



T.C
BİRÜNİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI
BİYOİSTATİSTİK TEZLİ YÜKSEK LİSANS PROGRAMI

YAPAY ZEKADA VERİ MADENCİLİĞİ YÖNTEMİ
KULLANARAK BİYOİNFORMATİK YAZILIM
GELİŞTİRİLMESİ: SİTRÜLİN İLE BİR UYGULAMA

Ayşe AKGÜL IŞIK

DANIŞMAN
Prof. Dr. M. Yusuf ÇELİK

Aralık, 2022



**T.C.
BİRÜNİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI
BİYOİSTATİSTİK TEZLİ YÜKSEK LİSANS PROGRAMI**

**YAPAY ZEKADA VERİ MADENCİLİĞİ YÖNTEMİ
KULLANARAK BİYOİNFORMATİK YAZILIM
GELİŞTİRİLMESİ: SİTRÜLİN İLE BİR UYGULAMA**

AYŞE AKGÜL IŞIK

**DANIŞMAN
Prof. Dr. YUSUF ÇELİK**

Aralık, 2022

T.C.
BİRÜNİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
YÜKSEKLİSANS TEZ ONAY SAYFASI



BEYAN

Bu tezin bana ait olduğunu, tüm aşamalarında etik dışı davranışımın olmadığını, içinde yer alan bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, kullanmış olduğum bütün bilgilere kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin yürütülmesi ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını beyan ederim.

Öğr. Gör. Ayşe AKGÜL IŞIK

İMZA

TEŞEKKÜR

Öncelikle sayın hocam Prof. Dr. M. Yusuf Çelik'e, tezimin süreci boyunca verdiği mentorluk, destek, düşünceli yorum ve tavsiyeleri için sonsuz teşekkürlerimi sunarım. Ayrıca değerli fikir ve deneyimleri ile manevi desteğini her zaman en üst düzeyde tutan, yol gösteren Prof. Dr. Fatma Çelik'e,

Tanıştığımız ilk andan itibaren her türlü iyi, kötü ve zor anımda yanımda olan, en büyük destekçilerimden Dr. Öğr. Üyesi Tuğba Elgün, Tuba Saraç ve Berivan Şahan'a

Çalışma hayatını severek paylaştığım, desteklerini her zaman hissettiğim çok sevgili arkadaşlarım Elif, Zehra ve Mert'e,

Hayatımın her alanında yanımda olan, karşılaştığım her zorlukta her zaman arkamda olan, bana benden daha çok güvendiği ve beni her zaman desteklediği için kendisine özellikle minnettar olduğum hayat arkadaşım Furkan Işık'a,

Beni bu günlere getiren, seçtiğim meslekten yaptığım işe kadar tüm seçimlerimde yanımda olan sevgili aileme,

en içten teşekkürlerimi sunuyorum...

İÇİNDEKİLER

İÇ KAPAK	-
ONAY SAYFASI	-
BEYAN.....	iii
TEŞEKKÜR	iv
İÇİNDEKİLER	v
SİMGE/SEMBOL VE KISALTMALAR LİSTESİ.....	vii
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ.....	ix
TÜRKÇE ÖZET VE ANAHTAR KELİMELER.....	x
İNGİLİZCE ÖZET VE ANAHTAR KELİMELER.....	xi
1. GİRİŞ VE AMAÇ	1
2. GENEL BİLGİLER	3
2.1. Yapay Zeka	3
2.2. Veri Madenciliği	4
2.2.1. Veri madenciliğinin tanımı	6
2.2.2. Veri madenciliği süreci ve aşamaları	6
2.3. Sağlıkta Veri Madenciliği	10
2.3.1. Veri madenciliği görevleri ve sağlık hizmetleri alanında kullanımı .	11
2.3.2. Sağlıkta bilgi yönetimi ve veri madenciliği	12
2.3.3. Veri madenciliğinin genetik disiplindeki önemi	13
2.4. Veri Madenciliği Algoritmaları	15
2.4.1. Yapay sinir ağları	16
2.4.2. Karar ağacı	18
2.4.3. Bulanık kümeler	19
2.4.4. Destek vektör makinesi.....	20
2.4.5. Bayes sınıflandırma algoritması	21
2.4.6. Genetik algoritma	21
2.4.7. Naive bayes algoritması	22
2.4.8. Random forest algoritması.....	23
2.5. Sitrülin Nedir?	23
2.5.1. Sitrülinin tanımı	23
2.5.2. Sitrülin metabolizması ve muhtemel metabolik yolları.....	25

2.5.3. Sitrülinin genler ile ilişkisi ve veri madenciliğindeki yeri	29
3. GEREÇ VE YÖNTEM	30
3.1. Araştırmanın Yapıldığı Yer ve Zaman, Türü	30
3.2. Veri Toplama Araçları	30
3.3. Veri Toplama Yöntemi	31
3.3.1. <i>STRING Gene ilişkisi</i>	31
3.3.2. <i>David Gen Pathway</i>	31
3.3.3. <i>Reactome</i>	31
3.3.4. <i>KEGG Pathway</i>	32
3.3.5. <i>Wiki Pathway</i>	33
3.4. Biyoinformatik Analiz	33
3.5. Veri Seti.....	35
3.5.1. <i>GSE1563</i>	35
3.5.2. <i>Kullanılan veri seti ve veri setinin düzenlenmesi</i>	35
3.6. Verilerin İstatistiksel Analizi	36
3.6.1. <i>Kümeleme analizi (Cluster analysis)</i>	37
3.6.2. <i>Volcano Plot</i>	37
3.6.3. <i>PlotMD</i>	37
3.6.4. <i>Quantil-Quantil Plot</i>	38
3.6.5. <i>BoxPlot</i>	38
4. BULGULAR	39
4.1. Analiz Sonuçları	39
5. TARTIŞMA	57
5.1. Tartışma	57
5.2. Sonuç ve Öneriler	61
6. KAYNAKLAR	63
7. EKLER	72
8. ÖZGEÇMİŞ	73
9. İNTİHAL RAPORU	74

SİMGE/SEMBOL VE KISALTMALAR LİSTESİ

ARG	Arginin
ARG1	Arginaz1
ASL	Argininosüksinat Liyaz
ASS1	Argininosüksinat Sentetaz
CPS1	Karbamoil Fosfat Sentetaz 1
DDAH2	Dimetil Arginin Dimetilaminohidrodaz 2
DVM	Destek Vektör Makineleri
FC	Fold Change
GA	Genetik Algoritmalar
GenBank Overview NCBI- GEO	National Center for Biotechnology Information
Gene Expression Omnibus	
GLN	Glutamin
k-NN	En Yakın Komşu Algoritması
KEGG	Kyoto Genler ve Genomlar Ansiklopedisi
NB	Naive Bayes
NO	Nitrik Oksit
NOS	Nitrik Oksit Sentaz
OTC	Ornitin Transkarbamilaz
OTCD	Ornitin Transkarbamilaz Eksikliği
P5CS	Δ 1-pirolin-5-karboksilat sentaz
PBL	Periferik Kan Lenfositleri
PCA	Temel Bileşen Analizi
TCA	Sitrik Asit

TABLOLAR LİSTESİ

Tablo No	Tablo Adı	Sayfa No
Tablo 2.4.1.	Veri Madenciliği Görevleri ve Veri Teknikleri Arasındaki İlişki	16
Tablo 2.4.2.1.	Karar Ağacının Özellikleri	19
Tablo 4.1.1.	Sitrülin İlişkili Genlerin (Homo Sapiens) Tanımları ve Görsel İfadeleri	41
Tablo 4.1.2.	Sitrülin İlişkili Genlerin Görev Aldıkları Yolaklar	43
Tablo 4.1.3.	Üre Döngüsü Bozukluğunda Sitrülin İlişkili Genlerin Z Skorları	50
Tablo 4.1.4.	Sitrülin İlişkili Genlerin Dahil Olduğu Biyosentez ve Metabolik Yolların İstatistiksel Analizi (Reactome Software)	51
Tablo 4.1.5.	Sitrülin İlişkili Genlerin Log Fc Değerleri	52

ŞEKİLLER LİSTESİ

Şekil No	Şeklin İsmi	Sayfa No
Şekil 2.2.1	Veri Madenciliği ve İş Zekası	6
Şekil 2.2.2.1	Veri Madenciliği Süreci	7
Şekil 2.3.1.1	Tıbbi Veri Madenciliği Disiplinleri	12
Şekil 2.5.1.1	L-Sitrülin Desteğinin Etkileri	24
Şekil 3.4.1	Sitrülin İlişkili Gen Veri Analizi	34
Şekil 3.5.2.1	R Software Programı Ara Yüzü	36
Şekil 4.1.1	Sitrülin İlişkili Genlerin Tespit Edilmesi İçin Overview-NCBI -NIH Kütüphanesinin Ekran Görüntüsü	39
Şekil 4.1.2	Sitrülin İlişkili Genlerin STRING GENE Analizi ve Bağlantıları	40
Şekil 4.1.3	Sitrülin <i>Biyosentez</i> Sürecinde Etkili Olan Genler Arasındaki İlişkinin Gösterilmesi	42
Şekil 4.1.4	Sitrülin <i>Metabolik</i> Sürecinde Etkili Olan Genler Arasındaki İlişkinin Gösterilmesi	42
Şekil 4.1.5	Sitrülin <i>Biyosentez</i> Sürecinde Etkili Olan Genlerin Analizi	43
Şekil 4.1.6	KEGG Pathway Analizi ile Arginin Biyosentezi	45
Şekil 4.1.7	KEGG Pathway Analizi ile Alanin, Aspartat ve Glutamat Metabolizması	46
Şekil 4.1.8	KEGG Pathway Analizi ile Amino Asitlerin Biyosentezi	47
Şekil 4.1.9.	Üre Döngüsü ve İlgili Hastalıklar	48
Şekil 4.1.10	Üre Döngüsü ve İlişkili Yollar (Wiki Pathway)	49
Şekil 4.1.11	Böbrek Transplantasyon ve Doku Hasarı Bulunan Klinik Hasta Datası Üre Döngüsü İlişkili 100 Genin Küme (Cluster) Analizi (Heat-Map Analizi)	53
Şekil 4.1.12	BoxPlot Analizi	54
Şekil 4.1.13	VolcanoPlot Analizi	55
Şekil 4.1.14	PlotMD Analizi	56
Şekil 4.1.15	QQ-Plot Analizi	56

TÜRKÇE ÖZET VE ANAHTAR KELİMELER

Işık, A.A. (2022). Yapay Zekada Veri Madenciliği Yöntemi Kullanarak Biyoinformatik Yazılım Geliştirilmesi: Sitrülin İle Bir Uygulama. Yüksek Lisans Tezi, Biruni Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.

Günümüzde sağlık alanında yapay zeka giderek büyük önem kazanmaya başlamıştır. Başta yapay zeka olmak üzere bilgisayar bilimi, makine öğrenmesi, veri görselleştirme, veri tabanı yönetimi ile algoritmalarını, istatistiği birleştiren, veri madenciliği disiplinler arası bir alanı oluşturmaktadır. Veriler arasından ihtiyaç duyulanan seçilmesi ve bunların işlenerek anlamlı bağıntıların araştırılması veri madenciliğinin asıl konusunu oluşturmaktadır. Bu çalışma ile istatistik temelli bir yaklaşımla, biyosentez ve metabolik yolu henüz tam olarak aydınlatılmamış ve insan metabolizmasında da önemli bir aminoasit/protein olan sitrülinin veri madenciliği ile R Software kullanılarak bir arayüz oluşturulması amaçlanmaktadır. Bu çalışmanın sitrülin ile ilişkili gen verileri GenBank Overview-NCBI-GEO veri tabanından elde edilmiştir. Veriler üzerinde Bayes Sınıflandırma Algoritması kullanılmış ve biyoinformatik analiz yapılmıştır. Yapılan analizler ile sitrülin ilişkili genlerin protein hedefine ne düzeyde etki ettiği sayısal olarak ilişkilendirilmiştir. İlişkilendirme sonrasında ise olası metabolik yollar belirlenmesinde String, Reactome Pathway Database, Wiki Pathway, KEGG ve David Gen Pathway veri tabanları kullanılmıştır. Sitrülin ilişkili genlerin görev aldığı yollar değerlendirildiğinde; biyosentez yolağında en fazla görev alan gen CPS1 iken, metabolik yolda en fazla görev alan genin ASS1 olduğu belirlenmiştir. Metabolik yolağında sitrülinin en fazla görev aldığı ve etkili olduğu yolağın ise %30 'luk oran ile üre döngüsü olduğu belirlenmiştir. Bu kapsamda klinik tablo örneği olarak belirlenen böbrek transplantasyon ve doku hasarı bulunan klinik hasta verileri üzerinden üre döngüsünde sitrülin ilişkili 100 genin cluster analizi (heat-map analizi) yapılmıştır. Sonuçlar PlotMD, VolcanoPlot, BoxPlot, QQ-Plot analizi ile yorumlanmıştır.

Anahtar Kelimeler: Yapay zeka, veri madenciliği, sitrülin, üre döngüsü, cluster analizi

İNGİLİZCE ÖZET VE ANAHTAR KELİMELER

Isik, A.A. (2022). Development of Bioinformatics Software Using Data Mining Method in Artificial Intelligence: An Application with Citrulline. Master Thesis, Biruni University Graduate Education Institute, Istanbul.

Artificial intelligence has started to gain importance especially in the field of health today. Data mining is an interdisciplinary field that combines artificial intelligence, computer science, machine learning, data visualization, database management, algorithms and statistics. The main subject of data mining is to select the needed data among the all data, and then to process them and create meaningful connections. In this study, it is aimed to create an interface using R Software with data mining of citrulline, which is an important protein/aminoacid in human metabolism, whose biosynthesis and metabolic pathway has not been fully elucidated yet, with a statistics-based approach. Citrulline-related gene data for this study were obtained from the GenBank Overview-NCBI-GEO database. Bayesian Classification Algorithm was used on the data and bioinformatic analysis was performed. The effects of citrulline-related genes on the protein were numerically shown and correlated with the analyzes. After the association, String, Reactome Pathway Database, Wiki Pathway, KEGG and David Gen Pathway databases were used to determine possible metabolic pathways. Considering the pathways in which citrulline-related genes are involved; While CPS1 was the most involved gene in the biosynthesis pathway, ASS1 was the gene most involved in the metabolic pathway. In the metabolic pathway analysis, it has been determined that the pathway in which citrulline is most involved and effective is the urea cycle with a rate of 30%. In this context, determined as an example of clinical picture, Cluster analysis (heat-map analysis) of 100 citrulline-related genes in the urea cycle was performed through data from clinical patients with kidney transplantation and tissue damage. Results were interpreted by PlotMD, VolcanoPlot, BoxPlot, QQ-Plot analysis.

Keywords: Artificial intelligence, data mining, citrulline, urea cycle, cluster analysis

1. GİRİŞ VE AMAÇ

Günümüzde yapay zeka uygulamaları özellikle tıpta; sanal ve fiziksel olmak üzere iki kategoride ele alınmaktadır. Fiziksel olarak kullanımı ameliyatların yapılmasına yardımcı olan robotlar, engelliler için akıllı protezler ve yaşlı bakımı ile ilgili alanları incelenmesi gerçekleştirilmektedir. Sanal kullanımı, elektronik sağlık kayıt sistemleri gibi uygulamalardan, tedavi kararlarında sinir ağı tabanlı rehberliğe kadar uzanmaktadır (Liao, 2003).

