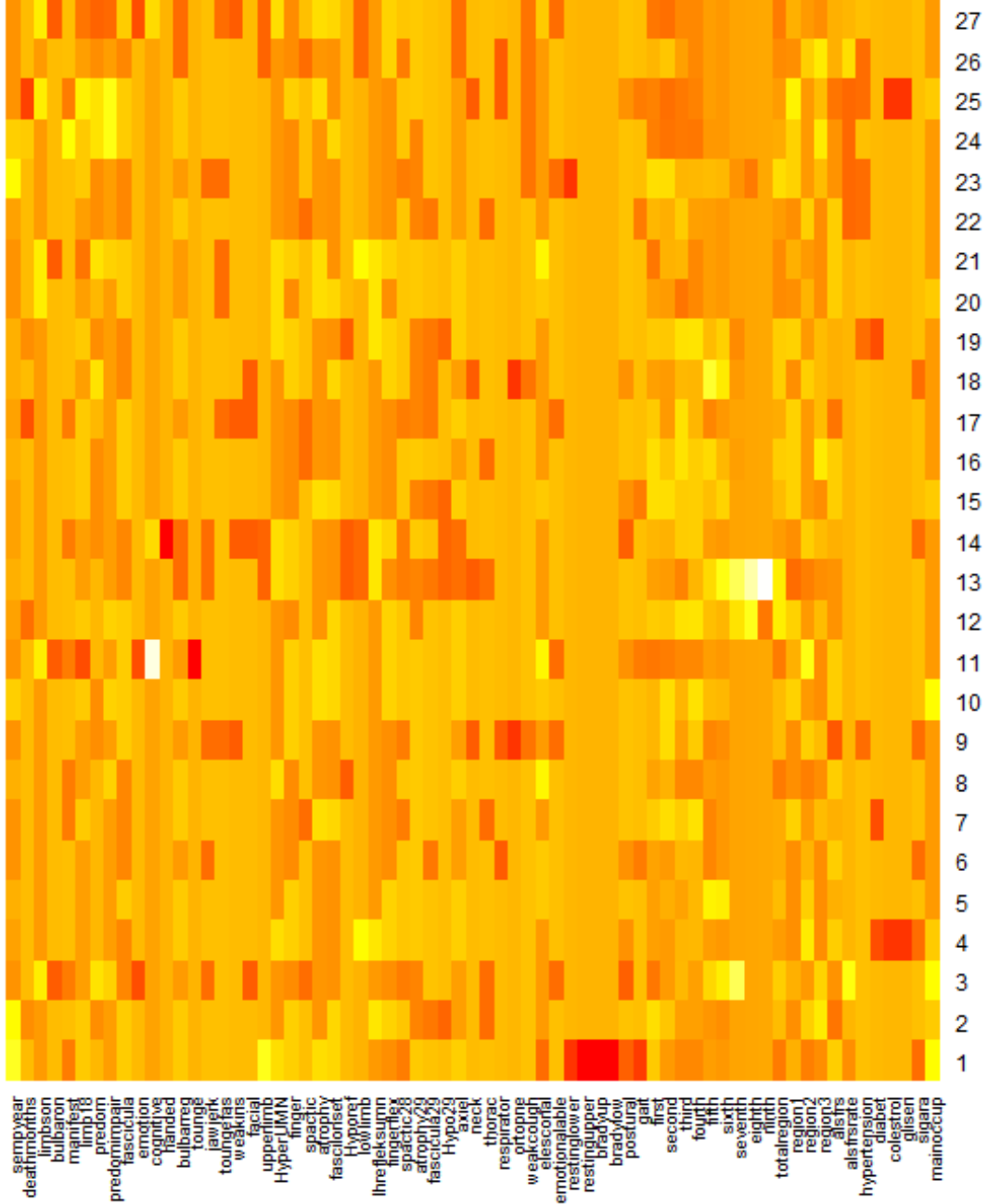
sonucunda oluşan dendrogram ile aynı olduğu görülmüştür. Dendrogram demografik ve genetik açıdan analiz edildiğinde 1. Grup içerisinde SOD1 mutasyonunun bulunduğu hasta sayısı fazla olsa da 2. Grup ile karşılaştırıldığında anlamlı bir değişiklik bulunmamaktadır. Demografik açıdan ise 1. Grupta bulunan kadın hasta sayısı 5, 2. Grupta bulunan kadın hasta sayısı 6 olmakla birlikte önemli bir farklılık göze çarpmamaktadır.



**Şekil 4.16.** 27 hastanın klinik verilerinin dendrogram görüntüsü

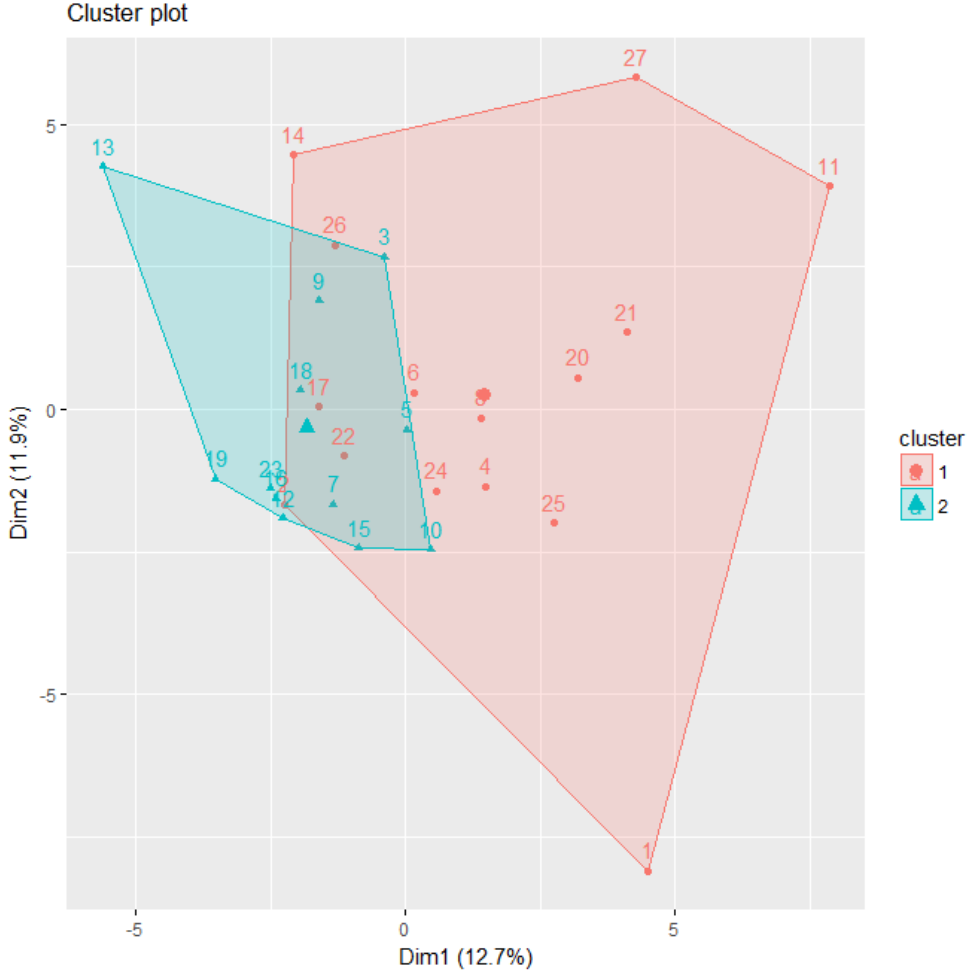
27 hastaya ait klinik verilerin incelendiği grubun ısı haritası Şekil 4.17’de verilmiştir. Bu görsel değerlendirildiğinde özelliklerin neredeyse tamamında ortalamaya kıyasla azalış

eğiliminde olduğunu söylemek mümkündür. Sadece 1 numaralı hastanın alt ve üst ekstremitelerde istirahat tremoru, alt ve üst ekstremitelerde hareket yavaşlığı değerlerinde ortalamaya kıyasla artış olduğu görülmektedir. Bunun haricinde 11 numaralı hastada dil spastisitesi ve 25 numaralı hasta hiperkolesterolemi ve trigliset yüksekliği özelliklerinin ortalamaya kıyasla uzaklık artışı olduğu görülmektedir.



Şekil 4.17. 27 hastanın klinik verilerinin ısı haritası (heatmap) görüntüsü

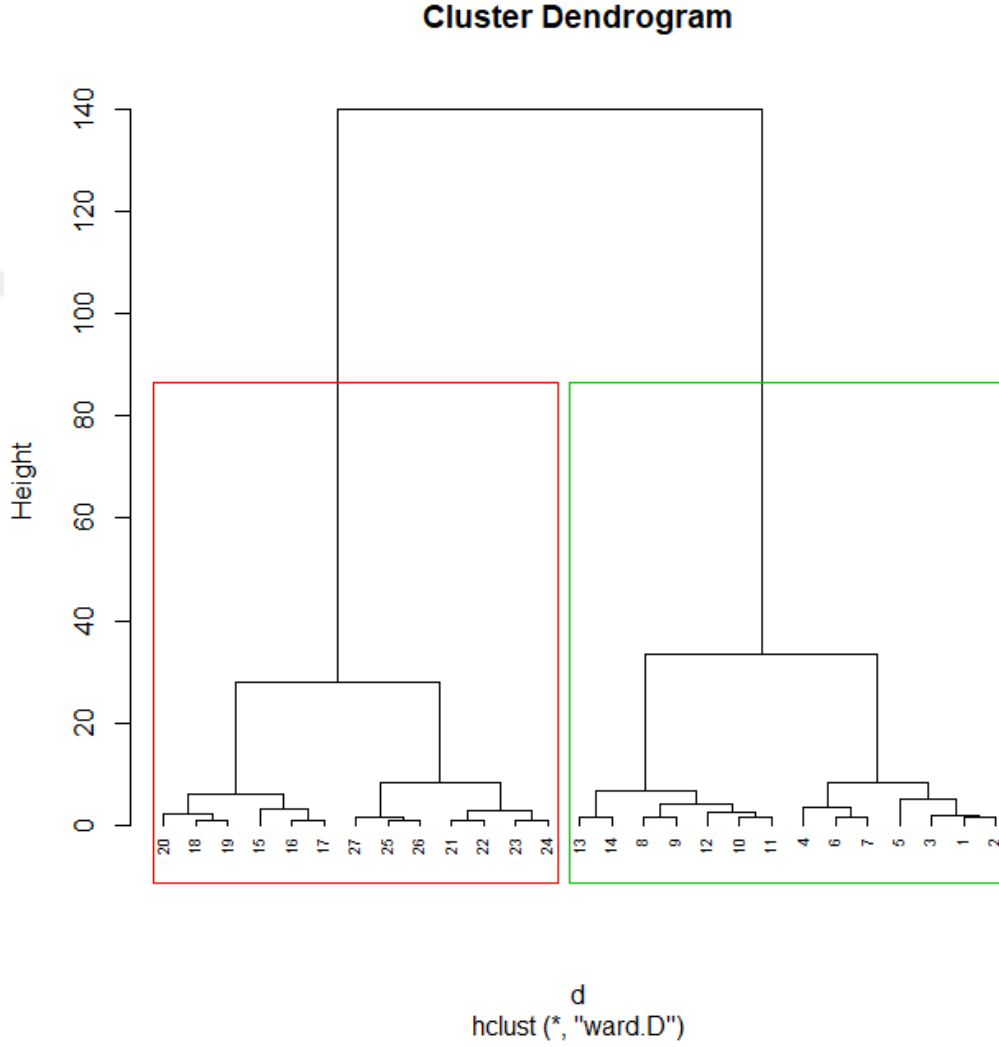
27 hastaya ait klinik verilerin analiz edilmesi sonucu oluşan dendogramlar neticesindeki kümeleme grafiği ve gruplar arasındaki kesişim alanı incelendiğinde hastaların özellikler açısından birbirinden önemli ölçüde ayrılmadığı ve Şekil 4.18.'de görüldüğü üzere hastaların sayıca çoğunluğunun bir alanda toplandığı görülmektedir. Bu nedenle genetik ve demografik özellikler açısından hastalar arasında anlamlı bir farklılık tespit edilememiştir.