Bu kapsamda veri, dijital bir ortamda olan ancak işlenmemiş ham bilgi olarak adlandırılmaktadır. Ayrıca dijital programlar aracılığı ile işlenen bilgilerin formülize edilmiş hali de veridir. Veriye ulaşım ölçüm veya gözlem sonucunda bilgilerin kayıt altında tutulması ile gerçekleştirilebilmektedir. Veri madenciliği yönteminin uygulanabilmesi için elde edilen büyük ölçekli verilerin içerdiği anlamlı bağıntıların analiz edilmesi gerekmektedir (Erdem, 2008). Veri madenciliği desteği ile, farklı verilerin aralarındaki bağlantı ve/veya etkileşimler açıklanmaktadır. Dolayısıyla veri kullanan kişilere bilgi vermek amacıyla veriler arasında ki doğru ilişkinin kurulması son derece önemlidir (Amin et al., 2019).

Başta yapay zeka olmak üzere bilgisayar bilimi, makine öğrenmesi, veri görselleştirme ve veri tabanı yönetimi ile matematik algoritmalarını, istatistiği birleştiren veri madenciliği disiplinler arası bir alanı oluşturmaktadır (Liao, 2003). Veriler arasından ihtiyaç duyulunun seçilmesi ve bunların işlenerek anlamlılıkların keşfedilmesi veri madenciliğinin asıl konusunu oluşturmaktadır. Günümüzde sağlık alanında veri madenciliği giderek büyük önem kazanmaya başlamıştır (Firouzi et al., 2018).

Sağlık hizmetlerinin en hızlı, en doğru, en yüksek kalitede ve ihtiyaca cevap verecek şekilde sunulabilmesi için sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşması ve bu bilgiyi karar destek sistemlerinden faydalanarak kullanması gerekmektedir. Karar analizin de ve bir olayın (örn., hastalık) olma olasılığının hesaplanmasında istatistik yöntemlerinden yararlanılan veri madenciliği algoritmaları arasında yer alan Bayes Sınıflandırma Algoritması bu alanında sıklıkla kullanılmaktadır (Yüksel ve ark, 2018). Sağlık alanında kullanılan bir diğer algoritma ise genetik algoritmadır. Süreç modellemesi yapan genetik algoritmalar (GA)

problemin çözümlenebilmesi için çok yönlü bir arama ile en başarılı olanı tespit etme prensibine dayanmaktadır (Tuğ, 2005).

Çok geniş yelpazesi olan sağlık alanındaki çalışmalardan olan genetik temelli (gen havuzu) araştırmalarda veri madenciliğinin yeri çok büyük önem teşkil etmektedir. Özellikle algoritma yöntemleri metabolik yolak analizlerinin temel taşıını oluşturmaktadır.

Proteine giden metabolik yolda genin ve ifadesinin iyi tanınması, bu çalışmalar için olmazsa olmaz niteliktedir. Bilindiği üzere proteinler aminoasitlerden meydana gelir. Aminoasitlerdeki herhangi bir eksiklik veya kusur protein düzeyinde hasara dolaylı olarak da sağlığınıza ciddi olumsuzlukları (hastalık, sendrom) hatta geri dönüşü olmayan durumlarda ölümlere neden olabilmektedir.

Arginin (ARG) metaboliti olan sitrülün, glutamin (GLN) metabolizmasının son ürünü olan bir amino asittir (Papadia et al., 2018). GLN ise, bağırsak tarafından çevrilebilen bir ornitin öncüsüdür. Glutamat, GLN ve prolin gibi diyet ile alınan aminoasitlerin metabolize olması sonucu ince barsaklarda üretilen sitrülün, ARG sentezinde birincil yapı taşı olarak görev almaktadır. Sitrülünün ARG'e metabolize edilmesi enterositlerin içerisinde gerçekleşmektedir. Enterositler, arginino-süksinat sentaz enzimi bulundurmadığından, sitrülün portal vene serbest halde bırakılmaktadır. Sitrülünün yaklaşık olarak %80'i böbreklerde ARG'e dönüştürülmektedir (Kaorea et al., 2012). Dolayısıyla metabolizmanın normal fonksiyonunda az miktarda alınan sitrülün karaciğer tarafından sistemik dolaşıma girer ve böylece böbreklerde ARG'e dönüştürülebilir. Sitrülünün ARG'e göre depo oranı daha fazladır bu nedenle biyoyararlanımı yüksektir (Curis and Nicolis, 2005).

Sitrülünün olası metabolik yolaklarının aydınlatılmasında, literatürde benzeri bulunmamasından dolayı bu çalışmayı özgün değer açısından ileriye taşımaktadır. Biyoinformatik analiz ile sitrülün ilişkili genlerin ekspresyon verileri dikkate alınarak protein-protein düzeyinde inceleme yapılarak olası yolakların belirlenmesi hedeflenmiştir. Yapılan çalışma ile bu verilerden yararlanarak sitrülün ilişkili metabolik yolakların aydınlatılması için istatistik temeline dayanan veri madenciliğinden yararlanarak R programı ile bir arayüz oluşturması amaçlanmıştır.

2. GENEL BİLGİLER

2.1. Yapay Zeka

Modern bilgisayarların ve yapay zekanın kurucularından biri olan Alan Turing (1950), “Turing testi” ile bir bilgisayarın akıllı davranışının, bilişle ilgili görevlerde insan düzeyinde performans elde etme yeteneği olduğu gerçeğini ortaya koymuştur (Batal, 2016). Yapay zekaya olan ilgi ise 1980’li ve 1990’lı yıllarda artış göstermiştir. Özellikle sağlık hizmetlerinde farklı klinik ortamlarda yazılım literatüründeki; bulanık uzman sistemler, Bayes ağları, yapay sinir ağları ve hibrit akıllı sistemler gibi yapay akıllı teknikler kullanılmaktadır. Yapay zeka araştırmalarına 2016 yılında yapılan en büyük yatırımın diğer sektörlerle kıyasla sağlık uygulamaları alanında olduğu saptanmıştır (Prim, 2016).

Yapay zekanın tarihçesi dönemlere ayrılarak değerlendirildiğinde genel görüş dört farklı dönemde incelenmektedir. Bu dönemler sınıflandırıldığında:

Tarih Öncesi Dönem: Binlerce yıl önce Yunan mitolojisinde Rüzgar Tanrısı zannedilen Daedalus’un "yapay-insan" denemesi olduğu bilinmektedir.

Karanlık Dönem (1965-1970): Bilgisayar uzmanları, düşünen bir mekanizma geliştirmeye çalışmış ve sadece verileri yükleyerek akıllı bilgisayarlar yapmayı hedeflemişlerdir. Bu dönemde çok az gelişmeler elde edilmiş ve kısa süreli bir bekleme dönemine girilmiştir.

Rönesans Dönemi (1970-1975): Bilim insanları, hastalıkların teşhisi gibi çok önemli sistemler geliştirmişlerdir. Rönesans dönemi, sonraki dönemlere ışık tutacak ve yapay zeka çalışmalarına ivme kazandıracak çalışmaların yapıldığı dönemdir.

Ortaklık Dönemi (1975-1980): Yapay zeka ile ilgilenen bilim insanları, dil ve psikoloji gibi diğer bilim dallarından da yararlanmaya başlanıldığı dönemdir.

Girişimcilik Dönemi (1980-?): Yapay zeka, geliştirildikleri laboratuvarların dışına çıkarılarak, gündelik hayatın içerisinde kullanılmaya ve geliştirilmeye başlanmıştır. Bu dönem, günümüzde halen devam etmektedir (Karaduman, 2019).

Yapay zeka veya bilimsel adıyla mantıksal analiz robotunun bir tanımında bilgisayarın veya bilgisayar kontrolündeki robotun insana özgü niteliklerden olan akıl yürütme, anlam çıkartma ve geçmiş deneyimlerden öğrenme gibi yüksek zeka gerektiren faaliyetleri, insan zekasının çalışma modelini örnek alarak gerçekleştirmesi

olarak ifade edilebilir. Birçok bilim insanı yapay zeka ile ilgili çok çeşitli tanımlarda bulunmuştur. Tanımlardan bazıları şu şekildedir (Kapucu, 2014);

Genesereth ve Nilsson'a göre yapay zeka; "Akıllı davranış üzerine bir çalışmadır. Ana hedefi, doğadaki varlıkların akıllı davranışlarını yapay olarak üretmeyi amaçlayan bir kuram oluşturmaktır." şeklinde ifade edilirken, Tesler'e göre bu tanım; "Şu ana kadar yapılmayanlardır." Axe göre yapay zeka; "Akıllı programları hedefleyen bir bilimdir." Slage'ye göre ise; "İnsanların yaptıklarını bilgisayarlara yaptırabilme çalışmasıdır." İfadelerini kullandıkları belirtilmiştir.

İnsan aklına benzetilmiş yapay beyin ve yapay bilinç sistemi, insan akli kadar esnek ve duyguları olan bir karar alma mekanizması, uzman sistemler kadar yetkin bir bilgi birikimi ile yapay zeka, gelecekte insan zekasına bir alternatif oluşturabileceği düşünülmektedir (Batal, 2016).

2.2. Veri Madenciliği

Sağlık sistemlerinin dijitalleşmesi ile klinikte büyük veri seline yol açtığı ve tıpta veri biliminin hızla büyümesine neden olduğu saptanmıştır (NelsonSanchez-Pinto et al., 2018).

Veriye ulaşım, ölçüm veya gözlem sonucunda bilgilerin kayıt altında tutulması ile gerçekleştirilmektedir. Veri madenciliğinin uygulanabilmesi için elde edilen büyük ölçekli verilerin işlenmesi, düzenlenmesi ve analiz edilmesi gerekmektedir (Erdem, 2008). Veri madenciliği kullanılarak farklı veriler arasında bağlantı ve etkileşimler keşfedilmektedir. Dolayısıyla veri kullanan kişilere bilgi vermek amacıyla veriler arasındaki doğru ilişkilerin kurulması sağlanmaktadır (Amin et al., 2019). Şekil 2.2.1'de veri madenciliği basamaklarını gösteren piramit verilmiştir.

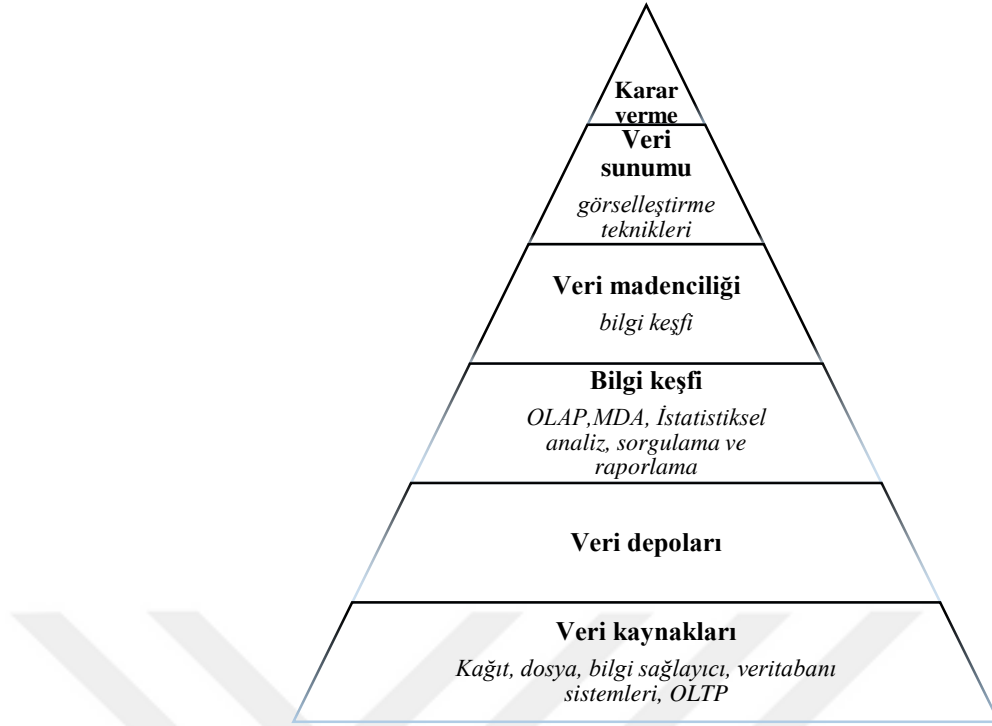
Büyük veri yığınlarının bulunduğu dijital ortamlardaki veriler birbirleri ile ilişkilendirilmektedir. Bu verilerin gelecek ile ilgili tahminde bulunması ve buna bağlı olarak gerekli algoritmalara başvurusu adımı veri madenciliği yöntemi adı verilir. Mevcut verilerin daha öncesinde keşfedilmemiş bilgilerini, belirli istatistikî yöntemler kullanarak ortaya çıkartılmasını veri madenciliği sağlamaktadır. Veri madenciliği, veriler arasındaki örüntüleri farklı algoritmaları kullanarak tespit etmeye çalışır (Yalçın, 2019). Günlük operasyonel ve stratejik karar vermeyi desteklemek için bilgi sistemleri; veri toplama, işleme ve dağıtma amaçların da kullanılmaktadır (Kayaalp ve Başarslan, 2018).

Büyük ham sağlık verileri koleksiyonunu, veri madenciliği yöntemi ile bilinçli karar ve tahmin yapılabilecek sürece yardımcı olabilecek bilgilere dönüştürmektedir. Tahmin modellerinin performansını iyileştirebilecek olan önemli özelliklerin doğru kombinasyonunu seçmek için kullanılmaktadır (Shouman et al., 2013).

Veri madenciliğinin tanımı ham verilerden anlamlı bilgilere geçiş sürecidir. Ham veriler arasından gizli olan bilgileri çıkarmak için veriler çeşitli veri madenciliği yöntemleri uygulanarak işlenmektedir. Veri madenciliğinin bir sonraki aşamasında işlenmiş olan veriler kullanılabilir hale gelmektedir. Veri madenciliği ve veri işlemede kullanılmak üzere birçok açık kaynak ve ticari uygulama bulunmaktadır (Michael and Gordon, 2000).

Veri madenciliğinin çeşitli tanımları literatürde yer almaktadır. Büyük miktardaki veriler içerisinde “değerli” olan bilgiyi elde etme süreci olarak adlandırılabilir (Yalçın, 2013). Benzer bir ifade ile veri tabanlarında bilgi keşfi, daha önce hiç keşfedilmemiş, potansiyel olarak yararlı ve anlaşılır olan bilgilerin büyük miktarlardaki verilerden çıkarılması sürecine veri madenciliği yardımcı olmaktadır. Veri madenciliği olarak değerlendirilen, Backend veritabanı yönetim sistemleri, istatistikler, yapay zeka, makine öğrenmesi, paralel ve dağıtık süreçler gibi veri analiz teknikleri bulunmaktadır (Han et al., 2022).

Gelecekle ilgili tahminler yapmamızı sağlayan bilgisayar programlarını kullanan veri madenciliği yöntemi; büyük miktardaki verilerden ilişki ve kuralların aranmasını sağlamaktadır. Geleceğin, en azından yakın geleceğin geçmişten çok farklı olmayacağı düşünülürse, geçmişte çizilen kurallar gelecekte de geçerli olacak ve gelecek için doğru tahminler yapmamızı sağlayacaktır. Veri tabanlarında bilgi keşfinin ilk adımları 1990'lı yıllarda atılmış ve veri madenciliği yeni teknolojiler ile birlikte yaygın olarak kullanılan yeni bir standart haline getirilmiştir (Alpaydın, 2000). Bilginin keşfi için belirli bir süreç gerekmektedir, süreci tamamlamak için ham veriyi kullanılabilir hale getirmek ve işlemek gerekmektedir (Kayaalp ve Başarslan, 2018).



Şekil 2.2.1: Veri Madenciliği ve İş Zekası (Cifci ve Hussain, 2018).