Şekil 4.18. 27 hastanın klinik verilerinin kümeleme grafiği görüntüsü

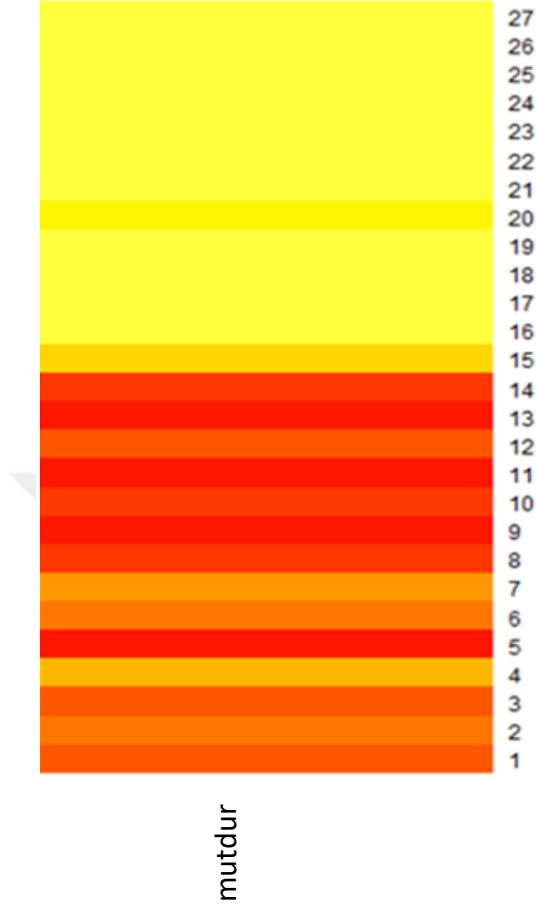
Mutasyonu belirli olan 27 hastanın genetik verilerine ilişkin yapılan analiz sonuçları Şekil 4.19.'da verilmiştir. Analizler sonucunda 2 ana grup oluşmuştur ve gruplar klinik ve demografik açıdan incelenmiştir. Genetik açıdan, 13 hastanın bulunduğu 1. Grubun tamamının SOD1 mutasyonunu gösterdiği tespit edilmiştir. Sadece 21 numaralı hasta SOD1 mutasyonunun yanı sıra C9ORF72 mutasyonunu da göstermiştir. 2. Grupta ise

SOD1 mutasyonunu gösteren hastalar olsa da 2, 3 hatta 4 mutasyonu da gösteren hastalar bulunmaktadır. Klinik ve demografik açıdan incelendiğinde ise hastalar arasında farklılık tespit edilememiştir.



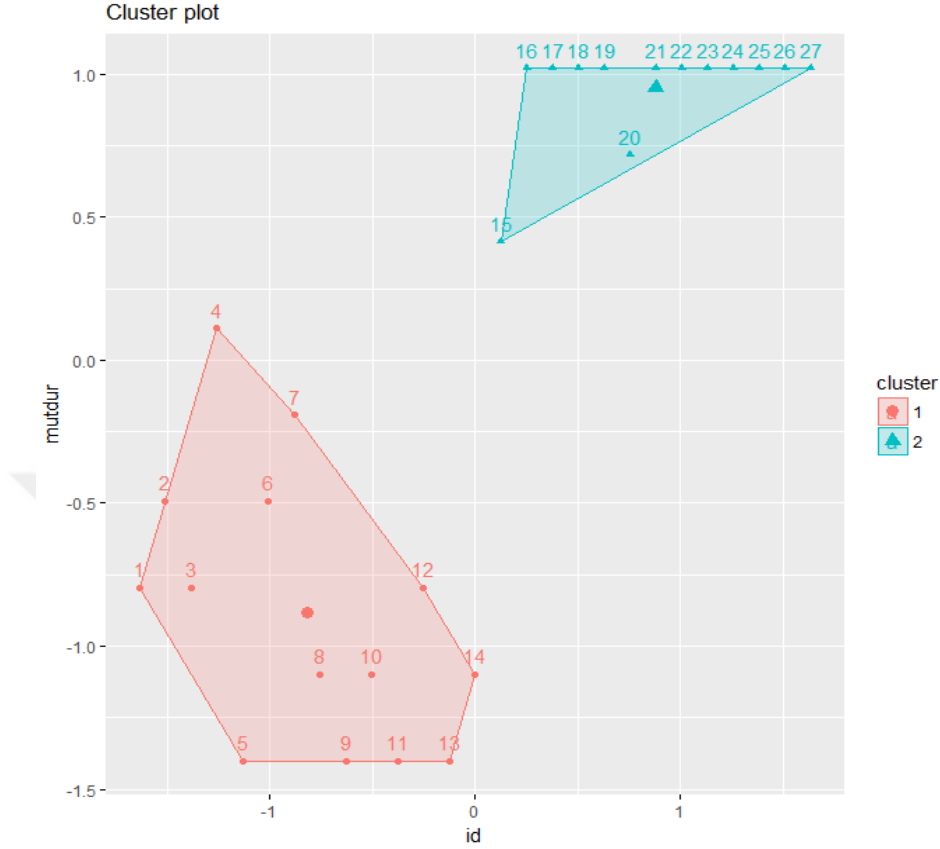
**Şekil 4.19.** 27 hastanın genetik verilerinin dendrogram görüntüsü

Mutasyonu tanımlanmış olan hasta grubuna ait genetik veriler incelendiğinde oluşan ısı haritası Şekil 4.20’de görülmektedir. Şekil incelendiğinde 5,9,11,13 nolu hastalar genetik özelliğinin ortalamasına göre daha çok benzerlik göstermişlerdir. Hastalara ait veri setinde genetik olarak sadece mutasyon tipine yönelik veri bulunduğu için tek bir özellik açısından ısı haritası oluşmuştur.



**Şekil 4.20.** 27 hastanın genetik verilerinin ısı haritası (heatmap) görüntüsü

27 hastaya ait genetik verilerin analiz edilmesi sonucu oluşan dendogramlar neticesindeki kümeleme grafiği ve gruplar arasındaki kesişim alanı incelendiğinde grupların birbirinden ayrıştığı Şekil 4.21.'de görülmektedir. Dendogram açıklamalarında da belirtildiği üzere birinci grupta bulunan hastalar arasında 1 hasta hariç hepsinde sadece SOD1 mutasyonu görülmesi ve incelenecek bir başka değişkenin olmaması grafiğin kesişim alanının oluşmamasına neden olmuştur.