2.2.1. Veri madenciliğinin tanımı

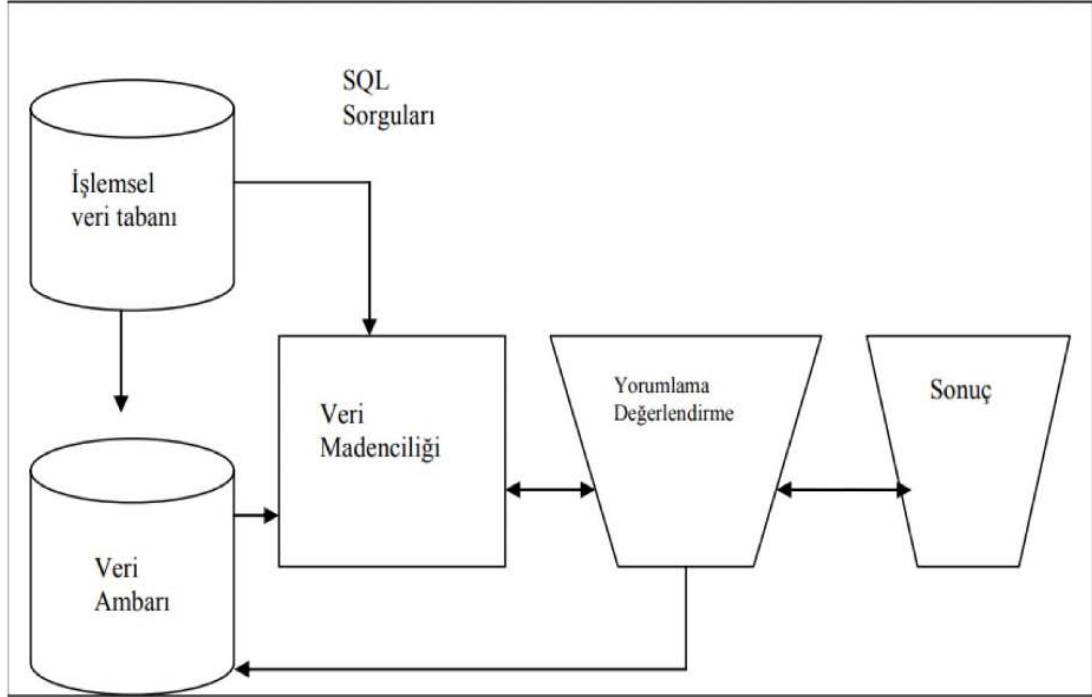
Veri tabanlarında Öz Bilgi Keşfi (Knowledge Discovery in Databases) adı altında, sürekli ve yeni arayışlara yöneltilmesinin nedeni geleneksel sorgu ve raporlama araçlarının veri yığınları karşısında yetersiz kalması olarak belirtilmektedir. Veri tabanlarında öz bilgi keşfi süreci içerisinde, model kurulması ve değerlendirilmesi aşamalarında ortaya çıkan veri madenciliği (data mining) en önemli bölümü oluşturmaktadır (Akpınar, 2000).

Veri madenciliğin Veri Tabanlarında Öz Bilgi Keşfi'nin temelini oluşturan keşif kısmının gerçekleştiği adım olarak değerlendirilebileceği gibi bağımsız süreç olarak da değerlendirilmektedir.

2.2.2. Veri madenciliği süreci ve aşamaları

Şekil 2.2.2.1'de veri madenciliği süreci görsel olarak özetlenmektedir. Şekilde işlemsel veri tabanı, veri ambarı, veri madenciliği, yorumlama değerlendirme ve sonuç arasında çeşitli bağları gösteren bir görsel yer almaktadır (Akpınar, 2000). Farklı bilgi sistemlerinde dağınık bulunan verilerden önemli bilgileri alabilmek için veri ambarı

oluşturmak kullanılan bir tekniktir. Veri analizi ve veri madenciliği tekniklerinin kullanılmasının bir nedeni ise veri ambarındaki bilgilerin daha iyi kullanılabilmesi içindir (Kayaalp ve Başarslan, 2018).



Şekil 2.2.2.1: Veri Madenciliği Süreci (Oğuzlar, 2004).

Beş temel sürecin sırasıyla izlendiği veri tabanlarındaki öz bilgi keşfi aşağıda belirtilmektedir (Akpınar, 2000);

1. Problemin belirlenmesi: Hangi işletme amacı için uygulamanın gerçekleştirileceğinin açık olarak belirlenmesidir. İşletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş ilgili işletme amaçlarını taşımalıdır.
2. Ön işleme / verileri hazırlama: Yeni veri setlerinin toplanması, betimlenmesi ve temizlenmesi her veri hazırlama sürecinin başında yer almaktadır. Süreç tamamlandıktan sonra toplanan veriler analiz edilebilir ve bu şekilde sonuçlara ulaşılabilir (Dasu ve Johnson, 2003). Veri madenciliğinde verinin kalitesi anahtar bir konuyu oluşturmaktadır. Güvenilirliğin artırılması için veri madenciliğinde, veri ön işleme yapılmalıdır (Oğuzlar, 2003). Veri hazırlama aşaması basamakları;

- Verilerin toplanması: Tanımlanan problem için gerekli olduğu düşünölen verilerin veri kaynaklarının belirlendiğı basamaktır (Akpınar, 2000).
- Verilerin temizlenmesi: Eksik verilerin tamamlanması, ilgisi olmayan değlerlerin belirlenmesi amacıyla düzeltilmesi ve verilerdeki tutarsızlıkların giderilmesi gibi işlem basamaklarından oluşmaktadır.
- Verilerin birleştirilmesi: Verilerin tek bir çatı altında (veri ambarında) birleştirilmesini sağlamak amacıyla farklı veri tabanlarının kullanıldığı aşamadır. Böylece farklı işletmelerin çatısı altındaki verilere sahip olma, bu verilerin işlenmesi, depolanması, iletilmesi ve veri ambarlarında birleştirilmesine yönelik ilerlemeler sağlanmaktadır. Bir diğer deyişle; farklı zaman boyutlarında elde edilmiş verilerin birbiri ile konuşmasını, iletişim kurmasını sağlayan, tutarlı ve doğru verilerin yer aldığı sistemdir (Şentürk, 2006).
- Verilerin dönüştürölmesi: Bazı verileri veri madenciliğinde kullanabilmek adına farklı formlara dönüştürmek gerekmektedir. Bunu sağlamak amacıyla da veri dönüştürme işlemi yapılmaktadır. Verileri uygun forma dönüştürme ve düzeltme, bir araya getirme, genelleştirip normalleştirme işlemlerini içermektedir.
- Verilerin indirgenmesi: Çekirdek ve küçük hacimli veri kümesi elde etmek amacıyla daha büyük ve karmaşık veri kümelerinin kullanıldığı işlem basamağıdır. Burada hedef verileri daha küçük boyutlara indirmektedir (Chien and Chen, 2008).

3. Veri madenciliğı modelinin kurulması ve değlendirilmesi: Belirlenen probleme yönelik en uygun modelin seçilebilmesi için, çok sayıda modelin denenmesi gerekmektedir. Dolayısıyla en iyi olduğu düşünölen modele varıncaya kadar veri hazırlama ve model kurma aşamaları, tekrar eden bir süreçtir.
4. Modelin kullanılması: Model kurulduğu ve geçerliliğı kabul edildiğinde doğrudan bir uygulama yapılabileceğı gibi, başka bir uygulamanın alt parçası olarak da kullanılabilir.
5. Modelin izlenmesi: Kurulan modelin sürekli olarak takip edilmesi ve gerekiyorsa yeniden düzenlenmesinin nedeni bütün sistemlerin özelliklerinde ki zaman içerisinde ürettikleri verilerde ortaya çıkan değışikliklerden kaynaklanmaktadır (Tüzüntürk, 2010).

Daha önce bilinmeyen kalıpları bulma tekniğı olan veri madenciliğı, gerekli verileri bulmak ve büyük miktarda veriyi aramak için kullanılan mantıksal bir süreç olarak bilinmektedir. İlk kez bulunan ve oluşturulan modeller, belirli kararları almada

işletmelerin gelişimi için kullanılmaktadır (Parker, 2004). Keşif, model tanımlama ve dağıtım bu model gelişimi için belirlenen ilk üç adımı oluşturmaktadır. Bu belirlenen üç adımın sırasıyla özellikleri aşağıdaki gibidir:

- İlk adım olan veri keşfinde, veriler temizlenip farklı forma dönüştürülür, önemli olan değişken tipleri ile daha sonra probleme dayalı verilerin doğası kararlaştırılır.
- Veriler araştırılıp, rafine edilip, belirli değişkenler için tanımlandıktan sonra model tanımlama basamağı için desen tanımlaması yapılmaktadır. Dolayısıyla en iyi tahmini yapan kalıpları belirlemek ve seçmek gerekmektedir.
- Son basamak olarak elde edilmesi istenen sonuç için kalıplar dağıtılmaktadır (Baharati and Ramageri, 2010).

Veri tabanlarında bilgi keşfi için sınıflandırma, kümeleme, regresyon, yapay zeka, sinir ağları, birliktelik kuralları, karar ağaçları, genetik algoritma, en yakın komşu yöntemi gibi çeşitli algoritma ve teknikler kullanılmaktadır (Han and Kamber, 2006). Veri madenciliği için üç temel işlem kullanılabilir;

-Sınıflandırma (Classification): Veri madenciliğinde en yaygın kullanılan sınıflandırma tekniğidir. Bu teknik, kayıt popülasyonunu genel olarak sınıflandırabilen bir model geliştirmek için bir grup önceden sınıflandırılmış örnek kullanılarak gerçekleştirilmektedir. Öğrenme ve sınıflandırma, veri sınıflandırma sürecine dahildir (Baharati and Ramageri, 2010). Yaygın kullanılan sınıflandırma modeli, sınıfı tanımlanmış veriler sayesinde sınıfı tanımlanmamış olan verilerin sınıfını tahmin etmek için kullanılmaktadır. Sınıflandırma modelini iki adım oluşturmaktadır. Tahmin için kullanılacak modeli belirlemek ilk adım iken ikinci adımda oluşturulan modelin sınıfı belli olmayan veri grupları üzerinde uygulaması yapılarak sınıfları tahmin edilmektedir (Han ve Kamber, 2006). Yapay Sinir Ağları (Artificial Neural Networks), Genetik Algoritmalar (Genetic Algorithms), K- En Yakın Komşu (K-Nearest Neighbour), Bellek Temelli Nedenleme (Memory Based Reasoning), Naive – Bayes, Lojistik Regresyon (Logistic Regression) ve Karar Ağaçlarıdır (Decision Trees) başlıca sınıflandırma teknikleri arasında yer almaktadır (Albayrak ve Yılmaz, 2009).

-Kümeleme (Clustering) modelleri: Benzer nesne gruplarının tanımlanması kümelemedir. Kümeleme tekniklerinin kullanım amaçları: nesne uzayındaki yoğun ve

seyrek bölgeleri daha fazla tanımlamak, veri öznitelikleri arasındaki genel dağılım modelini ve korelasyonları keşfedebilmektir (Baharati and Ramageri, 2010).

-Tahminleme: Tahmin için uygulanabilecek teknikler arasında regresyon metodu vardır. Bir veya birden fazla bağımsız değişken ile bağımlı değişken arasındaki ilişki regresyon analizi kullanılarak model oluşturabilir. Bilinen nitelikte olan değişkenler veri madenciliğinde bağımsız değişkenlerdir. Değişkenleri tahmin etmek için yanıt kullanılmaktadır. Çoğunlukla gerçek dünya problemi tahminden ibaret değildir. Sınıflandırma ve regresyon modellerinin ikisini de kullanarak, sinir ağları da model oluşturabilmektedir (Han and Kamber, 2006).

Veri madenciliği, genel olarak verilerin farklı bir bakış açısından analiz edilmesi ve kullanışlı bilgi haline dönüştürülmesi süreci olarak değerlendirilir. Teknik olarak değerlendirildiğinde veri madenciliği, büyük ve birbiriyle ilişkisi bulunan veri tabanları içerisindeki düzinelerce alan arasından korelasyonlar ve düzenler bulma sürecidir. Bu bölüme kadar yapılan birçok veri madenciliği tanımının nedeni, asıl konu üzerinde fikir birliği oluşturulmuş ortak bir tanımının bulunmamasından kaynaklanmaktadır (Akpınar, 2000).

2.3. Sağlıkta Veri Madenciliği

Bilginin içerik ve yapısal anlamda en hızlı değiştiği alanlar arasında sağlık sektörü yer almaktadır. Sağlık hizmetlerinin en hızlı, en doğru, en yüksek kalitede ve ihtiyaca cevap verecek şekilde sunulabilmesi için sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşması ve bu bilgiyi karar destek sistemlerinden faydalanarak kullanması gerekmektedir (Stolba and Tjoa, 2006). Veri Madenciliği büyük miktarda veri içerisinde, gizli kalmış, değerli, kullanılabilir bilgileri açığa çıkarmak ve stratejik karar destek sağlamak amacıyla kullanılan; verilerin analizini temel alarak karar verme modelleri yaratan bir yöntemdir. Sağlık hizmetlerinin sunumu, her düzeydeki sağlık kurumlarının yönetimi ve sağlık politikalarının oluşturulmasında bir karar destek aracı olarak Veri Madenciliği'nin kullanılması ile sağlık profesyonellerinin en optimal kararları almasına yardımcı olacağı düşünülmektedir (Koyuncugil ve Özgülbaş, 2009).

Sağlık Bakanlığı yaptığı değerlendirmeler ile sağlık alanında politika üretmek için ülkemizde hayati öneme sahip verilerin toplanmasında, saklanmasında ve analiz

edilmesinde ulusal veya uluslararası standartların olmadığı, özellikle veri toplama konusunda ciddi bir karmaşanın mevcut olduğu tespitinde bulunmuş ve “Sağlıkta Dönüşüm Programı” kapsamında “Karar Sürecinde Etkili Bilgiye Erişim: Sağlık Bilgi Sistemi” başlığı ile çalışmalar başlatmıştır. Veri Madenciliği çözümleri başvurusunun kaçınılmaz olmasının nedeni Sağlık Bakanlığı’nda toplanan verilerin analiz amaçlı göstergelere dönüştürülmesi için karar vermede yol gösterici modeller oluşturmaktadır. Bu ifadeye göre doğru sonuçlara ulaşmak için, doğru veri madenciliği çözümüne başvurulması gerekli görülmektedir. (<https://www.saglik.gov.tr/>) (Erişim Tarihi: 05.12.2022).

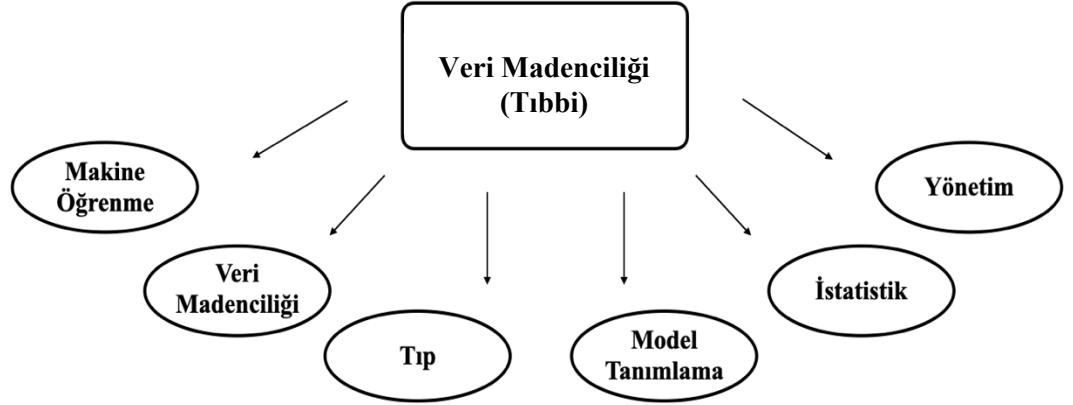
2.3.1. Veri madenciliği görevleri ve sağlık hizmetleri alanında kullanımı

Veri madenciliği, tıp ve veri madenciliği alanının katkı sağladığı büyük bir olgudur. Yüksek düzeyde doğruluğu ve hızı sağlama, sağlık hizmeti madenciliği üzerinde çalışırken temel bir kriterdir. Tıbbi verilerdeki bazı olumsuzluklar gün geçtikçe veri madenciliğinin önemini vurgular niteliktedir (Pradhan et al., 2020). Örneğin, boyutu yüksek olan tıbbi veriler, yönetiminin manuel olarak güçlüklerle gerçekleştirilebildiği veriler olması, insan hatalarında ki ciddi artış ve bu durumdan kaynaklanan olumsuz sonuçların olasılıkları, heterojen veriler, verilerdeki sonsuz düzenlilikler ve bilgilerin yetersiz olması vb. gibi durumlar söz konusu olan tıbbi veri madenciliğine olan ihtiyacı artırmaktadır. Zorlukların daha büyük ölçekte çözülmesine veri madenciliği uygulamaları yardımcı olabilmektedir (Dabas and Singh, 2021).