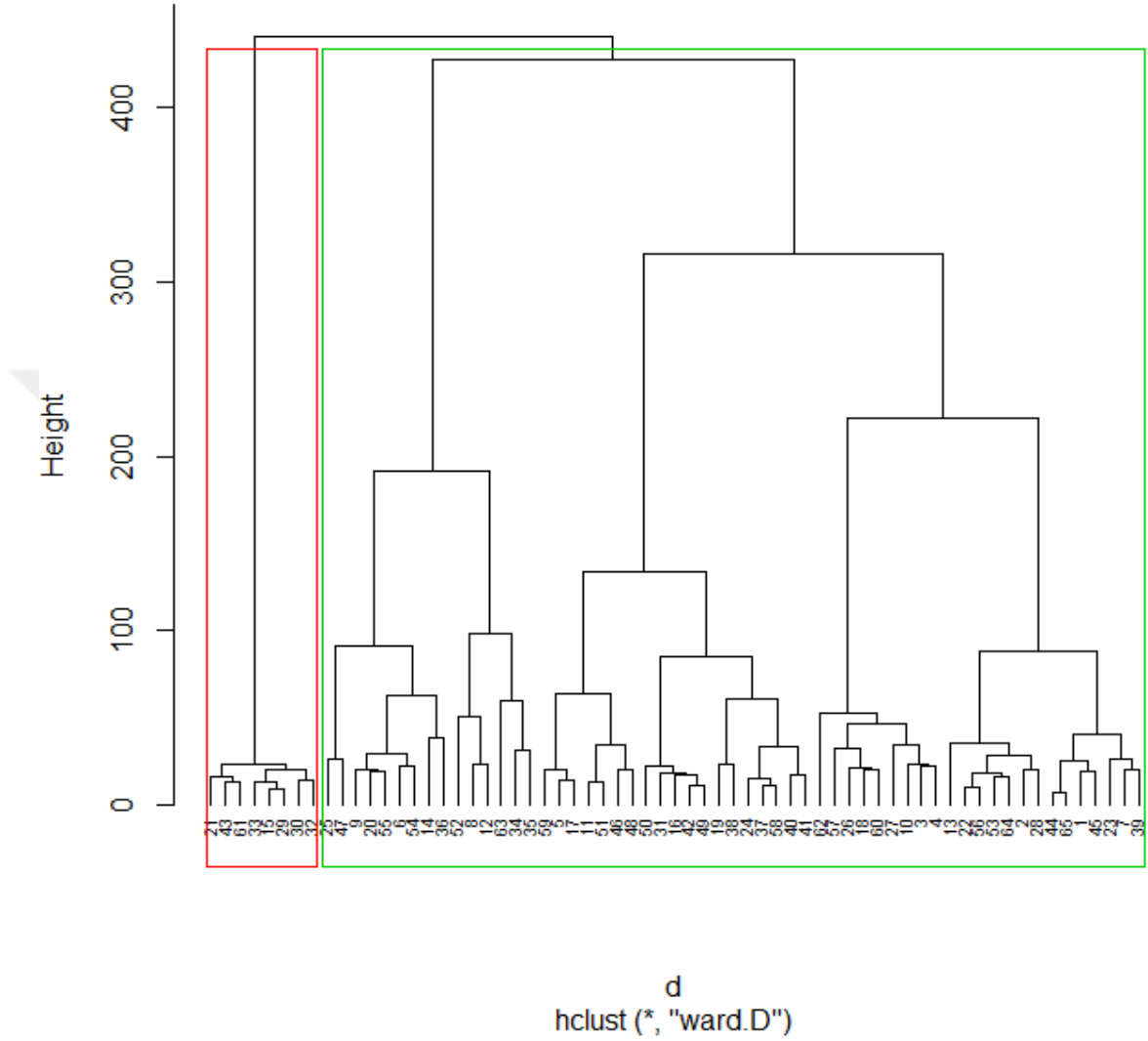


Şekil 4.21. 27 hastanın genetik verilerinin kümeleme grafiği görüntüsü

### 4.3. Toplam Hasta Grubu

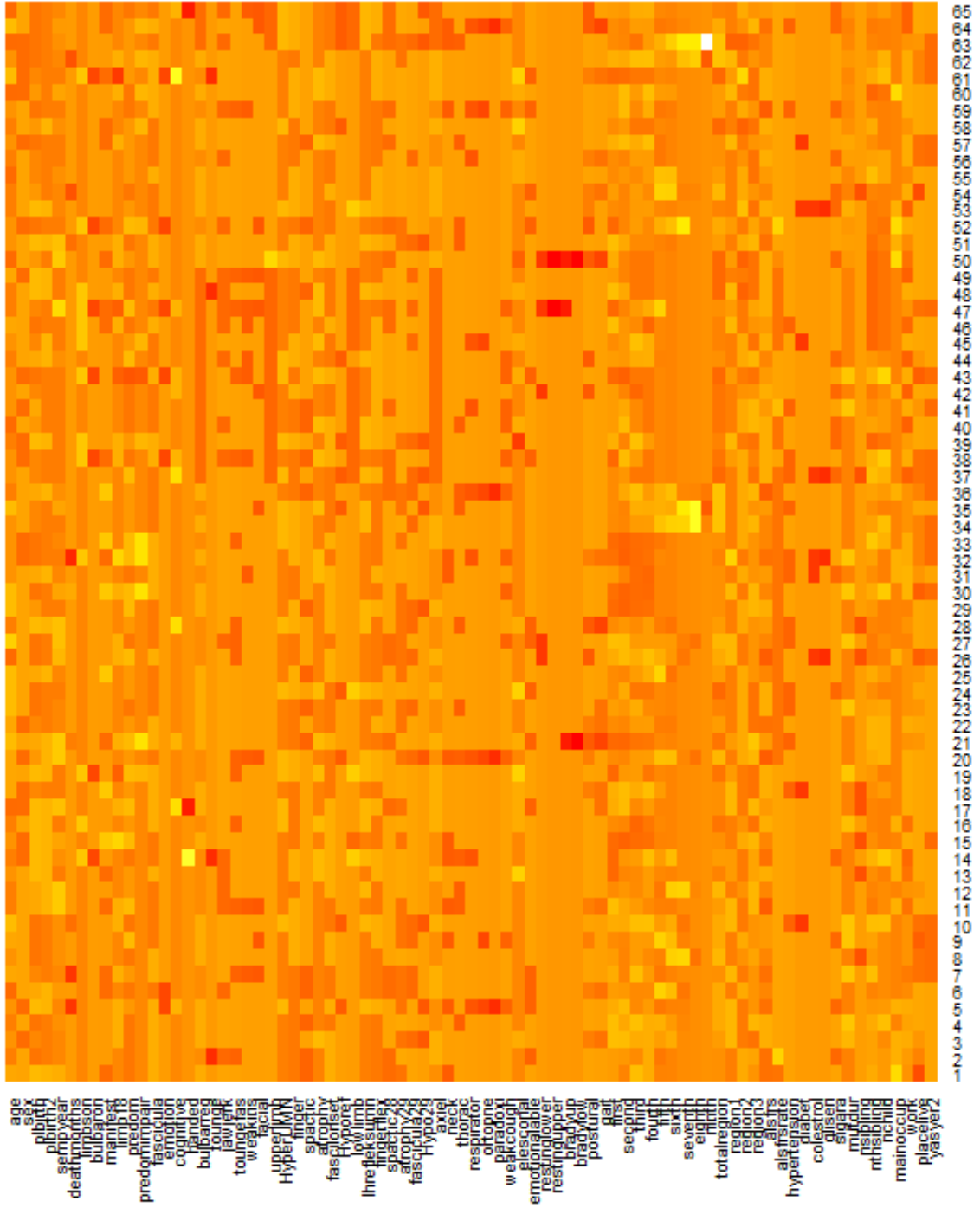
Son olarak toplam hasta grubu için genel gruplandırmayı görmek amacı ile demografik, klinik ve genetik bilgiler ayrımı gözetilmeksizin R Programında Hiyerarşik Kümeleme kullanılarak analiz yapılmıştır. Analiz sonuçları Şekil 4.22. , 4.23. ve 4.24 'de verilmiştir. Kümeleme analizi dendogram sonucu Şekil 4.22'de verildiği üzere, 8 üyeden oluşan 1. Grup ve 57 üyeden oluşan 2. Grup olmak üzere 2 gruptan oluşmuştur. Gruplar özelinde hastaların özellikleri incelendiğinde 1. Grupta bulunan hastaların tamamının klinik özelliklerinde çene hiperrefleksisi, istemsiz yüz kas kasılması, hiporefleksi, ortopne, paradoksik solunum, alt ve üst ekstremitte istiharat tremoru, tip I/II diyabet, sigara kullanımı yoktur. Ayrıca bu grupta yer alan hastalarda hastalığın yayıldığı bölge sayısı en fazla 5'dir. 2. Grupta ise üst ekstremitte istiharat tremoru olmayan hasta oranı %92 iken, alt üst ekstremitte istiharat tremoru olmayan hasta oranı %96'dır. Bununla birlikte alt ekstremitte hareket yavaşlığı olmayan hastaların oranı %98 iken üst ekstremitte hareket yavaşlığı bulunmama oranı %96 olarak tespit edilmiştir.

## Cluster Dendrogram



Şekil 4.22. 65 hastanın tüm verilerinin dendrogram görüntüsü

Son olarak incelenen grup her iki grubun da dâhil olduğu 65 hastadan oluşan toplam hasta grubudur. Bu grubun ısı haritası Şekil 4.23’de verilmiştir. Isı haritasında 65 hastaya ait tüm özelliklerine ait sayısal değerlerinin renklere dönüştürülmüş şekli görülmektedir. Bu görsel değerlendirildiğinde özelliklerin neredeyse tamamında ortalamaya kıyasla azalış eğiliminde olduğunu söylemek mümkündür. 48 ve 51 numaralı hastanın alt ve üst ekstremitelerde istirahat tremoru, alt ve üst ekstremitelerde hareket yavaşlığı değerlerinin ortalamaya kıyasla artışta olduğu görünse de tüm şekilde önemli bir artış veya azalış bulunmamaktadır.

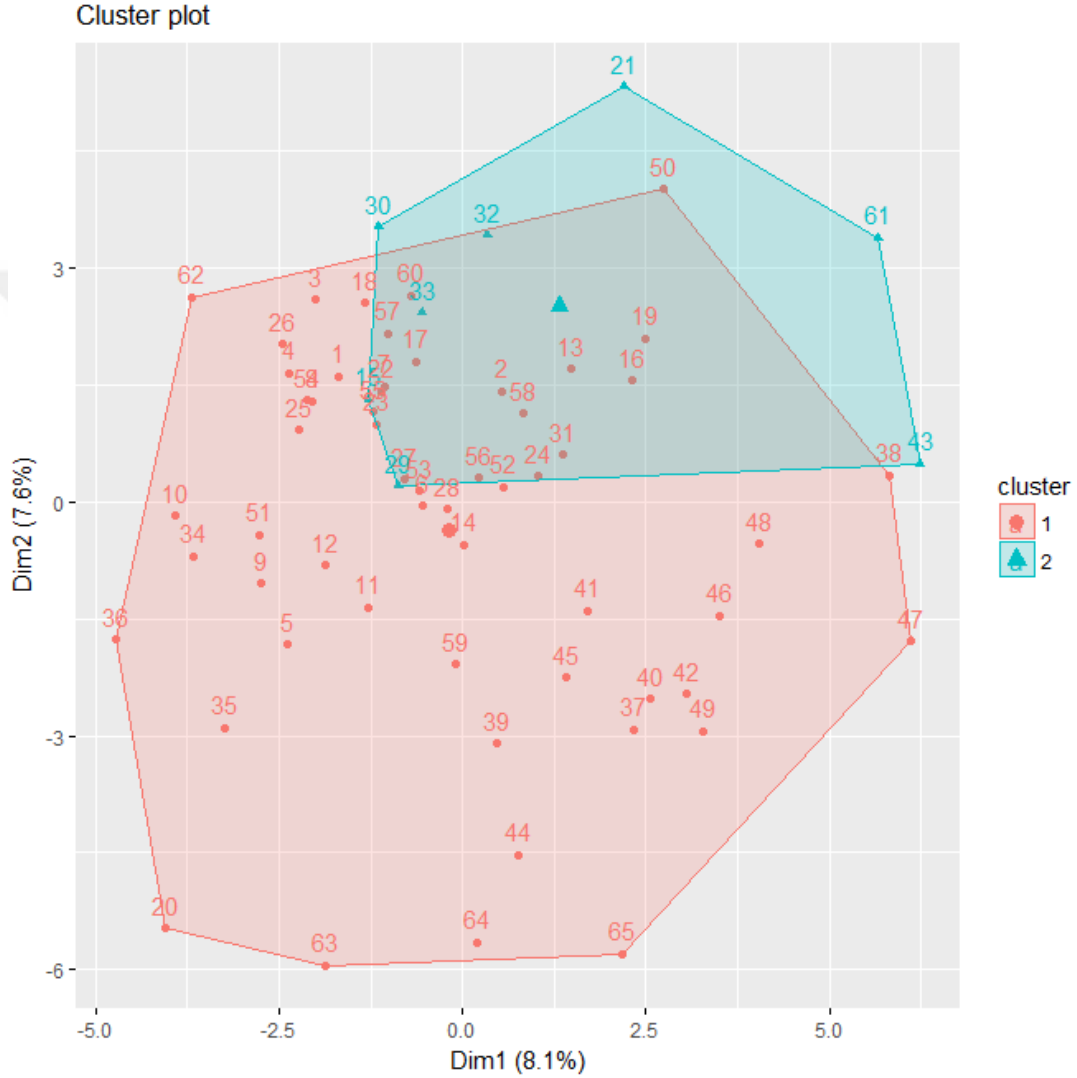


Şekil 4.23. 65 hastanın tüm verilerinin ısı haritası (heatmap) görüntüsü

Toplam hasta grubuna ait verilerin analiz edilmesi sonucu oluşan dendogramlar neticesindeki kümeleme grafiği ve gruplar arasındaki kesişim alanı incelendiğinde hastaların özellikler açısından birbirinden önemli ölçüde ayrılmadığı ve Şekil 4.24.'de



görüldüğü üzere hastaların sayıca çoğunluğunun bir alanda toplandığı görülmektedir. Bu nedenle klinik, demografik ve genetik özellikleri açısından hastalar arasında anlamlı bir farklılık tespit edilmemiştir.



Şekil 4.24. 65 hastanın tüm verilerinin kümeleme grafiği görüntüsü

#### 4.3.1. Toplam Hasta Grubu – Ki kare Analizi

Üzerinde çalışılan 65 kişinin oluşturduğu veri seti ağırlıklı olarak kategorik veriden oluşmaktadır. Ki kare analizi küçük hacimli örneklem sayısının olduğu, testte kullanılacak olan değerlerin yerine sıra numaralarının oluşturulduğu durumlarda sıkça kullanılan, anlaşılması ve uygulanması kolay bir istatistiksel analiz türüdür (Şehirlioğlu & Sarıbay, 2012). İki değişken arasında ilişkinin var olup olmadığını sistematik olarak incelemek için

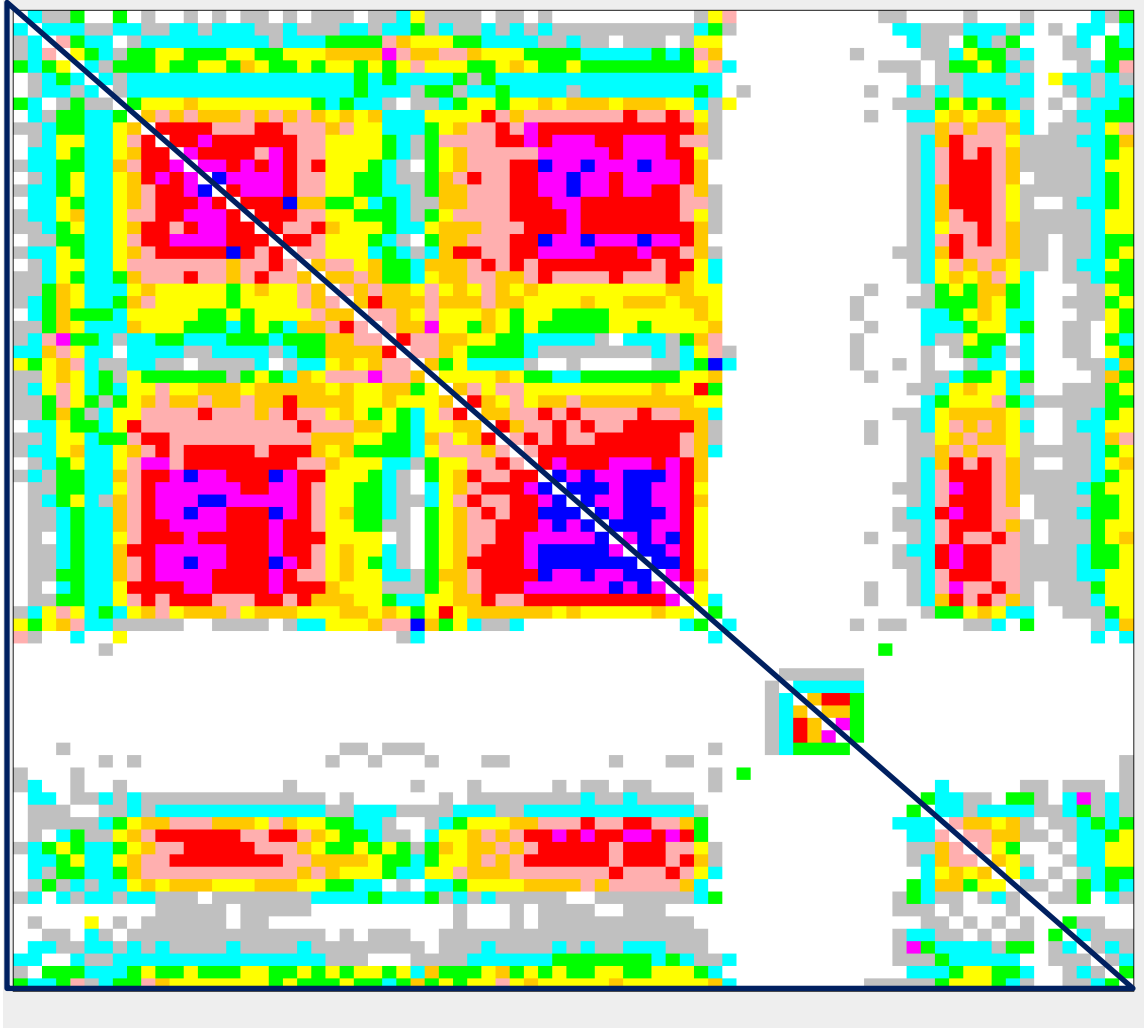
çalışmada Ki kare analizi uygulanmıştır. Ki kare analizi sonuçlarına göre anlamlı 242 ilişki oluşmuştur fakat bu özelliklerin sadece %10'u tez kapsamına alınmıştır. Yapılan analizin sonuçları Tablo 4.1.'de verilmiştir.

**Tablo 4.1.** Ki kare sonuç tablosu

Sıra No	Özellik 1	Özellik 2	Değer
1	yasyer2	plbirth	p=0.000
2	yasyer2	plbirth2	p=0.001
3	yasyer2	fascicula29	p=0.042
4	yasyer2	nthsibling	p=0.013
5	sex	region1	p=0.008
6	sex	work	p=0.000
7	plbirth	yasyer2	p=0.000
8	plbirth	plbirth	p=0.000
9	plbirth	fascilonset	p=0.035
10	plbirth	lhrefleksumn	p=0.023
11	plbirth	third	p=0.036
12	plbirth	region3	p=0.043
13	plbirth	mainoccup	p=0.040
14	plbirth	placelive	p=0.007
15	plbirth2	yasyer2	p=0.001
16	plbirth2	plbirth	p=0.000
17	plbirth2	lhrefleksumn	p=0.027
18	plbirth2	placelive	p=0.007
19	sempyear	limbson	p=0.024
20	sempyear	atrophy29	p=0.026
21	sempyear	thorac	p=0.016
22	sempyear	fifth	p=0.008
23	sempyear	sixth	p=0.001
24	sempyear	totalregion	p=0.000

#### 4.3.2 Toplam Hasta Grubu – Değişken Matrisi

Şekil 4.25'de gösterilen Hasta Yoğunluk Analizinin 2 boyutu hastaya ait değişkenleri göstermektedir. Bu da her kutuda hastaların yüzde kaçının aynı özelliklere ait olduğunu göstermektedir. Benzerlik oranı en fazla %100 en az %0'dır. Bu matris, hastaların özelliklerinin birbirine ne kadar benzedikleri konusunda görsel olarak genel bir fikir vermektedir.



Şekil 4.25. Değişken matrisi hasta yoğunluk analizi

Şekil 4.25’de verilen değişken matrisi tablosundaki gözlemlenen beyaz ve boş olan alanlar değişkenler açısından, benzerliğin az olduğunu veya olmadığını göstermektedir. Özelliklerin birbirleri ile ilişkilerinin incelendiği yöntemde %50’in üzerinde bulunan 786 ilişki oluşmuştur. Bu özelliklerden sadece benzerlik oranları %90’ın üzerinde olanlar Tablo 4.2.’de verilmiştir.

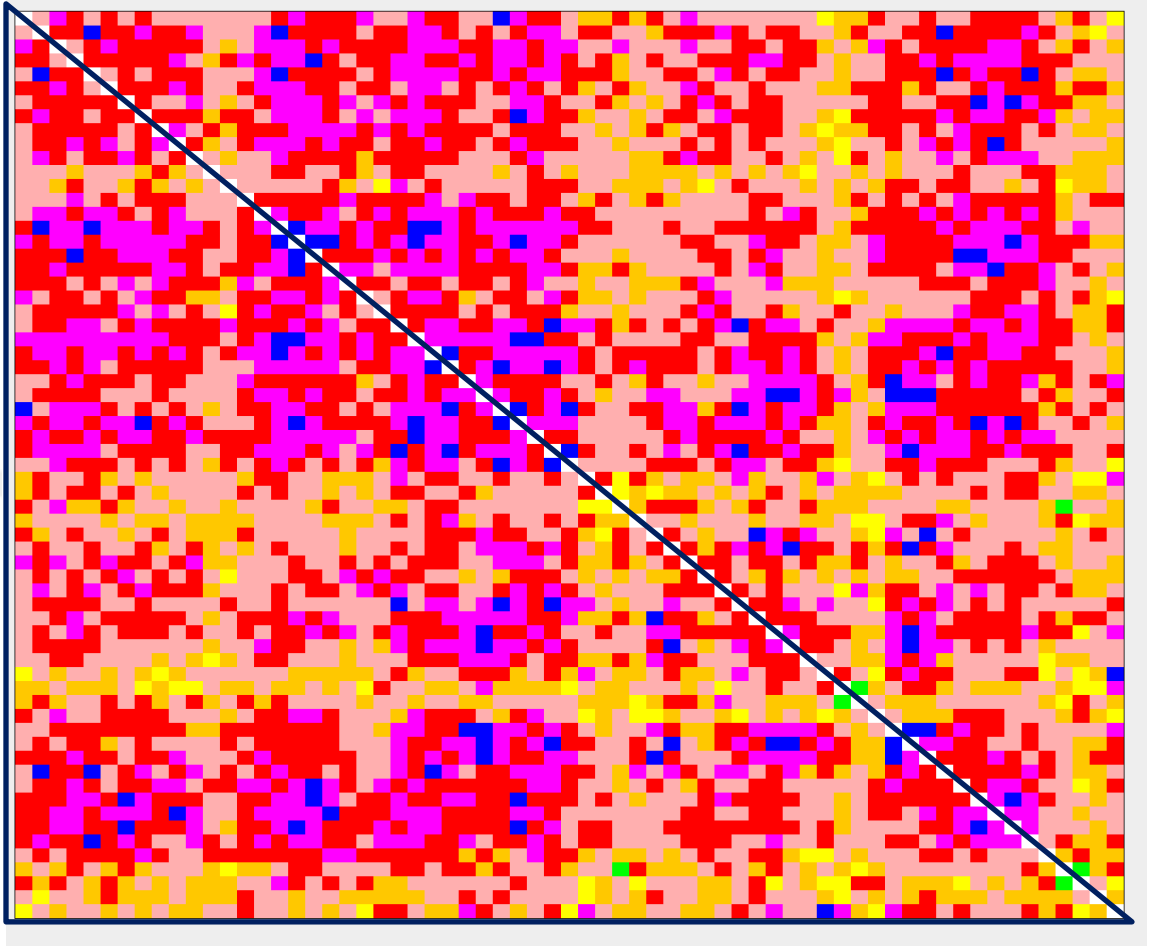
**Tablo 4.2.** Değişken matrisi hasta yoğunluk analizi sonuç tablosu

Sıra No	Özellik 1	Özellik 2	Değer
1	restingupper	bradyup	0.985
2	gliseri	colestrol	0.985
3	bradylow	bradyup	0.985
4	restingupper	bradylow	0.969
5	ortopne	paradoxi	0.954
6	restingupper	restinglower	0.954
7	bulbaron	emotion	0.938
8	bradyup	restinglower	0.938
9	bradylow	restinglower	0.923
10	handed	bradylow	0.923
11	handed	restingupper	0.923
12	ortopne	respirator	0.923
13	paradoxi	bradylow	0.908
14	gliseri	restingupper	0.908
15	handed	bradyup	0.908
16	handed	tounge	0.908
17	tounge	bradylow	0.908
18	tounge	restingupper	0.908
19	gliseri	bradylow	0.908
20	respirator	paradoxi	0.908
21	paradoxi	restingupper	0.908

Gruplar arasındaki benzerliklerin klinik veriler ile ilişkili olduğunu bu tablo ile gözlemlemek mümkündür. Tablo 4.2’de trigliset yüksekliği ve hiperkolesterolemi dışında görülen refleks değişiklikler, bulbar tutulum, kas güçsüzlüğü, fasikülasyon olup olmadığı, patolojik refleks gibi bulguların hepsi klinik bulguları oluşturmaktadır. Bu nedenle hastalığın belirleyiciliğini klinik bulgular oluşturmaktadır ve diğer faktörler ile ilişki kurulması için klinik bulgular olmazsa olmaz etmenlerdir.

#### **4.3.3. Toplam Hasta Grubu – Hasta Matrisi**

Şekil 4.26’da gösterilen Değişken Yoğunluk Analizinin 2 boyutu hastaları göstermektedir. Bu da her kutuda hastaların yüzde kaçının birbirine benzer olduğunu göstermektedir. Benzerlik oranı en fazla %100 en az %0’dır. Bu matris, hastaların birbirine ne kadar benzedikleri konusunda görsel olarak genel bir fikir vermektedir.



**Şekil 4.26.** Hasta matrisi değişken yoğunluk analizi

Hastalar arasındaki benzerliklerin görselleştirildiği matriste yer alan hasta ve hastaların benzerlik değerleri Tablo 4.3.'de verilmiştir. Matris sonucu %50'nin üzerinde ilişki gösteren 1502 değer bulunmaktadır. Hastalar arası benzerlik değerleri en yüksek %70 değerindedir. Ortaya çıkan bu değerler Tablo 4.3.'de verilmiştir. Bu değerler hastalar arasında yüksek bir ilişki olmadığını açıkça göstermekle birlikte hastalar ikişerli grup olarak incelendiğinde, hastaların ağırlıklı olarak klinik bulguları doğrultusunda benzedikleri söylenebilir.