Veri madenciliği, kısmen yeni bir teknoloji ve tam olgunlaşmamış olmasına rağmen çok sayıda şirket bu teknolojiyi kullanmaktadır. Bulunması güç olan kalıpları ve bağlantıları tespit edebilmek için veri madenciliğinden yararlanılabilir. Veri madenciliği teknolojisini kullanan çok sayıda popüler şirket, müşterileri hakkında daha fazla bilgi sahibi olma ve aynı zamanda akıllı pazarlama kararları verebilme olanaklarını sağlayabilmektedir (Baharati and Ramageri, 2010).

Literatürde yer alan çalışmalarda, her zaman yararlı olanın belirli bir veri kümesinde yer alan en faydalı bilgileri çıkarmak için tıbbi veriler üzerinde veri madenciliği tekniklerini kullanmak olduğu belirtilmektedir. Farklı disiplinlerin

alıřma konusu olan tıbbi veri madencilięi ile ilgili alanlar řekil 2.3.1.1’de yer almaktadır (Dabas and Singh, 2021).



řekil 2.3.1.1: Tıbbi Veri Madencilięi Disiplinleri (Türkeye evirilmiřtir) (Dabas and Singh, 2021).

2.3.2. Saęlık bilgi yönetimi ve veri madencilięi

Verilerin ierisindeki gizli olan kalıpları ve bilgileri keřfedip, bu bilgileri veri madencilięi sayesinde otomatik ıkarmak iin kullanılan bir alandır. Veriler gözlemlenerek özömlenemedięi iin gizli kalıplar olarak deęerlendirilmektedir (Koh and Tan, 2011). Bařlangıta yapay zeka olmak üzere, bilgisayar bilimi, makine ğrenmesi, veri görselleřtirme ve veri tabanı yönetimi ile matematik algoritmalarını, istatistięi birleřtiren veri madencilięi disiplinler arası bir alanı oluřturmaktadır. Veri tabanlarında bilgi keřfi iin teknoloji saęlayan veri madencilięi yöntemleri ile karar verme, problem özme, analiz yapabilme, planlama, teřhis, tespit, entegrasyon, önleme, ğrenme ve yenilik alanları farklı metodolojileri oluřturmaktadır (Liao, 2003).

Saęlık sektöründe kullanılan veri madencilięi alanları, müřteri memnuniyeti ve verimlilik aısından deęerlendirildięi zaman kurumda alıřan saęlık personellerinin performans izlenmesinin büyük bir etken olduęu düşünölmektedir (Verma et al., 2019).

Müřteri memnuniyetini saęlayabilmek, saęlık sektöründe alıřan yönetim ekibindeki en alt personelden, en üst personele kadar herkesi ilgilendiren bir süreçtir. Saęlık sektöründe istenilen bařarıyı elde etmek ve müřteri potansiyelini arttırmak iin

sağlık personellerinin değerlendirilmesi ve buna yönelik eğitilmesi gerekliliği bulunmaktadır (Fayyad et al., 1996). Sağlık sektöründeki kurum maliyetlerini düşürmek amacıyla veri madenciliğinin kullanılması sistem verimliliğini artırmaktadır. Hastanın hastanedeki yatış süresinin azaltılması ile birlikte kurumun gider yönetim sistemine katkı sağlayarak hastane maliyetlerinin düşürülmesi sonucunda kurumun sektördeki payı geniş alana yayılacaktır. Hizmet kalitesini arttırabilmek için sağlık sektörünün birçok alanında iyileştirme çalışmalarına ihtiyaç duyulacağı belirtilmiştir (Karabulut, 2021).

Sağlık sektörü alanında veri madenciliği yöntemleri ile tıbbi tedavi etkinliğini değerlendirmede katkı sağlayabileceği düşünülmektedir. Hastalık nedenlerini, görülen semptomları karşılaştıran veri madenciliği yöntemleri hangi tedavi şeklinin daha etkili olduğunu belirten bir analiz sunabilmektedir (Cios, 2012). Bu yöntemin birçok noktaya katkı sağladığı bilinmektedir, veri madenciliği uygulamaları sayesinde bazı değerlendirmelerin örnekler; aynı hastalığa sahip bireylerin farklı ilaçlarla tedavi edilmesinin sonuçları, yapılan tedavi yaklaşımları ile ilgili diğer veri madenciliği uygulamaları, tedavi sürecinde ortaya çıkan yan etkilerin değerlendirilmesi, farklı yanıt veren alt popülasyonları tedavi etmek için en etkili ilaç bileşiklerini belirlemeyi içermektedir. Öncelikli olarak popülasyondan belirli ilaçlara geçiş ve yan etkilerin riskini azaltmaya yönelik proaktif adımların gerçekleştirilmesidir (Han et al., 2022). Aynı zamanda bu uygulamaların sağlık hizmeti müşterisinin satın alma olasılığı olan diğer ürünleri, bir hastanın reçete edilen tedaviye uyma olasılığının olup olmadığını veya önleyici bakımın gelecekteki kullanımda önemli bir azalma sağlayıp sağlamayacağını tahmin etmek için de kullanılabilir (Diwani et al., 2013). Verilen bu örnekler sağlık alanı için kullanılan bilgi yönetiminde veri madenciliğinin önemini vurgulamaktadır.

2.3.3. Veri madenciliğinin genetik disiplindeki önemi

Günümüzde veri madenciliği genetik temelli yolların aydınlatılmasında belirli algoritmaları takip etmektedir (Albayrak ve Yılmaz, 2009). GA'nın temelini aşağıdaki basamaklar oluşturmaktadır ve bu basamaklar şu şekilde çalışmaktadır:

- Birinci basamak, çözüm grubunun oluşturulmasında mümkün olabilecek çözümlerin kodlandığı kısımdır. Kromozom çözümlerin kodları için kullanılır iken popülasyon çözüm grubuna verilen isimlendirmedir. Popülasyonda yer

alan birey sayısını tespit etmek bu basamağın ilk evresidir. Problemin karmaşıklığı ve aramanın derinliği popülasyon büyüklüğü seçimi için önem göstermektedir. Bu aşama sonrası rastlantısal olarak popülasyon belirlenmektedir.

- İkinci basamak, bütün kromozomların başarısının belirlenmesi ve başarıyı belirleyen fonksiyona uygunluk fonksiyonu adı verilmektedir. Bu fonksiyonun çalıştırılarak kromozomların uygunluklarının bulunmasına ise hesaplama denir. Genetik algoritmada bu fonksiyon probleme özel işleyen genetik algoritmanın beynini oluşturan tek bölümdür. Çoğunlukla uygunluk fonksiyonunun verimli ve hassas olması genetik algoritmanın başarısına bağlıdır.
- Üçüncü basamak, yeni bir popülasyon elde etmek için bu kromozomların eşleştirilmesi ve tekrar kopyalanıp değiştirme operatörleri kullanılarak hazırlanması gerekmektedir. Kromozomların uygunluk değerlerine göre eşleştirilmesi gerçekleştirilir. Seçimi yapmak için çeşitli yöntemler kullanılabilir.
- Dördüncü basamak, sabit büyüklükte bir popülasyon elde edebilmek amacıyla yeni kromozomlara yer açmak için eski kromozomlar ortadan kaldırılmaktadır.
- Beşinci basamak, yeni popülasyonun başarısının belirlenebilmesi için bütün kromozomların uygunlukları yeniden hesaplanmaktadır.
- Altıncı basamak, birden çok kez çalıştırılan GA ile çok sayıda popülasyon oluşturulup bu popülasyonların başarıları hesaplanmaktadır. Zaman var ise üçüncü basamağa geri dönülmektedir.
- Yedinci basamak, kromozom sonucunun en iyisini vermektedir. En iyi bireylerin popülasyon hesaplanmasında saklanmamış olması bu durumun nedenini oluşturmaktadır (Oğuz ve Akbaş, 1997).

Örüntü tanımada GA: GA'nın doğrudan hesaplanmasında sınıflandırıcı olarak kullanılmaktadır.

GA'ların ikinci yönteminde sonuçları optimize etmek için bir GA kullanılmalıdır, örneğin örüntü tanımlamada GA'ların çoğu uygulaması bazı parametreleri optimize etmektedir. Sınıflandırma doğruluğunu artıran en uygun özellik ağırlıkları kümesini bulmak için GA'lar uygulanmıştır. Temel Bileşen Analizi (PCA) gibi bir çıkarma

yöntemi öncelikle geleneksel bir özellik olarak uygulanır. Hemen sonrasında uygunluk fonksiyonunu GA'lar da hesaplamak için k-NN (En yakın komşu algoritması) gibi bir sınıflandırıcı kullanılmaktadır (Siedlecki and Sklansky, 1989). Bu algoritma GA'ları optimize etmek için sınıflandırıcıların kombinasyonlarının kullanıldığı başka bir alandır. Durum tabanlı sınıflandırmada prototip seçiminde GA'lar kullanılmaktadır (Pei et al., 1998).

GA'ların ikinci yöntemi olan veri kümesinden elde edilen sonuçları optimize etmek ve verilerin gözlemlerinin doğru değerlerini hesaplamak için veri madenciliği yöntemlerini uygulamanın daha etkili olduğu tespit edilmiştir (Verma and Verma, 2012). İş dünyasında GA'ların kapsamı büyüktür. İş karlarını optimize etmek ve sonuçlarını işlemek için çeşitli veri madenciliği teknikleri kullanılarak filtrelmesi gereken büyük miktarda veri bulunmaktadır. Birçok alanda uygulama yapılabilmektedir:

- Optimizasyon: Bir iş problemini belirli değişkenlerle ve iyi tanımlanmış bir kâr tanımı ile kârı optimize etmek amacıyla gerekli olan en uygun değeri otomatik saptamak için GA'lara ihtiyaç duyulmaktadır (Forrest, 1993).
- Tahmin: Diğer veri madenciliği algoritmalarını optimize etmeye yardımcı olmak için kullanılan GA, meta düzey operatörlerdir. Bir örnek ile değerlendirmek gerekirse piyasa analizinde optimal birliktelik kurallarını bulmak için kullanıldığı bilinmektedir.
- Simülasyon: Belirli iş sorunları bazen simülasyonda, ne kadar kâr olduğunu veya çözüm yolunun diğerinden daha iyi olup olmadığını iyi tanımlayamamaktadır. Dolayısıyla fazla mesai basit etkileşim kuralları aracılığıyla bazı iş insanlarının simüle etmek istedikleri çok sayıda varlık bulunmaktadır (Verma and Verma, 2012).

2.4. Veri Madenciliği Algoritmaları

Büyük veri setinden faydalı ve yararlı bilgileri elde edebilmek için kullanılan veri madenciliği, bir sınıflandırma yöntemi için karar verme veya tahminde bulunma gibi prosedürlerde uygulanmaktadır (Muhammad et al., 2017). Aynı zamanda

karmaşık verilerin arasındaki ilişkiyi analiz edebilmek için de kullanılmaktadır (Hussain et al., 2019) (Tablo 2.4.1).

Tablo 2.4.1. Veri Madenciliği Görevleri ve Veri Teknikleri Arasındaki İlişki

Veri madenciliği teknikleri	Özetleme	Karakterizasyon	Sınıflandırma	Kümeleme	Bağlantı	Aykırı değer analizi	Regresyon analizi
İstatistik	✓	✓	✓	✓	✓	✓	✓
Makine öğrenmesi		✓	✓	✓	✓	✓	✓
Sinir ağı		✓	✓	✓	✓	✓	✓
Veritabanı sistemi	✓	✓			✓	✓	
Genetik algoritmalar			✓	✓	✓	✓	
Bulanık küme ve mantık		✓	✓	✓	✓	✓	
Görselleştirme		✓	✓	✓		✓	✓

Kullanılan yöntemlerden bir tanesi kestirim yöntemidir. Bu yöntemin sınıflandırma yönteminden en önemli farkı kestirim algoritmalarının daha farklı prensip ile çalışıyor olmasıdır.

Kestirim yöntemi tahmin edilen sınıf üzerinde gerçekleştirilir dolayısıyla bu doğruluğu bekle-gör prensibi ile uygulanmaktadır. Doğruluk kestirimi, hız, dayanıklılık, ölçülebilirlik ve açıklanabilirlik gibi kriterler sınıflandırma ve kestirim yöntemlerini karşılaştırılmasında kullanılmaktadır (Gündoğdu, 2007). Sınıflandırma ve kestirim modelleri için kullanılan başlıca teknikler verilmiştir (Diler, 2016) ;

2.4.1. Yapay sinir ağları

İnsan beyninden esinlenerek geliştirilen Yapay Sinir Ağları (Artificial Neural Networks), ağırlıklı bağlantılar aracılığı ile birbirine bağlanan işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır (Tablo 2.4.1). Deneyim / tecrübelerden yararlanarak öğrenebilmesi en önemli özelliği arasındadır. İnsan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yeteneklerini herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile yapay sinir ağları, geliştirilmiştir. Bilgiler

arasında ilişkiler oluşturma özelliği öğrenmenin yanı sıra kullanılan yapay sinir ağları yeteneklerindendir (Uğur ve Kınacı, 2006). Yapay sinir ağları yapay zekanın önemli yan dalları arasında yer almaktadır. Yapay sinir ağlarının günümüzde en az uzman sistemler kadar yaygın bir uygulama sahasına sahip olması ile beraber her geçen gün bir başka uygulamayla hayat bularak kullanım alanını genişletmektedir (Batal, 2016). Karmaşık veya net olmayan verileri anlamlandırma konusunda önemli görevi olan yapay sinir ağları sayesinde diğer bilgisayar teknikleri ile fark edilemeyecek olan karmaşık kalıpları çıkarmak ve eğilimlerini tespit etmek amacıyla kullanılabilir (Bharati and Ramageri, 2010).

Yapay sinir ağları için belirlenen temel işlevlerin açıklaması aşağıdaki gibidir:

- Öngörü (Prediction) veya tahminleme: İleriki satışlar, hava tahminleri, at yarışları, çevresel risk, ...
- Sınıflandırma (Classification) ve Kümeleme (Clustering): Müşteri profilleri, tıbbi teşhis, ses ve şekil tanıma, hücre tipleri ...
- Kontrol (Control): Erken uyarı için uçaklarda ses ve titreşim düzeyleri, ... Ayrıca, Veri Birleştirme (Data Association), Kavramsallaştırma (Data Conceptualization) ve Filtreleme (Data Filtering) için de kullanılabilir.