**Tablo 4.3.** Hasta matrisi deęişken yoğunluk analizi sonuç tablosu

Sıra No	Hasta 1	Hasta 2	Deęer
1	p_17	p_20	0.772
2	p_1	p_26	0.759
3	p_17	p_14	0.759
4	p_56	p_23	0.759
5	p_59	p_25	0.759
6	p_39	p_44	0.747
7	p_41	p_42	0.747
8	p_50	p_49	0.747
9	p_34	p_16	0.747
10	p_55	p_9	0.734
11	p_47	p_49	0.734
12	p_41	p_49	0.734
13	p_66	p_45	0.734
14	p_61	p_3	0.734
15	p_24	p_26	0.734
16	p_43	p_50	0.722
17	p_19	p_3	0.722
18	p_23	p_26	0.722
19	p_46	p_49	0.722
20	p_17	p_59	0.722
21	p_14	p_20	0.722
22	p_19	p_23	0.722
23	p_61	p_56	0.722
24	p_43	p_66	0.709
25	p_30	p_24	0.709
26	p_31	p_34	0.709
27	p_19	p_58	0.709
28	p_17	p_54	0.709
29	p_24	p_23	0.709
30	p_43	p_41	0.709
31	p_1	p_32	0.709
32	p_16	p_30	0.709
33	p_13	p_26	0.709
34	p_40	p_49	0.709
35	p_34	p_30	0.709
36	p_30	p_26	0.709
37	p_40	p_46	0.709
38	p_2	p_1	0.709
39	p_50	p_47	0.709

## 5. TARTIŞMA

Amiyotrofik Lateral Skleroz (ALS) Hastalığının Genetik ve Klinik Veri İlişkisinin Veri Madenciliği Yöntemleri İle İncelenmesi başlıklı çalışmamızın evreni Akdeniz Üniversitesi Nöroloji Polikliniği'ne farklı şikayetler ile başvurmuş ve yapılan incelemeler sonucunda ALS teşhisi almış hastalardan oluşmuştur. 'Amiyotrofik Lateral Skleroz Motor Nöron Hastalığı'nın farkındalığının artırılması amacıyla faaliyet gösteren ALS-MNH Derneği, Türkiye'de toplam bilinen 6000 ile 8000 civarında ALS hastası olduğunu belirtmekle birlikte Türkiye'deki ALS hastalarının sayısının tam olarak bilinmediğini de yayınlamaktadır (ALS MNH Derneği, 2018). Türkiye'de ALS hastalığının görülme sıklığının ve hastalığın ortaya çıkma nedenlerinden olan genetik faktör oranının düşük olması nedeni ile çalışmaya dâhil edilebilecek hasta grubu sayısının az olması çalışmanın sonuçları için kısıt oluşturmuştur. Bu kısıtı gidermek için çalışma sürecinde yeni tanı alan veya mutasyon tespiti yapılan hastalar çalışmaya dâhil edilse de bu durum sonuçların genellemesine imkân vermemiştir.

Bu çalışmada bir grup verinin birden fazla küme özelliğine göre hangi kümeye ait olduğunun tespit edilmesi ve bilgi çıkarımı (knowledge extraction) yapılabilmesi için Hiyerarşik Kümeleme yöntemi kullanılmıştır. Yöntem mutasyonu tanımlı, mutasyonu tanımsız ve toplam hasta grubu olmak üzere 3 grup için kullanılmıştır. ALS'de mutasyonun rolünün hala tartışılan bir konu olması nedeni ile mutasyonun etkisini inceleyebilmek için öncelikle mutasyonu belirlenmiş hastalar ayrı bir grup olarak incelenmiştir. Sonrasında da mutasyonu belirli ve belirsiz olan hastalar bir arada ayrıca incelenmiş ve elde edilen sonuçlar arasındaki anlamlı ilişkiler incelenmiştir.

Arslan, Çolak ve Sarihan tarafından 2016 yılında yapılmış oldukları çalışma sonucunda tıbbi veri madenciliği veya tıbbi bilgi keşfinin bilgisayar destekli yaklaşımların hastalık öngörülmesinde etkili bir araç olduğunu ve veri kümelerindeki gizli ilişkileri araştırdığını ortaya koymuştur. (Arslan ve ark., 2016). Bunun ile birlikte Köktürk ve arkadaşlarını yapmış oldukları çalışmalarında veri madenciliği alanında yaygın olarak, grafik ve şekillerin incelenmesi aracılığıyla değerlendirme yapıldığını ve değişkenler arasındaki anlamlı bağlantılar tespit edilerek veri tanımlandığını belirtmişlerdir (Köktürk ve ark.,

2009). Çalışmamızda kullanılan dendogram, ısı haritası ve kümeleme grafikleri sonuçları görselleştirmede yoğun olarak kullanılmış ve veri değerlendirme sürecini kolaylaştırmıştır.

Zou ve arkadaşlarının yapmış olduğu çalışmada ailesel ALS'nin yanı sıra sporadik ALS'de ağırlıklı olarak SOD1 mutasyonu gösterdiği belirtilmiştir (Zou ve ark., 2016). Benzer şekilde, gerçekleştirilen tez çalışmasında mutasyonu tanımlı hasta grubundan 11 tanesinin sadece SOD1 mutasyonu göstermesi bu çalışmayı desteklemektedir. Bunun ile birlikte SOD1 mutasyonun gösteren hastaların klinik olarak diğer hastalardan ayırt edilememesi ve tanımlayıcı özelliklerinin belirlenememesi hastalığın kompleks bir hastalık olduğunu göstermektedir (Krakauer ve ark., 2002).

ALS hastalığına etki eden faktörlerde bahsedilen bir diğer başlık çevresel faktörlerdir. Alonso ve arkadaşları ile Oskarsson ve arkadaşlarının yapmış olduğu çalışmada sigaranın ALS riskini arttırdığı düşüncesi desteklenmektedir fakat yapmış olduğumuz çalışmada ALS hastalığı ile sigara kullanımı arasında bir ilişki tespit edilememiştir (Alonso ve ark., 2010; Oskarsson ve ark., 2016). Benzer şekilde literatürde fiziksel aktivitenin ALS hastalığı altında yatan nedenlerden olduğu yer almaktadır ancak çalışmamızda fiziksel aktivite ile hastalık arasında bir ilişki saptanamamıştır. Fakat hasta veri seti incelendiğinde, hastaların büyük bir kısmının beden aktivitesi yoğun mesleklerde çalıştığı da göze çarpmıştır.

R programında gerçekleştirilen hiyerarşik kümeleme analizlerinin yanı sıra hastalar ve özellikler arasındaki ilişkiyi tespit etmek amacı ile Java programlama dili kullanılarak Ki kare analizi uygulanmıştır. Bu analizin, gruplardaki gözlem sayılarının az olduğu durumlarda kullanılmasının uygun olduğu Şehirlioğlu ve Sarıbay'ın çalışmasında belirtilmiştir (Şehirlioğlu ve Sarıbay, 2012). Bulgular başlığı altında verilen Ki kare analizi ile elde edilen sonuçların uzman hekim tarafından değerlendirilmesine istinaden gözlem sayısının az olduğu durumlarda kullanılan bu analiz yönteminin uygunluğu desteklenmiştir.

R programının yanı sıra benzerlik ilişkilerini incelemek için Java programlama dili kullanılarak hasta yoğunluk ve özellik yoğunluk matrisi sonuçları elde edilmiştir. Bu



ilişkiler incelendiğinde; elde edilen benzerliklerin demografik veya genetik bulgulardan ziyade klinik bulgular bakımından olduğu tespit edilmiştir. Bu durum hastalık teşhisinde de kullanılan klinik verilerin belirleyici olduğunu, genetik ve demografik verilerin belirleyici olmadığını göstermiştir. Klinik uzman tarafından yapılan değerlendirme ve literatür bilgisi doğrultusunda, gruplandırmalar ve ilişkiler arasında literatürü destekleyici ilişkileri ortaya çıkardığı görülmüştür.

Bunun ile birlikte Türe ve arkadaşlarının yapmış olduğu çalışmada, Karar Ağacı yönteminin bir karara varmak için çeşitli adımlardan oluşan sistemlerde bu adımların olasılıksal değerlerini bir çizge üzerinde ağaç kullanarak göstermek amacı ile büyük veri tabanlarında sınıflandırma problemlerinin çözümü için karmaşık bilgilerde yararlı bir çözüm olduğu vurgulanmıştır (Türe ve ark., 2009). Fakat çalışmamızda R programı kullanılarak yapılan analizler neticesinde tüm veri setlerinde karar ağacında tek bir dal oluşmuştur. Farklı sayılarda eğitim veri seti kullanılsa da sonuç değişmemiştir ve uygulamada kullanılan veri seti üzerinde bu yöntemin anlamlı bir sonuç üretmediği gözlenmiştir.

## 6. SONUÇ VE ÖNERİLER

Bu çalışmada, Amiyotrofik Lateral Skleroz (ALS) Hastalığı bulunan kişilere ait klinik, demografik ve genetik özellikler arasındaki ilişkinin veri madenciliği yöntemleri ile incelemesi yapılmıştır.

Çalışmada, açık kaynak kodlu R programı ve Java programlama dili hakkında bilgiler verilmiş, çalışmada kullanılacak yöntemler tanıtılmış ve Akdeniz Üniversitesi Hastanesi Nöroloji Kliniği öğretim üyesi tarafından yürütülen Avrupa Birliği projesi kapsamında hastalardan elde edilen veriler üzerinde, seçilen yöntemlerin uygulaması R programı ile gerçekleştirilmiştir. Gerçekleştirilen analizler neticesinde elde edilen çıktılarda farklılıklar ve benzerlikler üzerinde durulmuştur.

Kullanılan veri seti, her biri 79 özellikten oluşan, mutasyonu tanımlanmış 27 hasta, mutasyonu henüz tanımlanmamış 38 hasta olmak üzere toplam 65 hastadan oluşmaktadır. Önerilen yöntemler 65 hasta üzerinde çalışılmıştır.

Hiyerarşik kümeleme yöntemi ile yapılan analizler sonucunda elde edilen bulgulara ait özet bilgiler Tablo 6.1.'de (Sayfa 78-80) verilmiştir. Bu bilgiler ortak görülen özelliklerin tek bakışta değerlendirilebilmesi için hazırlanmıştır ve gruplarda belirtilen özelliklerin benzerlik oranları Bulgular başlığı altında detaylı olarak verilmiştir.

**Tablo 6.1.** Kümeleme analizi bulguları özet tablosu

		<b>1.Grup Özellikleri</b>	<b>2.Grup Özellikleri</b>	<b>2 Grubun Kesişim Alanı Özellikleri</b>
<b>Mutasyonu Tanımlanmamış 38 Hasta</b>	<b>Tüm Veriler</b>	Kol ve bacak başlangıçlı, Duygusal kararsızlık yok, Baskın taraf distal, Seğirme var, Bilişsel belirti yok, Dil spastisitesi yok, Maksimum yayılan bölge sayısı 9, Alt ve üst istiharat tremoru yok, Alt ve üst ekstremitede hareket yavaşlığı yok,	Kol ve bacak başlangıçlı, Alt ekstremitede üst istiharat tremoru yok, Maksimum yayılan bölge sayısı 7, Kentsel alanda yaşamakta, Hiperkolesterolemi yok, Trigliserit yüksekliği yok, Tip I/II diyabet yok, Hiperkolesterolemi yok,	Alt ve üst ekstremitede istiharat tremoru yok, Alt ve üst ekstremitede hareket yavaşlığı yok
	<b>Demografik Veriler</b>	Kentsel alanda yaşamakta, Kol ve bacak başlangıçlı, Dil spastisitesi yok, Hiporefleksi yok, Boyun zayıflığı yok, Torasik kaslarda seyirme yok, Duygusal kararsızlık yok, Tip I/II diyabet yok, Hiperkolesterolemi yok, Trigliserit yüksekliği yok, Egzersiz yok	Baskın taraf distal, Üst ekstremitede	Hepsi erkek, 1 hasta hariç hepsi kentsel alanda yaşıyor, 1 hasta hariç hepsi büyük şehirde yaşıyor
	<b>Klinik Veriler</b>	Kol ve bacak başlangıçlı, Duygusal kararsızlık yok, Baskın taraf distal, Seğirme var, Bilişsel belirti yok, Dil spastisitesi yok, Maksimum yayılan bölge sayısı 9, Alt ve üst istiharat tremoru yok, Alt ve üst ekstremitede hareket yavaşlığı yok,	Kol ve bacak başlangıçlı, Alt ekstremitede üst istiharat tremoru yok, Maksimum yayılan bölge sayısı 7 Kentsel alanda yaşamakta Hiperkolesterolemi yok, Trigliserit yüksekliği yok, Tip I/II diyabet yok, Hiperkolesterolemi yok,	Maksimum yayılan bölge sayısı 7, Başlangıçta duygusal kararsızlık yok, Alt ve üst istiharat tremoru yok, Alt ve üst ekstremitede hareket yavaşlığı yok,

**Tablo 6.1 (Devamı).** Kümeleme analizi bulguları özet tablosu

		<b>1.Grup Özellikleri</b>	<b>2.Grup Özellikleri</b>	<b>2 Grubun Kesişim Alanı Özellikleri</b>
<b>Mutasyonu Tanımlanmış 27 Hasta</b>	<b>Tüm Veriler</b>	Ortopne yok, Paradoksik solunum yok, Dil spastisitesi yok, Boyun zayıflığı yok, Tip I/II diyabet yok,	Dil spastisitesi yok, Başlangıçta bilişsel belirti yok, Paradoksik solunum yok, Hiperkolesterolemi yok, Trigliserit yüksekliği yok, Kol ve bacak başlangıçlı, Alt ve üst ekstremitede atrofi ve güçsüzlük yok, ALSFRS-R düşme oranı 0,1-0,9 aralığı	Hepsi kentsel alanda yaşıyor, Bilişsel belirti yok, Alt ve üst istirahat tremoru yok, Alt ve üst ekstremitede hareket yavaşlığı yok, Üst motor nöronda hiporefleksi var
	<b>Demografik Veriler</b>	SOD1 mutasyonuna sahip, Sigara kullanımı yok	ALSFRS-R düşme oranı 0,1-0,9 aralığı	Hepsi kırsal alanda yaşıyor, Alt ve üst motor nöronda hiporefleksi var, Dilde fasikülasyon yok, Çene refleksi yok, Yüz kaslarında fasikülasyon yok, Ortopne yok, Atrofi ve güçsüzlük yok,
	<b>Klinik Veriler</b>	Ortak özellik yok	Ortak özellik yok	1 hasta hariç alt ve üst istirahat tremoru yok, 1 hasta hariç alt ve üst ekstremitede hareket yavaşlığı yok
	<b>Genetik Veriler</b>	1 hasta hariç tamamı SOD1 mutasyonuna sahip,	Ortak özellik yok	Tamamı SOD1 mutasyonuna sahip

**Tablo 6.1 (Devamı).** Kümeleme analizi bulguları özet tablosu

		<b>1.Grup Özellikleri</b>	<b>2.Grup Özellikleri</b>	<b>2 Grubun Kesişim Alanı Özellikleri</b>
<b>Toplam 65 Hasta</b>	<b>Tüm Veriler</b>	Çene refleksisi yok, İstemsiz yüz kas kasılması yok, Hiporefleksi yok, Ortopne yok, Paradoksik solunum yok , Alt ve üst ekstremitte istirahat tremoru yok, Tip I/II diyabet yok, Sigara kullanımı yok	Alt ve üst istirahat tremoru yok Alt ve üst ekstremitte hareket yavaşlığı yok	Ortak özellik yok

Kümeleme analizleri sonucunda;

- Mevcut veri setindeki grupların mutasyonu tanımlı veya mutasyonu tanımlı olmayan olarak değil SOD1 mutasyonu olanlar ve olmayanlar olarak gruplandırılabilceği tespit edilmiştir.
- SOD 1 mutasyonu tanımlanmış olan hastalar ile farklı mutasyonların tespit edildiği hastalar arasında klinik ve demografik bulgular açısından ayırıcı bir özellik tespit edilmemiştir.
- Literatürde sigara önemli bir faktör olarak belirtilirken çalışmamızda sigara kullanımı belirleyici bir faktör olarak tespit edilmemiştir. Elde edilen bu sonucun veri setinde sigara kullanan hasta sayısının azlığından kaynaklandığı düşünülmektedir.
- Literatürde ALS hastalığının erkeklerde yaygın görüldüğü belirtilmiştir, çalışmamızda da hasta grubu erkek ağırlıklı olmasına rağmen erkeklerin tutulumu ile kadınların tutulumu arasında bir fark tespit edilmemiştir.
- Günlük yaşam aktivitesi veya egzersiz açısından belirleyici bir farklılık gözlemlenememiştir. Sigara kullanımına benzer şekilde düzenli egzersiz yapan

hasta sayısının çoğunluğunun bilinmemesi bu sonucun elde edilmesine neden olduğu düşünülmektedir.

Veri madenciliği sınıflandırma modelleri arasında yer alan büyük ve karmaşık veri setleri üzerinde çalışılmasına olanak sağlayan karar ağaçları yöntemi ile yapmış olduğumuz çalışmada herhangi bir sonuç elde edilememiştir. Bunun nedeninin öğrenme kümesinde bulunan örnek sayısının yetersizliği, veri setindeki özelliklerin fazla olmasından kaynaklı olarak verinin genelleştirilmesi veya ALS hastalığının çok yönlü ve karmaşık bir hastalık olması olarak düşünülmektedir.

Ayrıca Java programlama dili kullanılarak geliştirilen araçlar ile elde edilen Ki kare ve yoğunluk matrisi sonuçlarına göre, istatistiksel olarak bazı anlamlı bilgiler bulunmuştur. Bu bilgiler doğrultusunda;

- Çalışmamızda demografik veriler ayırıcı bir faktör olarak tespit edilememiştir.
- Kullanılan yöntemler neticesinde ancak klinik bulguların ALS hastalığının anlamlılığını oluşturduğu söylenebilir.

Elde edilen tüm bu bilgiler ışığında, ALS hastalığı hem hasta hem de hasta yakınları açısından zor bir hastalıktır. Kısa sürede hastaların hayatını kaybetmesi nedeni ile erken teşhis konulması önem kazanmakla birlikte hastalığın teşhis edilmesi de oldukça zordur. Farklı çalışmalarda hastalığın ortaya çıkmasında klinik, genetik veya çevresel faktörlerin rolleri açıklanmaya çalışılsa da bu roller arasında doğrudan bir ilişki kurmak ve/veya bu ilişkileri tespit etmek oldukça zordur. Yine de çalışmamızda kullanılan ve geliştirilen metotlardan elde edilen bulgular, veri toplama süreci devam edip, verilerin sayısı arttıkça daha anlamlı sonuçlar elde edilebileceğinin bir göstergesidir.

Bu bilgiler doğrultusunda sonraki çalışmalar için öneriler:

ALS tanısı konmuş hastaların özelliklerinin hiyerarşik kümeleme ve karar ağacı yöntemleri ile incelendiği bu çalışmaya benzer şekilde diğer veri madenciliği yöntemleri de kullanılabilir.

Bu çalışma kapsamında kullanılan yöntemler, farklı özellikteki veri grupları ve farklı hastalık veri setleri üzerinde uygulanabilir.

Çalışmada R paket programı kullanılmıştır. Bu aracın dışında farklı veri madenciliği araçları/programları kullanılabilir ve/veya elde edilen çıktılar karşılaştırmalı olarak incelenebilir.

Bu çalışmada Java programlama dili ile geliştirilen metoda benzer yeni yöntemler geliştirilebilir ve bu yöntem ile farklı özellikteki veri grupları üzerinde çalışma yapılabilir.

Bunun ile birlikte yeni çalışmalarda ALS hasta veri seti kullanılarak, gerçek kişilerden alınan bilgilerin derlenmesi temeline dayanan ve zamanla kendisini geliştiren uzman sistem yazılımları kullanılabilir.

## KAYNAKLAR

- Ada M, Altunay F, Civelek M, Kaplan S, Koç P. Kümeleme Analizi. 2017: 1-8
- Ajrourd-Driss S, Siddique T. Sporadic and hereditary amyotrophic lateral sclerosis (ALS). *Biochimica et Biophysica Acta*. 2014;4: 679-684.
- Akküçük U. Veri Madenciliği: Kümeleme ve Sınıflama Algoritmaları. İstanbul: Yalın Yayıncılık; 2015, p:18
- Al-Azzawı MIM. The University Of Turkish Aeronautical Association Institute Of Science And Technology, Master Thesis, 2017, Ankara (Danışman: Assist. Prof. Dr. A Yuriy)
- Albayrak AS, Yılmaz K. Veri Madenciliği: Karar Ağacı Algoritmaları Ve İmkb Verileri Üzerine Bir Uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 2009;1: 31-52.
- Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Neurology*. 2013;11: 617-628.
- Alonso A, Logroscino G, Hernan M. Smoking and the risk of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2010;11: 1249-1252.
- ALS MNH Derneği. <https://www.als.org.tr/>: <https://www.als.org.tr/turkiye-de-als-mnh-hastaligi-adresinden-alindi>. 2018
- Altun M. Veri Madenciliği ve Uygulama Alanları Doktora Semineri Raporu. Akdeniz Üniversitesi Eğitim Bilimleri Bölümü EYTEPE ABD Doktora Programı. 2017: 1-27
- Antonio N, Almeida A, Nunes L. Predicting Hotel Booking Cancellations to Decrease Uncertainty and Increase Revenue. *Tourism & Management Studies*. 2017;2: 25-39.
- Argüden Y, Erşahin B. Veri Madenciliği. Ar-Ge Danışmanlık: <http://www.arge.com/wp-content/uploads/2013/02/VeriMadenciligi.pdf> adresinden alındı. 2008



Arslan A, Çolak C, Sarihan M. Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine*. 2016;130: 87-92.

Baydoğan M, Orbay B. R ile Programlamaya Giriş ve Uygulamalar. *İnternet Teknolojileri Derneği*: <http://inet-tr.org.tr/inetconf19/sunum/16.pdf> adresinden alındı. 2014

Bilge U, Bozkurt S, Durmaz S. Application of data mining techniques for detecting asymptomatic carotid artery stenosis. *Computers & Electrical Engineering*, 2010; 5: 1499-1505.

Bramer M. *Principles of Data Mining*. 2 nd ed. London: Springer-Verlag London; 2013, p:2-3

Chae Y, Kim H, Tark K, Park H, Ho S. Analysis of healthcare quality indicator using data mining and decision support system. *Expert Systems with Applications*. 2003;2: 167-172.

Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R. *CRISP-DM 1.0 Step-by-step data mining guide*. USA: SPSS Inc; 2000, p:10-30

Cheng P, Dummer R, Levesque M. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. *Medical intelligence*: 2015;145:1-5.

Chiò A, Benzi G, Dossena M, Mutani R, Mora G. Severely increased risk of amyotrophic lateral sclerosis among Italian professional football players. *Brain*. 2005;3: 472-476.

Clifton C. *Introduction to Data Mining*. <http://www.crisp-dm.org/CRISPWP-0800.pdf> adresinden alındı. 2004

Condradi S, Ronnevi LO, Vesterberg O. Lead concentration in skeletal muscle in amyotrophic lateral sclerosis patients and control subjects. *Journal of Neurology*.1978;11: 1001-1004.