Ayrıca yapay sinir ağlarının endüstriyel uygulamalar, finans uygulamaları, askeri ve savunma uygulamaları, tıp ve sağlık uygulamaları, mühendislik uygulamaları, robot bilim, görüntü işleme, örüntü tanıma dışında iletişim sanayi, eğlence amaçlı tahmin gibi özel uygulama alanları da bulunmaktadır. Bir yapay sinir ağı hücresinin girdi, ağırlık, toplama işlevi ve çıktısı temel prensibini oluşturmaktadır (Uğur ve Kınacı, 2006).

Yapay sinir ağları için belirtilen bir diğer önemli tanım “Yapay sinir ağları, deneysel bilgileri doğal olarak depolama eğilimi olan ve bu depolanan bilgilerin kullanılmasına imkan veren basit elemanlardan oluşan paralel dağıtılmış bir işlemcidir.” olarak ifade edilmiştir. Yapay sinir ağlarının çalışma prensibinin şematize edilmesi ile bilgisayarların en basit haliyle önce bir öğrenme sürecinden geçtiği, ardından öğrendiği bu bilgileri karşılaştığı olaylarda tepki olarak kullanması mantığı ile çalıştığı saptanmıştır (Batal, 2016). Veri ile ilgili kalıpların ve eğilimlerin belirlenmesinde sinir ağları en iyi yöntemlerden biridir ayrıca tahminleme veya tahmin ihtiyaçları için uygun bir metod olarak kullanılmaktadır (Bharati and Ramageri, 2010).

2.4.2. Karar ağacı

Normal bir ağaç kök, dal ve yapraklardan oluşmaktadır. Karar Ağacında da aynı yapı izlenmektedir. Kök düğümü, dalları ve yaprak düğümlerini içermektedir. Bir özniteliğin test edilmesi düğümdedir, testin sonucu dal ve sınıf etiketinde bulunmaktadır (Gershman et al., 2010; Jadhav et al., 2016). Kök düğüm, tüm düğümlerin ebeveynidir ve adından da anlaşılacağı gibi, karar ağacındaki en üst düğümdür. Karar ağacı, her düğümün bir özelliği (nitelik), her bağlantının (dal) bir karar (kural) gösterdiği ve her yaprağın bir sonucu (kategorik veya sürekli değer) gösterdiği bir yöntemdir (Jadhav et al., 2016).

Karar Ağacı, insanın karar verme sürecine benzemektedir ve anlaşılması kolaydır. Girdi olarak kesikli veya sürekli veri olup olmadığını her iki durumu da çözebilmektedir (Anyanwu and Shiva, 2009). Bilimsel literatürde karar ağaçları C 4.5 olarak bilinmektedir ve ID3 olarak bilinen algoritmanın geliştirilmesi ile ortaya çıkmıştır (Quinlan, 1993). Karar Ağacının özelliklerinden olan ID3 algoritması, sadece WEKA aracı üzerinde simüle edilebilmektedir ve veri setinin veri tipi sadece kategorik olmasıdır (Priyam et al., 2013). C4.5' in, ID3 algoritmasından farkı normalizasyon işlemlerinin bu algoritmaya dahil edilmesidir (Quinlan, 1993). ID3 simülasyon için sürekli veri seti almamaktadır. Benzer şekilde, CART ve C4.5, ID3 ile aynı özelliklere sahiptir. Tek fark C4.5 ve CART, simülasyon amacıyla sürekli veri setini girdi olarak alabilmektedir (Nadhammai and Hemalatha, 2014). Tablo 2.4.2.1' de detaylı gösterilmiştir (Çavuşoğlu ve Kaçar, 2019). Bilgi kazancı değerleri algoritma ile hesaplanmakta ve oran olarak bu değerler kullanılmaktadır. Karar ağaçları oluşturulurken aynı zamanda alt ağaçlar oluşturulmaktadır ve bu alt ağaçların farklı seviyelere taşınması mümkün olabilmektedir.

Veri setinde bulunan problemlili verilerin ağaçtan silinebilmesi için karar ağacı algoritmalarında dal budama işlemleri gerçekleştirilmektedir. Karar Ağacı oluşturmak için tek bir düğüm tespit edilerek işleme başlanmaktadır. Veri örneklerinin tamamının bir sınıfta yer alması durumunda düğüm yaprak olarak tespit edilmektedir. Dolayısıyla tek bir sınıfı temsil etmektedir. Fakat düğümde farklı sınıflara ait olan özelliklerin bulunması durumun da en iyi bölümlmeyi yapabilecek olan özellik belirlenir ve dallanma karar ağacında devam eder. Tüm özelliklerin bilgi kazancı değerleri hesaplandıktan sonra en yüksek bilgi kazancı değerine sahip olan ağaç karar düğümü olarak kabul edilmektedir. Kabul edilen karar düğümünün altında yeni bir alt ağaç

oluşturulması zorunluluğu ile işlem basamakları devam edebilir. Tüm elemanlar için aynı değer elde edildiyse belirlenen alt gruplarda ilerleme durur ve belirlenen son değer çıkış değeri olarak kabul edilmektedir. Ayırt edici olan bir nitelik belirlenemiyorsa ve alt grupta tek düğüm mevcut ise ilerleme kesilmektedir (Liao and Vemuri, 2002).

Tablo 2.4.2.1. Karar Ağacının Özellikleri (Çavuşoğlu ve Kaçar, 2019).

Karar ağacı algoritmaları	Veri türü	Sayısal veri bölme yöntemi	Olası program
CHAID	Kategorik	N/A	SPSS
ID3	Kategorik	Kısıtlama yok	WEKA
C4.5	Kategorik, sayısal	Kısıtlama yok	WEKA
CART	Kategorik, sayısal	İkili bölmeler	CART 5.0

2.4.3. Bulanık kümeler

Tarihte ilk olarak 1966’da bulanık kümeler Belmann tarafından kullanılmıştır. Daha sonraki yıllarda ilk bulanık kümeleme algoritması 1969 yılında Ruspini ve pek çok sınıflandırma şeması ise 1970 yılında Gitman ve Levene tarafından geliştirilmiştir (Gustafson and Kessel, 1978).

Bulanık kelimesi net olmayan kesin ayırt edilemeyen gibi ifadelerle tanımlanmaktadır. Rasgelelik ve bulanıklık kategorilerini içeren bir belirsizlik durumudur (Ulutagay, 2004). Bir olayın ortaya çıkması durumunda gerçekleşen belirsizlik durumunun sayısal ölçümü rasgelelik olarak tanımlanırken, sözel belirsizlik, bulanıklık olarak ifade edilmektedir. Bulanık kümeleme yönteminin temel mantığı olan sözel belirsizlik bir önermenin $[0,1]$ arasında ölçülebilir fonksiyon haline getirilmesi olarak tanımlanmaktadır (Joopudi et al., 2013). Bulanık kümeleme yönteminde kullanılan sınıflandırma değerleri yani küme üyelikleri bir olasılık dağılımı göstermektedir. Sınıflandırma değerleri / küme üyeliklerine bu olasılıkların yorumlanması ile karar verilmektedir. Bir küme elemanına ait olasılık değeri ilgili elemanın ait olduğu küme için aidiyet derecesini göstermektedir. Aşağıda fonksiyon olarak tanımlaması yer almaktadır (Ruspini, 1970).

- $f : X \rightarrow [0,1]$

Sonuç olarak belirtilen bilgilere göre bir elemanın bir kümeye ait olma olasılığından söz edilirken aynı kümeye ait olmama olasılığından da söz edilebilmektedir. Örnek olarak bir elemanın bir “A” kümesine ait olma olasılığı %80 ise olmama olasılığı %20 olarak belirlenmektedir. Olasılığın %20 olması halinde bulanık kümeleme mantığına göre aynı anda başka bir kümeye ait olması durumu saptanabilmektedir. Bulanık kümeleme algoritmasının klasik kümeleme algoritması ile arasındaki en önemli fark, klasik kümeleme analizinde bir eleman kümeye dahildir ya da değildir. Dolayısıyla bir elemanın aynı anda ait olması ve olmaması olasılığı söz konusu olmamaktadır (Kandemir, 2018).

2.4.4. Destek vektör makinesi

En popüler makine öğrenme algoritmalarından biride Destek Vektör Makeneleridir (DVM) (Kartal, 2020). İlk olarak DVM'lerini Vapnik kullanmıştır ve makine öğrenmesi araştırma topluluğu tarafından ilgi görülen bir konu olmuştur (Vapnik, 1999). Yapay Sinir Ağları ile yakından ilişkisi bulunan DVM'nin çift katmanlı ve ileri beslemeli bir yapay sinir ağına sahip olduğu bilinmektedir. DVM'i 1990'lı yıllarda geliştirilmiştir ve yüksek performanslı algoritmalar için en iyi yöntemlerden biri haline gelmiş ve son derece popüler olmuştur. İstatistiksel öğrenme teorisine dayanan DVM'ler parametrik olmayan bir sınıflandırma tekniği olarak kullanılmaktadır. İkili sınıflandırılmalar için geliştirilen DVM'leri yardımıyla birkaç örnekleme verisi kullanılarak doğru sınıflandırma sonuçlarına ulaşmak mümkün olabilmektedir (Kartal, 2020).

Son literatür çalışmaları, DVM'lerinin genellikle diğer veri sınıflandırma algoritmalarından, sınıflandırma doğruluğu açısından daha yüksek performans sağlayabildiğini belirtmektedir (Chang and Lin, 2011). DVM'ler için kullanılan veriler iki boyutta doğrusal olarak ayrılabilen yapı özelliği göstermemektedir. Verilerin sınıflandırılmasındaki boyut $z = x^2 + y^2$ formülü kullanılarak üç boyuta yükseltilmiştir. Dolayısıyla kullanılan çekirdek çeşidi kernel rbf olarak belirlenmiştir (Atasoy ve Demiröz, 2021).

Çalışmalarda, metin sınıflandırma, elle yazılmış rakam tanıma, ton tanıma, görüntü sınıflandırma, nesne algılama, mikro-dizi gen ifadesi veri analizi ve veri sınıflandırma gibi çeşitli uygulamalarda DVM'lerin kullanıldığı saptanmıştır (Chang and Lin, 2011).

2.4.5. Bayes sınıflandırma algoritması

İstatistik ve olasılık bilimleri oldukça önemli teoremlerdir (Yüksel ve ark, 2018). Olasılığa dayalı sınıflandırma yöntemi Bayes teoremini içerisinde barındırmaktadır (Kızılkaya ve Oğuzlar, 2018). Bu teoriyi ilk açıklayan istatistikçi, teolog ve matematikçi İngiliz Thomas Bayes'in (1702-1761) adına atfen Bayes Teoremi denilmektedir. Karar analizinde ve bir olayın (örn., hastalık) olma olasılığının hesaplanmasında kullanılmaktadır (Yüksel ve ark, 2018). Bayes teoremi, bir sonucun nedenini ararken koşullu olasılıktan yararlanıp, sonucun hangi olasılıklarla, hangi nedenlerden kaynaklanmış olabileceğini bulmaya destek olmaktadır (Kızılkaya ve Oğuzlar, 2018).

Özetlemek gerekirse “bu teorem bir rassal değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi göstermektedir”. Bayes teoremi bu şekli ile bütün istatistikçiler için kabul edilir bir ilişkiyi açıklamaktadır. Bu teorem için ayrıca Bayes kuralı veya Bayes kanunu adları da kullanılmaktadır (Wang and Druzdzel, 2013)

Bayes teoremi için belirtilen formül:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)}$$

Belirsizlik altında tahmin, akıl yürütme ve tanısal karar verme için olasılık dağılımı ve olasılık yasalarından ağ oluşturan algoritmaları BayesNet algoritması gerçekleştirmektedir. Algoritma da diğer bölümlere atanan değerlerin, verilen bir bölümün koşullu olasılığının hesaplanması için kullanıldığı saptanmıştır. Veri kümelerinden çözümlenen algoritmalar için, veri kümesinin özneliliklerini belirten düğümler kullanılmaktadır (Singh, 1997).

2.4.6. Genetik algoritma

GA'ların temel ilkeleri John Holland tarafından belirlenen, doğal seçim faktörlerine dayanan bir optimizasyon ve arama metodunu oluşturmaktadır (Holland, 1992) (Tablo 2.4.1). Evrimsel hesaplamaların bir bölümünü oluşturan GA öğrenme süreci içerisinde ki evrime dayanan bir yaklaşım ile saptanmaktadır. Süreç modellemesi yapan GA problemin çözümlenebilmesi için çok yönlü bir arama ile en

başarılı olanı tespit etmektedir. Güçlü olanların kazandığı, zayıf olanların elendiği bir algoritma temeline dayanmaktadır (Tuğ, 2005).

Popülasyon büyüklüğü, popülasyon seçimi, kodlama ve çaprazlama ve mutasyon oranı ile türleri gibi genel parametreler GA'da bulunmaktadır. Mutasyon oranlarının düşük olması beklenirken genellikle çaprazlama oranlarının yüksek olması beklenmektedir. Genellikle genetik GA'nın değerini popülasyonun büyük olması yükseltmemektedir. Sonucun iyi olması için popülasyon büyüklüğünün bazen fazla olması da beklenmektedir (Aşık, 2019).

Klasik optimizasyon algoritmaları, GA'dan farklıdır. Genetik algoritmada bulunan farklar dört temel alt başlık altında incelenebilir;

- Değişken kümesinin kodlanmış hali ile uğraşan GA değişkenlerin kendisi ile ilgilenmemektedir
- GA'nın aramaya tek bir noktadan başlamaz, bir nokta ailesinden başlamaktadır. Dolayısıyla yerel bir optimuma takılmadan çalışılabilmektedir
- GA hedef fonksiyonunun türevlerini ve bazı ek bilgilerini kullanmamaktadır. GA doğrudan hedef fonksiyonunun kendisini kullanmaktadır
- GA da deterministik olmayan rastgele geçiş kuralları kullanması gibi önemli farklılıkları oluşturmaktadır (Oğuz ve Akbaş, 1997).

2.4.7. Naive bayes algoritması

Belirlenmiş olan veri özellikleri ile farklı veri öğelerini ayırt etmek amacıyla Naive Bayes (NB) algoritması kullanılmaktadır. Dolayısıyla olasılık temelli bir algoritma olarak kullanıldığı saptanmıştır. Yöntem algoritması için sınıflandırma, tanılama, tahmin veya karar verme sürecini Bayes teoremine göre yürütmektedir (Yüksel ve ark, 2018).

NB algoritmasının da, her bir nitelik değişkenini, bağımsız değişken olarak kabul etmektedir. NB algoritmasının, karmaşık olan gerçek veri üzerinde kullanılmasının nedeni sınıflandırıcı, denetimli öğrenmede çok etkili bir şekilde kullanılabilir olmasından kaynaklanmaktadır.

Karakterizasyon için büyük önem taşıyan ve sınıflandırma için gerekli olan eğitim verilerinin çok az ölçüm gerektirmesi NB'nin en büyük avantajıdır (Kamel et al., 2019).

2.4.8. Random forest algoritması

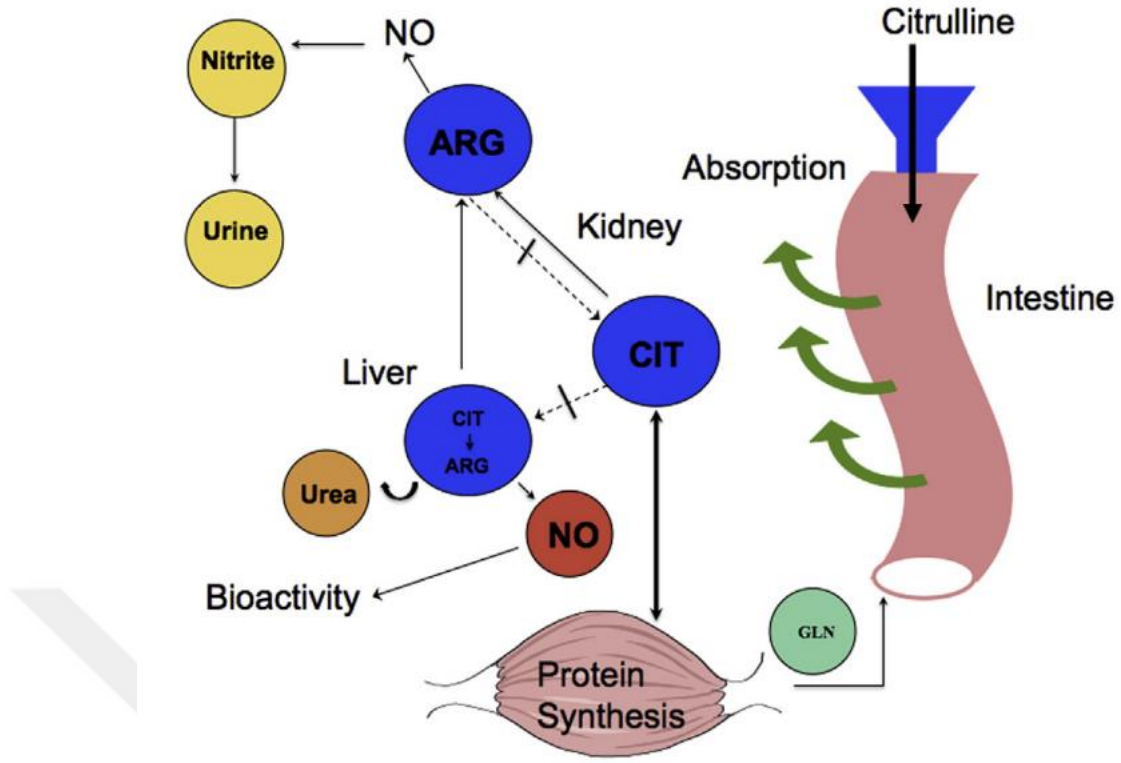
Eğitim sürecinde çok sayıda karar ağacı oluşturarak ve çıktı vererek kontrol edilebilen regresyon ve sınıflandırma için topluluk öğrenme tekniği rastgele orman algoritmasına dayanmaktadır. Rastgele orman, eğitim veri kümesine uygun olan herhangi bir karar için, kullanılabilecek olan en iyi algoritmadır (Mishra and Ratha, 2016).

2.5. Sitrülin Nedir?

GLN metabolizmasının bir son ürünü ve ARG metaboliti olan sitrülin bir amino asittir. İsmi Latince karpuz anlamına gelen *citrullus*'tan almıştır (Fearon,1939). Koga ve Otake 1914 yılında karpuzdan ekstrakte ederek sitrülin tespit etmiştir. Wada tarafından 1930 yılında sitrülin tanımlanmıştır (Cynober et al., 1995).

2.5.1. Sitrülinin tanımı

ARG metaboliti olan sitrülin, GLN metabolizmasının son ürünü olan bir amino asittir (Papadia et al., 2018). GLN, bağırsak tarafından çevrilebilen bir ornitin öncüsüdür. Ek olarak ARG enterositlerde sitrüline metabolize edilmektedir (Cynober et al., 1995). Enterositlerde, argininosüksinat sentaz enzimi bulunmadığından, portal ven içerisinde sitrülin olarak salınır. Karaciğer tarafından sadece az miktarda net sitrülin alımı olması sebebiyle normal karaciğer fonksiyonu varlığında sitrülin sistemik dolaşıma girer ve böbreklerde ARG'e dönüşür. Dolayısıyla tüm vücut seviyesindeki sitrülin net üretimi neredeyse yalnızca ince bağırsak epitelinde gerçekleşmektedir (Şekil 2.5.1.1) (Curis et al., 2005).



Şekil 2.5.1.1: L-Sitrülin Desteğinin Etkileri

Kısaltmalar: ARG L-Arginin; CIT L-Sitrülin, GLN Glutamin, NO Nitrik Oksit (Papadia et al., 2018).

- Sitrülinin elde edilmesi

Sitrülin, ilk olarak karpuz suyunda tanımlanan ve sonra kazeinin triptik parçalanması ile elde edilen esansiyel olmayan bir amino asittir. Kabakgiller familyasına ait acı kavun, salatalık, kokulu kavun, su kabağı, lif kabağı, kış kavunundan da izole edilmiştir. Ayrıca genç ceviz fidelerinde de yüksek konsantrasyonlarda bulunmaktadır, fakat çimlenme sırasında azot translokasyonundaki rolü nedeniyle, cevizin kendisinde ihmal edilebilecek kadar az oranda bulunmaktadır. Sitrülin aynı zamanda beyaz japon huş ağacının özünde de bulunduğu tespit edilmiştir (Rimando and Perkins-Veazie, 2005).

Bilinen en zengin sitrülin kaynağı karpuzdur ve bu amino asidin kuraklık toleransında önemli bir role sahip olduğu düşünülmektedir.

Kalahari çölünden toplanan yabani karpuzun sulandıktan sonra yapraklarında sitrülinin, toplam amino asit içeriğinin %50' sine kadar biriktiği ve beş gün boyunca bu seviyede devam ettiği karpuzdaki sitrülin miktarını saptamak amacıyla yapılan bir çalışmada görülmüştür (Kawasaki et al., 2000). Adaptif mekanizmaya neden olan bu

yeteneğın yalnızca sitozolde bulunan bir enzim ile ilgili olduđu düşünölmektedir (Yokota et al., 2002). Oluşan sitrölin, bir hidroksil radikal temizleyicisi gibi hareket ederek yaprakları kuraklıđa bađlı oksidatif strese karşı koruyabilmektedir (Akashi et al., 2001).

Sitrölinin karpuz gövdesinde ve kabuklarında bulunduđu bildirilmiş olmasına rađmen, raporlar kalorimetrik ölçömlere dayandıđı için amino asidin düzeyi yüksek deđerlerde tahmin edilmiş olabileceđi de bu çalışmanın sınırlılıkları arasındadır (Rimando and Perkins-Veazie., 2005).

2.5.2. Sitrölin metabolizması ve muhtemel metabolik yolakları

Bilindiđi kadarı ile glutamat, GLN ve prolin gibi diyet ile alınan aminoasitlerin metabolize olması sonucu ince bađırsaklarda üretilen sitrölin, ARG sentezinde birincil substrat olarak görev almaktadır. İnce bađırsaklarda ARG'e dönüştürölen sitrölin, daha sonrasında böbrek proksimal tübüllerine taşınmaktadır (Yavuz, 2006). Sitrölinin yaklaşık olarak %80'i böbreklerde ARG'e dönüştürölmektedir (Kaorea et al., 2013). Dolayısıyla metabolizmanın normal fonksiyonunda az miktarda alınan sitrölin karaciđer tarafından sistemik dolaşıma girer ve böylece böbreklerde ARG'e dönüştürölebilir. Sitrölinin ARG'e göre depo oranı daha fazladır bu nedenle biyoyararlanımı yüksektir (Curis et al., 2005).

Sitrölinin metabolize edilmesine ilişkin yolaklar ile ilgili muhtemel bazı mekanizmalar açıklanmıştır ve bu yolaklar üzerinden metabolizmadaki önemi ve yeri anlaşılmađa çalışılmaktadır. Ancak son derece karmaşık olan ilişkiler yumađını açıklayabilmek için detaylı verilere ihtiyaç duyulmaktadır. Bilinen ve sadeleştirilmiş şekli ile mevcut yolaklar aşağıdadır.

Nitrik oksit yolađı

Sitrölin etkisiyle artan ARG ve nitrik oksit (NO), kaslarda daha fazla kan akışının gerçekleşmesini sağlamaktadır. Kaslarda artan kan akışı ile yüksek oksijen ve besin transferi gerçekleşmektedir. Dolayısıyla kaslarda daha fazla enerji üretilmektedir (Kıyıcı ve ark., 2017).

NO, kalp ve dolaşım sistemi üzerine hücresele haberci olarak etki gösterir ve bu sistemler için önemli bir vazoproteksiyon molekülüdür. Ayrıca iyi bir hidroksil

radikali temizleyicisi olması nedeniyle sitr lin g    antioksidan etki g stermektedir (Rimando and Perkins-Veazie, 2005).

Ergojenik destek sportif alanda kullanılan L-Sitr lin, v cut i inde ARG'e d n  en ve NO miktarını artıran bir  r n olarak kullanılmaktadır. L-Sitr lin, karaci erde ve b breklerde  re d ng s n n  nemli bir bile eni olan n tr, esansiyel olmayan bir alfa-amino asittir. Sitr linin kas h crelerine ula masıyla daha verimli sonu lar alındı ı  alı malarda belirtilmi tir. Sitr lin, bu  zellikleri ile egzersiz performansını pozitif y nde etkilemektedir (Crenn et al., 2000).

NO, ins lin duyarlılı ı, enerji substratlarının oksidasyonu ve ayrıca imm n yanıtta kilit rol oynamaktadır. İns lin salgılayan  -h creleri, ARG'den NO ve L-Sitr lini sentezleyen yapıcı ve uyarılabilir enzimlere sahiptir.

Biyoenerji yola ı

Hayvan ve insan  alı malarından elde edilen verilere g re, sitr linin protein sentezi  zerindeki olumlu etkilerinin rapamisinin memeli hedefi (mTOR) yola ı ile oldu unu g stermektedir (Allerton et al., 2018). Fiziksel aktivite s recinde, ATP'nin tekrar sentezlenmesi sonucu, artan periferel ve merkezi amonyak  retimi v cutta yorgunluk ile sonu lanmaktadır. Dolayısıyla amonya ın detoksifikasyonu, egzersiz performansının artırılması i in  nemli bir a amadır. Ornitin ve ARG v cutta farklı fizyolojik etkileri mevcuttur. Bu amino asitler, protein sentezini artırarak b y me hormonunun etki g stermesini sa lamaktadır. Ayrıca bu amino asit destek  r nlerinin kas hipertrofisine katkı sa lamasıda beklenmektedir (Sureda et al., 2010).

Laktik asit yola ı

Sporcularda performans azalmasına neden olan fakt rlerden biri de laktik asit olu umuna ba lı yorgunluktur. Y ksek yo unlukta ger ekle tirilen farklı fiziksel aktivite  e itleri, kanda artan laktik asit seviyelerine neden olabilir. Laktik asit birikimi, egzersiz devam ederken ortaya  ıkan anaerobik yorgunlu a neden olur. Performans kapasitesi d  ebilir ve spor performansı  zerine etkisi vardır.

Karpuzun i erdi i sitr lin, anaerobik glikoliz i leminin bir yan  r n  olan laktik asit birikimini azaltabilir. Yapılan ara tırmalara g re, fiziksel aktivite  ncesi sitr lin deste i, submaksimal fiziksel aktiviteye sahip ki ilerde laktik asit ve oksijen

satürasyonu üzerinde etkilidir (Fastabiqi et al., 2020).

Krebs döngüsünün yolağı

Anaerobik koşullarda, yüksek yoğunluklu fiziksel egzersiz devam ederken kaslar kasılır ve anaerobik glikoliz sürecinde ATP' ye izin verir. Bu durum, hem kanda hem de kasta laktik asit seviyesinin artmasına neden olmaktadır. Fazla çalışmış bir kas grubu, yüksek laktat konsantrasyonlarına rağmen iyi kasılmaktadır. Oksijen kaynağı alındıktan sonra, laktik asit tekrar enerjiye dönüştürülecek olan pirüvik aside dönüştürülecektir. Böylece pirüvik asit, piruvat olarak kullanılabilecek bir enerji kaynağı olduğundan piruvatın kendisi Krebs döngüsüne katılır ve elektron taşınması sonrasında enerji, su (H₂O) ve karbondioksit (CO₂) üretilir (Fastabiqi et al., 2020).

Üre döngüsü yolağı

Karbamoil fosfat sentetaz I (CPS1), mitokondriyal bir enzimdir; pirimidin biyosentezinde görevli ve sitozolik bir enzim olan Karbamoil fosfat sentetaz II'den farklıdır. Karbamoil fosfat, aktiflenmiş bir karbamoil vericisi olarak düşünülebilir; dört enzimatik basamak gerektirerek üre döngüsüne girer.

- Üre döngüsünün birinci basamağında karbamoil fosfat, mitokondri içinde, karbamoil grubunu ornitin transkarbamoilaz tarafından katalizlenen bir reaksiyonla ornitine verir ve sitrülün oluşturur. Mitokondride oluşan sitrülün, sitozole geçer.
- Üre döngüsünün ikinci basamağında ATP gerektiren ve argininosüksinat sentetaz (ASS1) tarafından katalizlenen reaksiyonla sitrülünün üreido grubu ile aspartatın amino grubu sitozolde kondense olur ve argininosüksinat oluşur. Ürenin ikinci amino grubu aspartattan sağlanmaktadır.
- Üre döngüsünün üçüncü basamağında argininosüksinat, argininosüksinat liyaz (ASL) vasıtasıyla reversibl olarak parçalanır ve serbest arginin ile fumarat oluşturur. Fumarat, üre döngüsünü sitrik asit (TCA) döngüsüne bağlamaktadır, TCA döngüsü de üre döngüsüne aspartatargininosüksinat şantı ile bağlanır.
- Üre döngüsünün son basamağında ARG, sitozolik arjinaz enzimi etkisiyle üre ve ornitine parçalanır. Böylece ornitin yeniden oluşmuş olur ve yeni bir üre döngüsü için mitokondriye geçer.

Üre sentezinin hız sınırlayıcı reaksiyonları, CPS1, ornitin transkarbamoilaz ve arjinaz tarafından katalizlenen reaksiyonlardır. CPS1, asetil-CoA ve glutamattan sentezlenen N-asetil glutamat tarafından allosterik olarak aktive edilir; ARG de N-asetilglutamat sentazı aktive ederek N-asetil glutamat oluşumunu artırır.

Üre döngüsü vasıtasıyla azot akımı, diyetin bileşimi ile değişmektedir. Diyet başlıca protein ise yakıt için amino asitler kullanılır ve amino gruplarından fazla miktarda üre oluşur. Şiddetli açlık sırasında kas proteinin yıkılması metabolik yakıtın çoğunu karşılar ve üre oluşumu artırır (<https://www.mustafaaltinisik.org.uk/89-1-19.pdf>), (Erişim tarihi: 01.09.2022).

Arginin biyosentez yolağı

Protein, NO, üre ve kreatinin sentezi için bir öncü olan ARG, hayvan hücrelerinde en yaygın kullanılan amino asitlerdendir. Hücre proliferasyonu, hücre sinyal iletimi, bağışıklık fonksiyonunun korunması ve anti-inflamasyon gibi birçok biyolojik fonksiyonda ARG görev almaktadır.

Yarı esansiyel bir aminoasit olan ARG, sayısız biyolojik süreçte yer almaktadır. Çeşitli enzimatik reaksiyonlar için substrat olan ARG, vücutta bilinen üç ana yol aracılığı ile metabolize edilmektedir. Bu yollar sırası ile:

1. Arginaz, ARG'i L-ornitine metabolize eder,
2. ARG dekarboksilaz, ARG'i agmatine metabolize eder ve
3. Nitrik oksit sentazı (NOS), NO ve sitrülün oluşturmak için ARG kullanır (Brosnan and Brosnan, 2004 ; Gambardella et al., 2020).