Çam S. Veri Madenciliğinde Kümeleme Analizi ve Sağlık Sektöründe Bir Uygulama. *C.Ü. İktisadi ve İdari Bilimler Dergisi*. 2014;2: 85-96

Çataloluk H. Gerçek Tıbbi Veriler Üzerinde Veri Madenciliği Yöntemlerini Kullanarak Hastalık Teşhisi. B.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2012, Bilecik (Danışman: Yrd. Doç. Dr. K Metin)

Çelik Ş. Kümeleme Analizi İle Sağlık Göstergelerine Göre Türkiye'deki İllerin Sınıflandırılması. Doğu Üniversitesi Dergisi. 2013;2: 175-194.

Çetin E. Data Mining Techniques and A Banking Application. The Graduate School of Natural and Applied Sciences Industrial Engineering, M.Sc. Thesis, 2011, İstanbul (Danışman: Assoc. Prof. Dr. Y Mehmet)

Das, K, Nag C, Ghosh M. Familial, Environmental, and Occupational Risk Factors in Development of Amyotrophic Lateral Sclerosis. North American Journal of Medicine Sciences. 2012;8: 350-355.

Demirhan A, Güler İ. Bilişim ve Sağlık. Bilişim Teknolojileri Dergisi. 2011;3:13-20.

Diler S. Veri Madenciliği Süreçleri ve Karar Ağaçları Algoritmaları İle Bir Uygulama. Y.Y.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2016, Van (Danışman: Yrd. Doç. Dr. D Saygın)

Du Y, Wen Y, Guo X, Hao J, Wang W, He A, Fan Q, Li P, Li L, Liang X, Zhang F. A Genome-wide Expression Association Analysis Identifies Genes and Pathways Associated with Amyotrophic Lateral Sclerosis. Cellular and Molecular Neurobiology. 2017;3: 635-639

Durairaj M, Ranjani V. Data Mining Applications In Healthcare Sector: A Study. International Journal Of Scientific & Technology Research Volume. 2013;10: 29-35.

Edelstein H. Introduction to Data Mining and Knowledge Discovery. 3 rd ed. Two Crows Corporation; 1999, p:26-27

Eisen A, Krieger C. Amyotrophic lateral sclerosis : a synthesis of research and clinical practice. New York: Cambridge University Press. 2006;3: 296

Evangelista D, Colonna G, Miele M, Cutugno F, Castello G, Desantis S, Costantini S. CDMS (Clinical Data Mining Software): a cytokinome data mining system for a predictive medicine of chronic inflammatory diseases. *Protein Engineering, Design & Selection*. 2010;12: 899–902.

Gartner G. Data Mining. Gartner: <https://www.gartner.com/technology/home.jsp> adresinden alındı. 2007

Giuliano D, Quintieri M, Serra S, Pignolo L. Clinical signs and early prognosis in vegetative state: A decisional tree, data-mining study. *Brain Injury*. 2009;7-8: 617-623.

Güllüoğlu SS. Tıp ve Sağlık Hizmetlerinde Veri Madenciliği Çalışmaları: Kansere Teşhisine Yönelik Bir Ön Çalışma. *Online Academic Journal of Information Technology*. 2011;5: 1-7.

Gündoğdu ÖE. Veri madenciliğinde Genetik Algoritmalar. K.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2007, Kocaeli (Danışman: Yrd. Doç. Dr. D Nevcihan)

Hamidou B, Couratier P, Besancon C, Nicol M, Preux P, Marin B. Epidemiological evidence that physical activity is not a risk factor for ALS. *Eur J Epidemiol*. 2014;7: 459-475.

Hand DJ, Symth P. Mannila H. Principles of Data Mining. 1st ed. Cambridge: The MIT Press; 2001, p:17-19

Han J, Kamber M. Data Mining: Concepts and Techniques. 3rd ed. Waltham: Morgan Kaufman Publishers; 2006, p:1-10.

Hastings M, Goedert M. Circadian clocks and neurodegenerative diseases: time to aggregate?. *Current Opinion in Neurobiology*. 2013;5: 880-887.

Hatipoğlu B. Dershanede Eğitiminin Üniversiteye Yerleşmedeki Etkisinin Veri Madenciliği İle İrdelenmesi. İ.A.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2013, İstanbul (Danışman: Prof. Dr. A Zafer)

Haux R. Health information systems – past, present, future. *International Journal of Medical Informatics*. 20006;3-4: 268-281.

Jackson HA, Cashy J, Frieder O, Schaeffer A. Data Mining Derived Treatment Algorithms From the Electronic Medical Record Improve Theoretical Empirical Therapy for Outpatient Urinary Tract Infections. *The Journal of Urology*. 2011;6: 2257-2262.

Jiang YH, Bressler J, Beaudet A. Epigenetics and human disease. *Annual Review of Genomics and Human Genetics*. 2004;5: 479-510.

Kamel F, Umbach D, Munsat T, Shefner J, Hu H, Sandler D. Lead Exposure and Amyotrophic Lateral Sclerosis. *Epidemiology*. 2002;3: 311-319.

Koca TT. Amyotrofik Lateral Skleroz: Klinik Özellikler ve Güncel Tedavi Yaklaşımları . *Arşiv Kaynak Tarama Dergisi*. 2015;2: 182-194.

Koyuncugil AS, Özgülbaş N. Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri Dergisi*. 2009;2: 21-32.

Koyuncugil AS, Özgülbaş N. Donor Research and Matching System Based on Data Mining in Organ Transplantation. *J Med Syst*.2010;3: 251-259.

Köktürk F, Ankaralı H, Sümbüloğlu V. Veri Madenciliği Yöntemlerine Genel Bakış. *Türkiye Klinikleri*. 2009;1: 20-25

Krakauer M, PJ W, Hofman A. Genetic epidemiology of amyotrophic lateral sclerosis. *Clinical Genetics*. 2002;2: 83-101.

Kuşçu H. Nesneye Yönelik Programlama. Trakya Üniversitesi: [http://hilmi.trakya.edu.tr/ders\\_notlari/Java/JAVA\\_Kitabi.pdf](http://hilmi.trakya.edu.tr/ders_notlari/Java/JAVA_Kitabi.pdf) adresinden alındı. 2009

Lakshmanan, B, Srinivasan V, Ponnuraja C. Data Mining with Decision Tree to Evaluate the Pattern on Effectiveness of Treatment for Pulmonary Tuberculosis: A Clustering and Classification Techniques. *Scientific Research Journal*. 2015;6: 43-48.

Larose D. *Discovering Knowledge in Data: An Introduction to Data Mining*. 1st ed. New Jersey: A John Wiley & Sons, Inc; 2005, p:5-21

Leblond C, Kaneb H, Dion P, Rouleau G. Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Experimental Neurology*. 2014;262: 91-101.

Lorr M. *Cluster Analysis for Social Sciences*. 1st ed. San Francisco: Jossey-Bass; 1983, p:461-462

Ludolph A, Drory V, Hardiman O, Nakano I, Ravits J, Robberecht W, Shefner J. A revision of the El Escorial criteria - 2015. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. 2015;5-6: 291-292

Marangi G, Traynor B. Genetic causes of amyotrophic lateral sclerosis: New genetic analysis methodologies entailing new opportunities and challenges. *Brain Research*. 2015;1607: 75-93.

Mehta P, Kaye W, Raymond J, Ruoming W, Larson T, Punjani R, Heller D, Cohen J, Peters T, Muravov O, Horton, K. Prevalence of Amyotrophic Lateral Sclerosis — United States, 2014. *Morbidity and Mortality Weekly Report*. 2018;7: 216-218.

Mehta P, Kaye W, Raymond J, Ruoming W, Larson T, Punjani R, Heller D, Cohen J, Peters T, Muravov O, Horton, K. Prevalence of amyotrophic lateral sclerosis - United States, 2010-2011. *Morbidity and Mortality Weekly Report*. 2014;7: 1-14.

Milli Eğitim Bakanlığı. *Biyomedikal Cihaz Teknolojileri / Tıbbi Bilişim*. 1st ed. Ankara: Milli Eğitim Bakanlığı; 2011, p: 5-11

Nikam S. *A Comparative Study of Classification Techniques in Data Mining Algorithms*. *Oriental Journal Of Computer Science & Technology*. 2015;1: 13-19.

Nısbet R, Elder J, Miner G. *Handbook of Statistical Analysis and Data Mining Applications*. 1st ed. Canada: Academic Press Elseiver; 2009, p:126-127

Obenshain MK. Application of Data Mining Techniques to Healthcare Data. *Infection Control and Hospital Epidemiology*. 2004;8: 690-695.

Office of Communications and Public Liaison. Amyotrophic Lateral Sclerosis (ALS) Fact Sheet. National Institute of Neurological Disorders and Stroke: <https://www.ninds.nih.gov> adresinden alındı. 2013

Olson D, Delen D. *Advanced Data Mining Techniques*. 1st ed. Berlin: Springer-Verlag Berlin Heidelberg; 2008, p:12-13

Oskarsson B, Horton D, Mitsumoto H. Potential Environmental Factors in Amyotrophic Lateral Sclerosis. *Neurol Clin*. 2016;4: 1-13.

Pamphlett R, Jew SK. Heavy metals in locus ceruleus and motor neurons in motor neuron disease. *Acta Neuropathol Commun*.2013;81: 1-15.

Phan T, Chen J, Singhal S, Ma H, Clissold B, Ly J, Beare R. Exploratory Use of Decision Tree Analysis in Classification of Outcome in Hypoxic–Ischemic Brain Injury. *Frontiers in Neurology*.2018;9:126-131.

Philippe C, Helene B, William C. Genetics of amyotrophic lateral sclerosis. *Revue Neurologique*. 2017;3: 254-262.

Poyraz O. Tıp'ta Veri Madenciliği Uygulamaları: Meme Kanseri Veri Seti Analizi. T.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2012, Edirne (Danışman: Doç. Dr. U Erdem)

Pupillo E, Messina P, Logroscino G, Zoccolella S, Chio A, Calvo A, Corbo M, Lunetta C, Micheli A, Millul A, Vitelli E, Beghi, E. Trauma and amyotrophic lateral sclerosis: a case-control study from a population-based registry. *European Journal of Neurology*.2012;12: 1509-1517.

Renton A, Chio A, Traynor B. State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*. 2013;1: 17-23.

Senio Y, Hasegawa T, Koya T, Sakagami T, Mashima I, Shimizu N, Muramatsu Y, Muramatsu K, Suzuki E, Kikuchi T, Nigata Respiratory Disease Study Group. A Cluster

Analysis of Bronchial Asthma Patients with Depressive Symptoms. *Internal Medicine* 2018; 9073: 17-24

Shaw C, Al-Chalabi A, Leigh N. Progress in the Pathogenesis of Amyotrophic Lateral Sclerosis. *Current Neurology and Neuroscience Reports*. 2001;1: 69-76.

Silahtar G. Veri Madenciliği Kavram ve Algoritmaları. 1st ed. İstanbul: Papatya Yayınevi; 2013, p:39-42

Strickland D, Smith, S, Dolliff G. Amyotrophic Lateral Sclerosis and Occupational History A Pilot Case-Control Study. *Arch Neurol*; 1996;8: 730-733.

Sutedja N, Veldin, J, Fischer K, Kromhout H, Heederik D, Huisman M, Vokke JH, van den Berg, L. Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: A systematic review. *Amyotrophic Lateral Sclerosis*. 2009;5-6: 301-309.