Memeli karaciğerinde üre döngüsü veya ARG biyosentez yolu eşsiz bir yol olarak tanımlanmaktadır. Üre döngüsü amonyak detoksifikasyonunun ana yoludur öte yandan tekrar sentez yoluyla ARG üretilmektedir. CPS1, ornitin transkarbamilaz (OTC), ASS1, ASL ve arginaz 1 (ARG1) üre döngüsünde kullanılan enzimlerdendir. Bir enzimin eksik olması durumunda, üre döngüsü bozukluğu meydana gelebilmektedir (Nagamani et al., 2012). Tamamen farklı olan iki enzimin, farklı gen organizasyonu modeli gösteren ARG biyosentezi sonucun

da görülen örneklerden biri; ornitin gibi bir anahtar ürünün oluşumunu katalize etmesidir (Xu et al., 2000).

2.5.3. Sitrülinin genler ile ilişkisi ve veri madenciliğindeki yeri

ARG öncüsü olan sitrülin, amino asit olması dışında geniş bir alana sahiptir ve etkilerinin fazla olması nedeniyle bilim dünyasında büyük ilgi ve merak uyandırmıştır. Dolaşım sistemindeki sitrülin, aslında hepatik bozulmayı önlemek amacıyla ARG'in maskelenmiş bir rolüdür. Sitrülin tedavisinin kısa bağırsak sendromu, orak hücre anemisi, hiperlipidemi, kardiyovasküler hastalıklar, kanser kemoterapisi, erektil disfonksiyon, atletik antreman, yara iyileşmesinin hızlanması ve nörodejeneratif hastalıklar üzerinde faydalı etkileri olduğu saptanmıştır (Kaore, 2019).

3. GEREÇ VE YÖNTEM

3.1. Araştırmanın Yapıldığı Yer ve Zaman, Türü

Çalışmanın gerçekleştirildiği kurum Biruni Üniversitesi alt yapısında yer alan bilgisayar laboratuvarıdır. Açık erişimle ulaşılan National Center for Biotechnology Information Gene Expression Omnibus (GenBank Overview-NCBI-GEO) kütüphanesinden sitrülün ilişkili gen bilgisini içeren tüm *Homo sapiens* (insan) verileri, Biruni Üniversitesi Girişimsel Olmayan Etik Kurulu'nda 17.12.2021 tarihinde ve 2021/64-16 sayılı karar no ile onayı alındıktan sonra incelenmiştir (Ek-1). Veri madenciliği yöntemi ile çalışılacağından tüm ilişkili gen kombinasyonları değerlendirmeye alınmıştır.

- **Araştırmaya alınma ve dışlanma kriterleri:** GenBank Overview-NCBI-GEO kütüphanesinden elde edilen sitrülün ilişkili genler çalışmanın alınma kriterleri arasında yer almaktayken, bu kütüphanede yer almayan genler çalışmanın dışlanma kriterleri arasında tutulmuştur.

3.2. Veri Toplama Araçları

Çalışmada GenBank Overview-NCBI-GEO kütüphanesi, String (Functional protein association networks), Reactome Pathway Database, Wiki Pathway, Kyoto Genler ve Genomlar Ansiklopedisi (KEGG) ve David Gen Pathway veri tabanları kullanılmıştır.

NCBI-GEO kütüphanesinin “GENE” sekmesi bu çalışmadaki sitrülün ilişkili gen verilerine ulaşmak için kullanılmıştır. Gene sayfasındaki arama çubuğuna anahtar kelime olarak “Citrulline” yazılarak ilişkili genler belirlenmiştir. Böylece sitrülün ile ilişkili gen kodlarına ulaşılmıştır. Filtreleme yapılarak *Homo sapiens* (insan) türü seçilmiştir.

3.3. Veri Toplama Yöntemi

3.3.1. *STRING Gene ilişkisi*

String veri tabanı farklı genlerin birbirleriyle olan ilişkilerini analiz etme yöntemi olarak kullanılmaktadır. Genler “*STRING CONSORTIUM*” veri tabanı ile gen-protein ilişkisi incelenmiştir. Genler arasındaki bağlar ile ilişkilendirilen genler hakkında bilgi verilmektedir. Genlerin doğrudan ya da dolaylı etkileri çalışılmaktadır. Çıkarılan genlerin ağırlıklı protein etkileşimlerinin grafiksel gösterimi, biyolojik işlemlerde analizini kolaylaştıran ve fonksiyonel bağlantının yüksek seviyeli bir görünümünü sağlamaktadır (Mering et al., 2003).

Çalışmada bu basamaklardan yararlanılarak NCBI’den insan çalışmaları için kullanılan sitrülün ilişkili genler analiz edilmiştir. İlişkili pathwayler tespit edilerek detaylı değerlendirme yapılmıştır.

3.3.2. *David Gen Pathway*

Gen ID’ leri, gen isimlerine ve görevlerini öğrenebilmek için “*DAVID Bioinformatics Resources*” veri tabanında analiz edilmektedir. *DAVID* veri tabanını yolak (pathway) analizi görebilmek için kullanılmaktadır. Çalışmada sitrülünün ilişkili olabileceği muhtemel metabolik ve biyosentez yollarının aydınlatılması için *DAVID gen pathway* analizinden yararlanılmıştır.

3.3.3. *Reactome*

Reactome, insan üzerindeki yolların, reaksiyonların, süreçlerin manuel olarak küratörlüğünde olan hakemli gözden geçirilmiş veritabanının açık bir kaynağı olarak kullanılmaktadır (Peter et al., 2019). Biyolojik yollarda yer alan mekanizmaları ve küçük molekülleri analiz etmek, yorumlamak ve görselleştirmek için kullanılan Reactome biyoinformatik araçları da içeren ücretsiz, açık kaynaklı veri tabanını içermektedir (Nicholson et al., 2016).

Sistem biyolojisi ile temel ve klinik araştırmaları desteklemek üzere verileri görselleştirme, entegrasyon sağlama ve analiz edebilmek için biyolojik bilgileri arama, düzenleme ve kullanma aşamasında farklı biyoinformatik araçlar sunmaktadır (Leiman and Taylor, 2019). Reactome kavramı sayesinde tek bir düzende tanımlama

yapılabilmektedir. Biyolojik olan süreçlerden sinyalizasyon, metabolizma, transkripsiyonel düzenleme tek formatta değerlendirilebilmektedir. Reactome dışında diğer veri tabanları (KEGG, Panther, vb.) kullanılarak da fonksiyonel etkileşimler elde edilebilmektedir (Kitano, 2002).

Bu çalışma ile Reactome pathway analizden faydalanarak sitrülün ilişkili muhtemel yolların detaylı analizinin yapılması gerçekleştirilmiştir.

3.3.4. KEGG Pathway

KEGG 1995 yılında, orijinal olarak genomdaki genleri manuel olarak oluşturulan yol haritalarına, haritalama prosedürü olan KEGG yol haritalaması ile tamamen sıralanmış genomların biyolojik olarak yorumlanması için entegre bir veri tabanı kaynağı olarak geliştirilmiştir (Kanehisa et al., 2017).

Genomdaki bir dizi genin, hücredeki bir yol veya bir kompleks gibi, daha yüksek dereceli bir biyolojik işlevi temsil eden etkileşimli moleküllerden oluşan bir ağ ile bağlama işlemi KEGG'deki fonksiyonel atama ile gerçekleştirilmektedir.

Üç veri tabanından oluşan KEGG;

- YOL etkileşen moleküller ağı açısından daha yüksek dereceli fonksiyonların temsili,
- GENES tamamen sıralanmış tüm genomlar ve bazı kısmi genomlar için gen kataloglarının toplanması,
- LIGAND kimyasal bileşiklerin toplanması için hücrede, enzim moleküllerinde ve enzimatik reaksiyonlarını oluşturmaktadır (Goto et al., 2000).

Daha önce bildirilenler ile KEGG sisteminin genel mimarisi temelde birdir (Kanehisa and Goto, 2000).

KEGG pathway veritabanı, sinyal iletimi ve diğer hücresel bilgileri göstermektedir. Aynı zamanda biyolojik etkileşimlerin haritalarının bir derlemesidir. Bu bilgiler doğrultusunda diferansiyel olarak eksprese edilen sitrülün ilişkili genlerin KEGG yolları ile yüksek verimli ekspresyon verilerine karşılık gelen genlerin fonksiyonel yapısı analiz edilmiştir.

3.3.5. Wiki Pathway

WikiPathways projesi, bilimsel bilginin toplanması için herkesin fikri üzerine 2007 yılında kurulmuştur. Projenin genel amacı, araştırma topluluğu tarafından toplanan ve küratörlüğünü yaptığı biyolojik yol modelleri için WikiPathways bir veri tabanı olarak bilinmektedir (Martens et al., 2021). Çalışmada sitrülün için ortak olan diferansiyel olarak eksprese edilmiş genler, WIKI yolları ile analiz edilmiştir.

3.4. Biyoinformatik Analiz

Proteinlerin ekspresyon değerleri biyoinformatik analiz ile değerlendirilmektedir. Biyoinformatik analiz sonucunda sayısal bir veri olarak bir genin ne kadar etkisi olduğu verilmektedir.

Aynı anda çok sayıda örneklem sıralanması için yeni nesil dizileme kullanılmaktadır. Dolayısıyla değerlendirilen her örnek için kapsamlı ekspresyon profilleri belirlenmiştir. Biyoinformatik analiz için potansiyel zamana ihtiyaç duyulmaktadır (Eyileten et al., 2018).

Çalışmada uygulanan analiz şeması Şekil 3.4.1’ de gösterilmiştir. Öncelikle protein ifadesinden sorumlu sitrülün verisi CELL.gz formatında NCBI veri tabanından elde edilmiştir. Bu sitrülün verilerini analiz edebilmek için *R version 3.6.2 (2019-12-12)*—“*Dark and Stormy Night*” Copyright © 2019 the R Foundation for Statistical Computing Platform : x86_64-apple-darwin15.6.0 (64-bit) yazılımının da okutulmuştur. Okuma paketi yardımıyla veriler okutulmuştur. Okuma çeşitleri, analizi yapılan verilere göre değişiklik gösterebilir, bazen ikisi de bir arada kullanılabilir. *R* veri tabanını açtıktan sonra kütüphaneleri `install.packages(“ ”)` olarak indirilmesi gerekir, kodlar o kütüphaneler sayesinde çalışmaktadır. İlk seferinde indirmek yeterlidir, tekrardan *R* ekranı açtıktan sonra `library(“ ”)` yapılması yeterli olacaktır. Her kodun kendine özgü kütüphanesi bulunmaktadır, mutlaka onlar çalıştırılmalıdır. Sitrülün verilerini *GenBank Overview-NCBI-GEO* veri tabanından çağırmak için “`getGEO`” kodu kullanılmalıdır ve bunun için kütüphane olarak “*BioBase*” ve “*GEOquery*” kullanılmalıdır. Bu kodlar sayesinde *GenBank Overview-NCBI-GEO* ‘den sitrülün verilerini çağırdıktan sonra verinin içerisindeki bilgiler incelenmiştir.

Başka bir okuma çeşidi olarak CELL.gz formatında sitrülün verisini indirdikten sonra *R* veri tabanında “read.cellfiles” yapılmalıdır. Bu kod indirilen sitrülün verilerini, *R* veri tabanının okumasını sağlayacaktır. İndirilen sitrülün verilerinin içerisinde hasta hakkında bilgiler bulunmaktadır. Bunlar; yaş bilgisi, hastalık bilgisi ya da gen bilgilerini içermektedir.



Şekil 3.4.1: Sitrülün İlişkili Gen Veri Analizi

Çalışma kapsamında analiz edilen gen verisi “Sitrülün ilişkili genler” olarak değerlendirilmiştir. Veriyi okuma şeklinde ayrı ayrı normalizasyon yaparak, verilerin filtrelenmesine yardımcı olmuş ve veriyi daha okunabilir şekilde göstermiştir. Veriler “log2” ölçeğinde normalleştirilmiştir. Hasta bilgisi ile kullanılan veri içeriği; “üre döngüsü” olacak şekilde matrix oluşturmuştur. Kodları yazarken “eBayes” yöntemi kullanılmıştır. Bu yöntemin kullanılmasında ve hesaplanmış t-istatistikleri, diferansiyel ifadelerin log oranları, standart hataların ortak bir değere göre ampirik Bayes moderasyonu ile hesaplanması gerçekleştirilmiştir. Sonuç olarak çıkan veri Bayes ile çalışma bölgesinin gözlenen ve beklenen sayısal veriler değerlendirilerek, sitrülün ilişkili metabolik yolların aydınlatılması sağlanmıştır.

3.5. Veri Seti

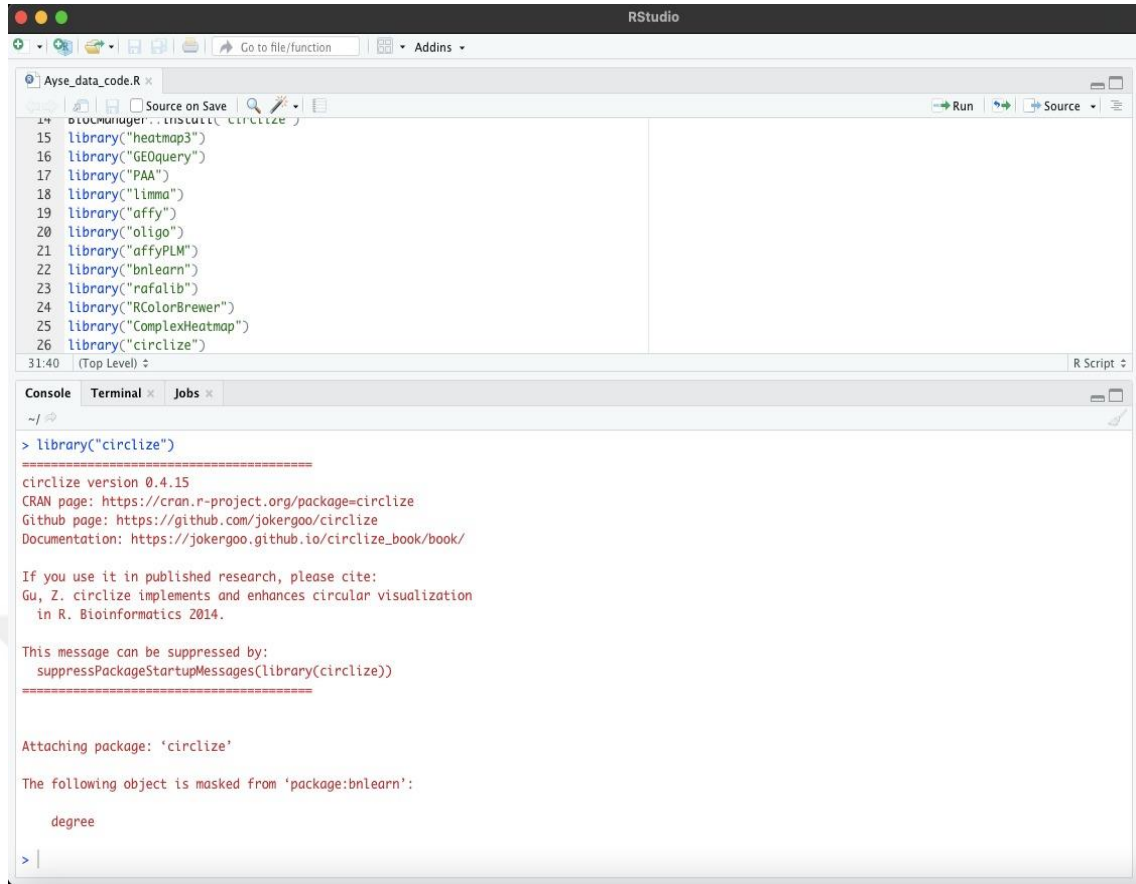
Veri seti terimi, bir veya daha fazla kayıt içeren bir dosyayı ifade eder. Veri kümeleri, metrikler ve varlıklarla ilişkili ölçümleri içermektedir. Bazı veri kümelerinde, hangi varlıkların ölçüldüğü ve tam olarak hangi metriklerin kullanıldığı net olmamakla birlikte, ölçümlerin yanlış yorumlanması da söz konusu olabilmektedir (Rosli et al., 2016). Dolayısıyla bu çalışma için uygun bulunan veri setleri belirlenmiş ve bunların arasından GSE1563 veri seti kullanılmıştır.