Şehirlioğlu AK, Sarıbay E. Parametrik Olmayan Hipotez Testleri. Dokuz Eylül Üniversitesi: <http://kisi.deu.edu.tr/kemal.sehirli/poi2010.pdf> adresinden alındı. 2012

Şimşek Erdem N. Amyotrofik Lateral Skleroz Hastalarında Solunum ve Yutma Fonksiyon Bozukluklarının İlişkisi. A.Ü. Sağlık Bilimleri Enstitüsü, Uzmanlık Tezi, 2013, Antalya (Danışman: Prof. Dr. U Hilmi)

Tang, QS, Sun , WJ, Qu M, Guo DF. Analysis of syndrome discipline of generalized anxiety disorder using data mining techniques. *Journal of Chinese Integrative Medicine*. 2012;9: 975-982.

Telcioğlu B. Veri Madenciliğinde Genetik Programlama Temelli Yeni Bir Sınıflandırma Yaklaşımı Ve Uygulaması. E. Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2007, Kayseri (Danışman: Yrd. Doç. Dr. Ö Lale)

The ALS Association. About ALS. The ALS Association: <http://www.alsa.org/about-als/what-is-als.html?referrer=https://www.google.com.tr/> adresinden alındı. 2005

Ture M, Tokatli F, Kurt İ. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients . *Expert Systems with Applications*. 2009;2: 2017-2026.

Ümit, FS. Kümeleme Analizi: İstihdamın Sektörel Yapısı Açısından Avrupa Ülkelerinin Karşılaştırılması. *İstanbul Üniveristesi Sosyal Bilimler Dergisi*. 1997;2: 50-59.

Van Bommel JH. The Structure of Medical Informatics. *Med Inform*. 1998;3-4: 175-180.

Wang MD, Gomes J, Cashman N, Little J, Krewski D. A Meta-Analysis of Observational Studies of the Association Between Chronic Occupational Exposure to Lead and Amyotrophic Lateral Sclerosis. *Journal of Occupational and Environmental Medicine*. 2014;12: 1235-1242.

Wang MD, Little J, Gomes J, Cashman N, Krewski D. Identification of risk factors associated with onset and progression of amyotrophic lateral sclerosis using systematic review and meta-analysis. *NeuroToxicology*. 2017;61: 101-130.

Wang R, Strong D, Guarascio L. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*. 1996;4: 5-33.

Yaz HF. Çok Değişkenli İstatistiksel Tekniklerden Kümeleme Analizi; Spss İle Bir Uygulama. [www.academia.edu](http://www.academia.edu) adresinden alındı. 2014: 1-22

Yılmaz M. Süperoksit Dismutaz Genlerinin Amyotrofik Lateral Skleroz Hastalığının Oluşumundaki Rolünün Anlaşılması. K.Ü Sağlık Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2008, Kocaeli (Danışman: Prof. Dr. S Ali)

Yu B, Pamphlett R. Environmental insults: critical triggers for amyotrophic lateral sclerosis. *Translational Neurodegeneration*. 2017;6: 6-15.

Zhang Y, Guo SL, Han LN, Li TL. Application and Exploration of Big Data Mining in Clinical Medicine. *Chinese Medical Journal*. 2016;6: 731-738.



Zou ZY, Zhou ZR, Che CH, Liu CY, He RL, Huang HP. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Neurogenetics*. 2016;7: 540-549.

Zufiri'a M, Gil-Bea F, Ferná'ndez-Torro'n R, Poza JJ, Muñoz-Blanco JL, Rojas-Garci'a R, Riancho J, Munain AL. ALS: A bucket of genes, environment, metabolism and unknown. *Progress in Neurobiology*. 2016;142: 104-129.



## EKLER

### Ek 1: Terimler Sözlüğü

Terim No	Terim	Açıklama
1	age	Doğum Tarihi
2	sex	Cinsiyet
3	plbirth	Doğum Yerinin Büyüklüğü
4	plbirth2	Doğum Yeri Özelliği
5	sempyear	Semptom İlk Kaç Yıl Önce Görülmüş
6	deathmonths	İlk Semptomdan Kaç Ay Sonra Ölmüş
7	limbson	Ekstremitte Başlangıcı (Kol ve Bacak Başlangıç)
8	bulbaron	Bulbar Başlangıç
9	manifest	Başlangıç Belirtileri (UMN, LMN)
10	limb18	Uzuv
11	predom	Baskın Taraf
12	predomimpair	Baskın Bozukluk
13	fascicula	Başlangıçta Seyirme
14	emotion	Başlangıçtaki Duygusal Oynaklık
15	cognitive	Başlangıçtaki Bilişsel Belirtileri
16	handed	Baskın El
17	bulbarreg	Bulbar Bölge
18	tounge	Dil Spastisitesi
19	jawjerk	Çene Refleksi
20	toungefas	Dilde Fasikülasyon (LMN)
21	weakiris	Çene Refleksi (LMN)
22	facial	Yüz Kaslarında Fasikülasyon
23	upperlimb	Üst Ekstremitte
24	HyperUMN	Hiperrefleksi
25	finger	Parmak Fleksiyonu /Hoffman Belirtisi
26	spactic	Spastisite
27	atrophy	Atrofi ve Güçsüzlük
28	fascilonset	Başlangıçtaki Seyirme
29	Hyporef	Hiporefleksi

30	lowlimb	Alt Ekstremitte
31	lhrefleksumn	Hiporefleksi
32	fingerflex	Parmak Fleksiyonu/Babinski İşareti
33	spactic28	Spastisite
34	atrophy29	Atrofi Ve Güçsüzlük
35	fascicula29	Başlangıçtaki Seyirme
36	Hypo29	Hiporefleksi
37	axiel	Aksiyel Kaslar
38	neck	Boyun Zayıflığı
39	thorac	Torasik Kaslarda Seyirme
40	respirator	İstirahatte Solunum Yorgunluğu
41	ortopne	Ortopne
42	paradoxi	Paradoksik Solunum
43	weakcough	Zayıf Öksürük
44	elescorial	El Escorial Kriterlerine Göre Gruplandırma
45	emotionalable	Duygusal Kararsızlık
46	restinglower	Üst Ekstremitte İstirahat Tremoru
47	restingupper	Alt Ekstremitte İstirahat Tremoru
48	bradyup	Hareket Yavaşlığı -Upper
49	bradylow	Hareket Yavaşlığı -Lower
50	postural	Postürel Değişkenlik
51	gait	Geniş Tabanlı Yürüyüş
52	first	1. Bölge
53	second	2. Bölge
54	third	3. Bölge
55	fourth	4. Bölge
56	fifth	5. Bölge
57	sixth	6. Bölge
58	seventh	7. Bölge
59	eighth	8. Bölge
60	ninth	9. Bölge
61	totalregion	Toplam Yayılan Bölge Sayısı
62	region1	Bölge 1 - Özellik

<b>63</b>	region2	Bölge 2 - Özellik
<b>64</b>	region3	Bölge 3 - Özellik
<b>65</b>	alsfrs	Toplam ALSFRS-R
<b>66</b>	alsfrsrate	ALSFRS-R Düşme Oranı
<b>67</b>	hypertension	Hipertansiyon
<b>68</b>	diabet	Diyabet Tip I/II
<b>69</b>	colesterol	Hiperkolesterolemi
<b>70</b>	gliseri	Trigliserit Yüksekliği
<b>71</b>	sigara	Sigara Kullanımı
<b>72</b>	mutdur	Mutasyon Durumu
<b>73</b>	nsibling	Ana Baba Akrabalıđı
<b>74</b>	nthsibling	Hasta Kaç Kardeř
<b>75</b>	nchild	Hasta Kaçınıcı Çocuk
<b>76</b>	csayisi	Çocuk Sayısı
<b>77</b>	mainoccup	Düzenli Fiziksel Egzersiz
<b>78</b>	work	Meslek
<b>79</b>	placelive	Yaşadıđı Yer
<b>80</b>	yasyer2	Yaşadıđı Alan (Kırsal/Kentsel)

## ÖZGEÇMİŞ

### Kişisel Bilgiler

<b>Adı</b>	Nesrin	<b>Uyruğu</b>	TC
<b>Soyadı</b>	ÇELİK	<b>Tel no</b>	507-0577609
<b>Doğum tarihi</b>	10.01.1992	<b>e-posta</b>	nesrincelik@akdeniz.edu.tr

### Eğitim Bilgileri

<b>Mezun olduğu kurum</b>	<b>Mezuniyet yılı</b>
<b>Yüksek Lisans</b> Akdeniz Üniversitesi	Devam ediyor
<b>Lisans</b> Mehmet Akif Ersoy Üniversitesi	2015
<b>Lisans</b> Anadolu Üniversitesi	2015
<b>Lise</b> Nahit Mentеше Anadolu Meslek Lisesi	2009
<b>İlk Öğretim</b> İhsan İzgi İlk ve Orta Öğretim Okulu	2005

### İş Deneyimi

<b>Görevi</b>	<b>Kurum</b>	<b>Süre (yıl-yıl)</b>
<b>Proje Geliştirme ve Üniversite Sanayi İşbirliği Uzmanı</b>	Akdeniz Üniversitesi AKİŞMER TTM	12.2015 - Halen
<b>Stajyer</b>	Akdeniz Üniversitesi AKİŞMER TTM	02.2015- 06.2015
<b>Stajyer</b>	Efor Bilgi İşlem	2008-2009

<b>Yabancı Dilleri</b>	<b>Sınav türü</b>	<b>Puanı</b>
<b>İngilizce</b>	YÖKDİL	58,7

## Proje Deneyimi

Proje Adı	Destekleyen kurum	Süre (Yıl-Yıl)
Küçük Ölçekli Altyapı Mali Destek Programı Proje Yazımı	Batı Akdeniz Kalkınma Ajansı (BAKA)	2018
Akdeniz Üniversitesi Araştırma Altyapısı Makine - Teçhizat ve Hizmet Envanteri Altyapısının Hazırlanması Proje Yazımı ve Yürütülmesi	Akdeniz Üniversitesi BAP	2018-2019
1601 Teknoloji Transfer Ofisleri Kapasite Geliştirme Programı Proje Yazımı ve Yürütülmesi	TÜBİTAK	2017-2019
1601 Teknoloji Transfer Ofisleri Kapasite Geliştirme Programı Yürütülmesi	TÜBİTAK	2015-2016
Antalya OSB Firmaları Ar-Ge ve İnovasyon Kapasite Ölçümü Fizibilite Çalışması BAKA Doğrudan Faaliyet Projesi Yürütülmesi	Batı Akdeniz Kalkınma Ajansı	2015-2015
2237 Proje Yazım Atölyesi	TÜBİTAK	2016-2017
1503 Proje Pazarları Destekleme Programları- Bitki Islahı Proje Pazarı	TÜBİTAK	2016
IVT Elektrik Hattı Üzerinden İnternet Bağlantısı	Ulusal Ajans	2009-2009

## Burslar-Ödüller:

## Yayınlar ve Bildiriler:

Kitapci O., Sakarya M.C., Dörtyol İ.T., Sezgin İ.V., Çelik N. , Koçyiğit M., et al., "ANTALYA FİRMALARINA YÖNELİK ENDÜSTRİ 4.0 DURUM TESPİTİ", 2018

Çelik N. Bilge U.,Uysal H.," Amiyotrofik Lateral Skleroz (ALS) Hastalarının Klinik Verilerinin Veri Madenciliği Yöntemleri ile İncelenmesi", X. ULUSLARARASI TIP BİLİŞİMİ KONGRESİ, Antalya, TÜRKİYE, 12 Ekim 2017- 15 Ekim 2017, Poster Sunumu

Bilge U., Çelik N., "An Analysis of Gene Variations in Craniosynostosis", ICCEE 2016, Barselona, İSPANYA, 7 Aralık 2016 - 11 Şubat 2017, pp.37-37