3.5.1. GSE1563

Data seti olarak GSE1563 kullanılmıştır. GSE1563 veri setinde 62 örnek bulunmaktadır. Böbrek transplantasyon ve doku hasarı bulunan klinik hasta verisi üre döngüsü (sitrülin) ilişkili 100 gen için değerlendirilmiştir (Flechner et al., 2004).

3.5.2. Kullanılan veri seti ve veri setinin düzenlenmesi

R programı kullanılarak ilgili kütüphaneden data setleri okutularak biyoinformatik analiz yapılmıştır. Sitrülin verilerini *GenBank Overview-NCBI-GEO* veri tabanından çağırmak için “getGEO” kodu kullanılmalıdır ve bunun için kütüphane olarak “BioBase” ve “GEOquery” çalıştırılmalıdır. Bu kodlar sayesinde *GenBank Overview-NCBI-GEO* ‘den sitrülin verilerini çağırdıktan sonra verinin içerisindeki bilgiler incelenmiştir. Sitrülin verilerini analiz edebilmek için *R version 3.6.2 (2019-12-12)—“Dark and Stormy Night” Copyright © 2019 the R Foundation for Statistical Computing Platform : x86_64-apple-darwin15.6.0 (64-bit)* yazılımının da okutulmuştur (Şekil 3.5.2.1).



Şekil 3.5.2.1: R Software Programı Ara Yüzü

3.6. Verilerin İstatistiksel Analizi

Farklı istatistiksel yöntemler kullanılarak yapılan analizler daha iyi kanıt sağlayabilmek için kullanılmaktadır. Veri analizleri, NCBI’da kullanılan data setlerinde önerilen diferansiyel tanımlama tekniğinin performansının mevcut yöntemlerden daha iyi olduğunu göstermektedir. Bu nedenle biyoinformatik analiz yöntemi ile muhtemel metabolik yolak analizine katkıda bulunmuştur.

Anlatımında farklılık görülen genlerin belirlenmesi için her veri seti bağımsız olarak R yazılım programı üzerinden istatistiksel olarak analiz edilmiştir. Gen anlatımı farklılık gösteren genlerin p-değerlerine ($p < 0,05$) göre kat değişimleri FC (fold change) ve değişikliğin yönü (gen anlatımı artan/azalan) analiz edilmiştir. Gen anlatım düzeyi azalan genler $FC < 0,5$; artan genler $FC > 2$ olarak belşirlenmiştir.

Beş farklı, Cluster, PlotMD, VolcanoPlot, BoxPlot, QQ-Plot, analiz yöntemi de LogFc değerlerinin grafiksel gösterimine katkı sağlamıştır. Böylelikle bu beş ayrı grafik analizi sonucunda etkin bulunan genlerin data setlerindeki baskınlık derecesine detaylı yorumlama yapılabilmiştir.

3.6.1. Kümeleme analizi (Cluster analysis)

Gen için içerik oluşturmada “Toptable” kodu kullanarak model uyumundan en uygun ve en üst sıradaki genlerin bir tablosu çıkartılmıştır. Seçilen gen sayısı aralığı toptable yaptıktan sonra cluster oluşturabilmek için rastgele seçilmiştir. Kümeleme analizi (Heatmap) bir ısı haritasıdır. Cluster analizin de hangi gen aralığında anlamlı çıktı ise o gen aralığı kullanılmıştır. Çıkan genlerin logFc oranlarına bakılmış, önemli fark veriyor ise kullanılmıştır. Oluşturulan matris değerlerinin, renk olarak ifade edildiği verinin grafiksel bir gösterimi verilmiştir. Veri matrisinde kontrol ve sitrülün desteği alan grubun genleri hiyerarşik kümeleme sonucundadır.

Veri matrisinde "yüksek" ve "düşük" değerleri vermektedir. Sağ tarafta gen isimleri, aşağı kısımda destek alan grubun kodları bulunmaktadır. Genlerin karşısında “cluster tree” (küme ağacı) dediğimiz bir yakınlık ilişkisini farklı şekilde gösteren bir çizim örneğidir.

Veri matrisini bu şekilde görselleştirmek, her örnek küme için karakteristik görünen değişkenleri bulmaya yardımcı olmuştur. Cluster analizi üzerindeki bilgiler gen analizi için ekstra sonuç vermektedir.

3.6.2. Volcano Plot

Çalışmada VolcanoPlot analizi kullanılmıştır. LogFc değerlerine göre değişimlerinin istatistiksel önemi olarak grafiksel dağılım yapmaktadır. Değerli logFc’deki kat farklarını VolcanoPlot ile genlerin ayrımını iyi bir şekilde yapabilmektedir. LogFc değerlerinin yüksek çıktığı düzenlenmiş genler sağa doğru, düşük değerdeki düzenlenmiş genler sola doğru ve istatistiksel olarak önemli farklara sahip genler yukarı doğru gösterilmiştir.

3.6.3. PlotMD

PlotMD bir istatistiksel analizdir, ortalama grafiği olarak tanımlanmaktadır. İki farklı veri nesnesi için hesaplanan M ve A değerleri ile grafik oluşmaktadır. LogFc değerlerinin sonuçlarıyla renk ayrımı vererek ortalama yoğunluğu vermektedir. PlotMD grafiği ve küme ekspresyon değerlerini ifade etmek için lcpm değerleri

kullanılmıştır. Verileri iyi bir ayırım ile yapabilen ve büyük farkları ortaya koyabilen kat değişimini göstermektedir (Kumar et al., 2018).

3.6.4. Quantil-Quantil Plot

Güçlü bir grafik yöntemi sağlayan çoklu kuantil grafikler, iki veya daha fazla popülasyonun dağılımını karşılaştırmak için kullanılmaktadır. Verilerin t dağılımlarını göstermek için Quantil-Quantil Plot grafiği kullanılmaktadır. Çalışma verilerini istatistiksel açıdan değerlendirirken bu grafik kullanılmıştır (Boon et al., 2013).

3.6.5. BoxPlot

Kutu grafiği, bir veri kümesinde gizlenmiş genel kalıpları ortaya çıkarmak için bir araçtır ve büyük veri kümesinin özelliklerini gösterebilmek için kullanışlı bir grafik türüdür. BoxPlot grafikleri tarafından verilerin merkezini ve dağılımını göstermek için kullanılan öğeler; Medyan, Çeyrekler Arası Aralık (IRQ) ve Whiskers' dir (Tareen et al., 2019).

4. BULGULAR

4.1. Analiz Sonuçları

Çalışmada GenBank Overview-NCBI-GEO kütüphanesinden *Homo sapiens* filtresi uygulanarak sitrülün ilişkili genler belirlendi (Şekil 4.1.1).

NIH National Library of Medicine
National Center for Biotechnology Information

Gene (citrulline) AND "Homo sapiens"[porgn: _txid9606] Search

Gene sources: Genomic, Categories: Alternatively spliced, Annotated genes, Protein-coding, Sequence content: CCDS, Ensembl, RefSeq, RefSeqGene, Status: Current, Clear all, Show additional filters

Search results: Items: 1 to 20 of 23. See also 11 discontinued or replaced items. Showing results for (citrulline) AND "Homo sapiens"[porgn: _txid9606]. Your search for (citrulline) AND "Homo sapiens"[porgn: _txid9606] retrieved no results.

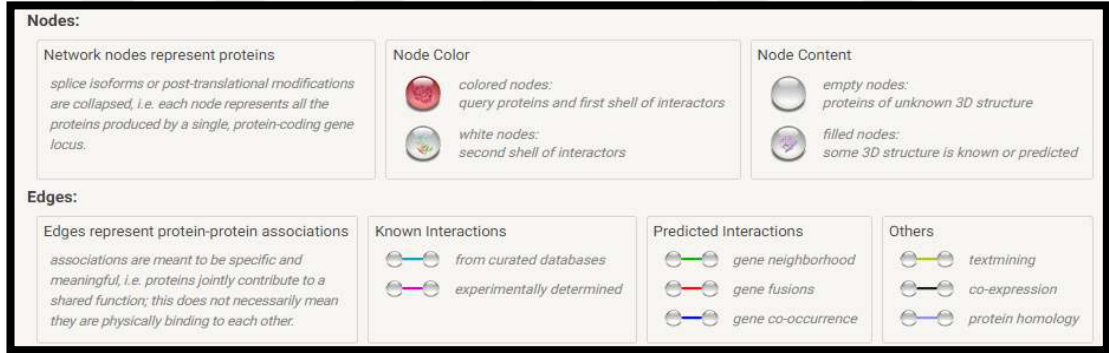
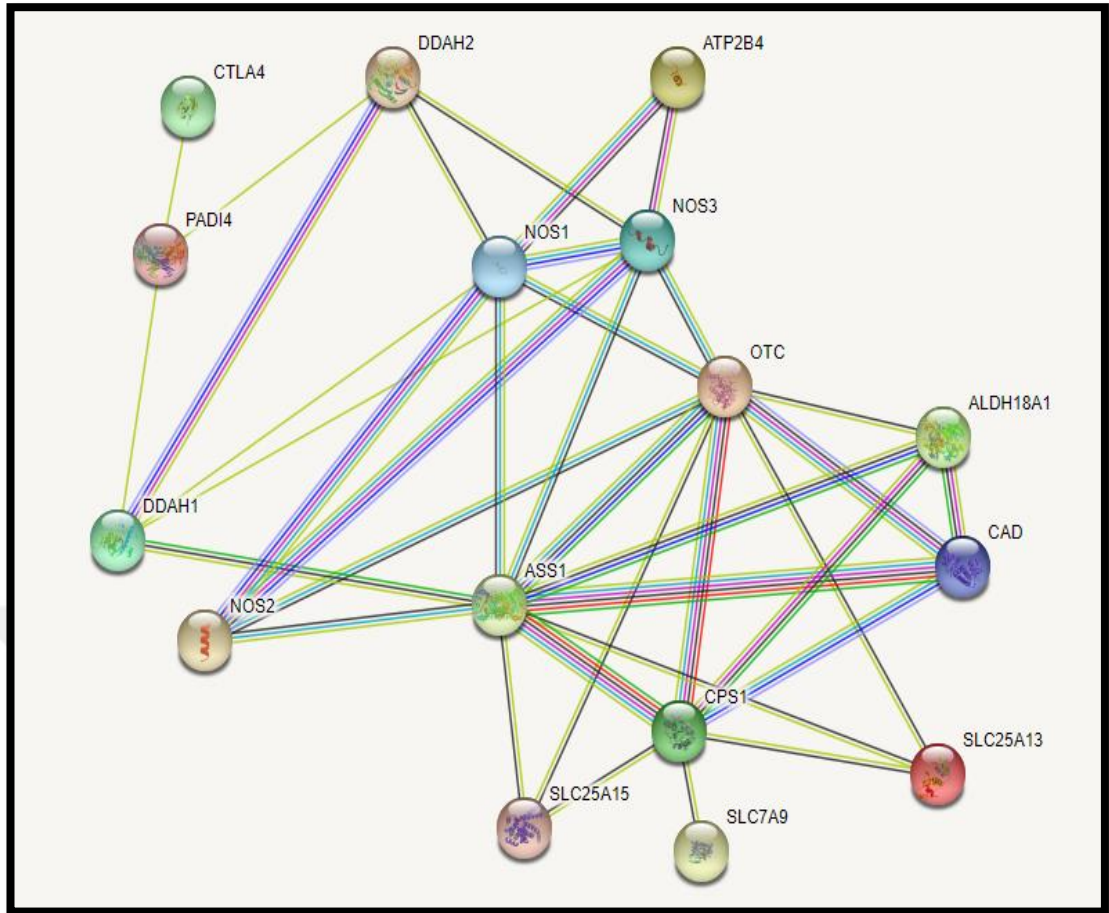
Name/Gene ID	Description	Location	Aliases	MIM
☐ ASS1 ID: 445	argininosuccinate synthase 1 [Homo sapiens (human)]	Chromosome 9, NC_000009.12 (130444707..130501274)	ASS, CTLN1	603470
☐ NOS3 ID: 4846	nitric oxide synthase 3 [Homo sapiens (human)]	Chromosome 7, NC_000007.14 (150991017..151014588)	ECNOS, eNOS	163729
☐ NOS2 ID: 4843	nitric oxide synthase 2 [Homo sapiens (human)]	Chromosome 17, NC_000017.11 (27756766..27800529, complement)	HEP-NOS, INOS, NOSA, NOS2	163730
☐ NOS1 ID: 4842	nitric oxide synthase 1 [Homo sapiens (human)]	Chromosome 12, NC_000012.12 (117208142..117361626, complement)	IHPS1, N-NOS, NC-NOS, bNOS, nNOS	163731
☐ CTLA4 ID: 1493	cytotoxic T-lymphocyte associated protein 4	Chromosome 2, NC_000002.12	ALPS5, CD	123890

Search details: citrulline[All Fields] AND "Homo sapiens"[porgn] AND alive[prop] Search See more...

Recent activity: Turn Off Clear
Q (citrulline) AND "Homo sapiens"[porgn] AND (alive[prop]) (23) Gene
Q citrulline AND (alive[prop]) (2205) Gene
Porphyrionomonas gingivalis SerB mutant infection effect on GDSBrowser

Şekil 4.1.1: Sitrülün İlişkili Genlerin Tespit Edilmesi İçin Overview-NCBI-NIH Kütüphanesinin Ekran Görüntüsü

Belirtilen CTLA4, PADI4, DDAH1, NOS2, DDAH2, NOS1, NOS3, ATP2B4, OTC, ASS1, SLC25A15, CPS1, SLC7A9, SLC25A13 CAD ve ALDH18A1 kodlu genlere ulaşılmıştır. Genler STRING clouster veri bankasında analiz edilmiştir (Şekil 4.1.2).



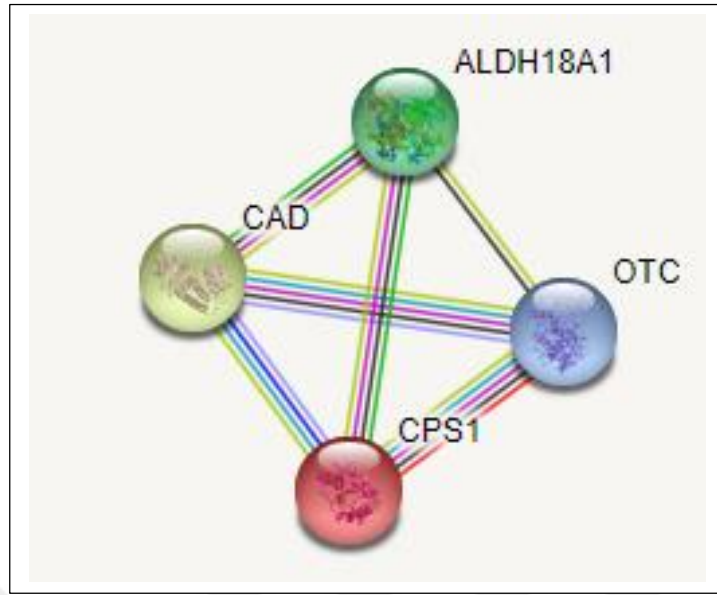
Şekil 4.1.2: Sitrulin İlişkili Genlerin STRING GENE Analizi ve Bağlantıları

STRING protein-protein interaction veri tabanından sitrulin ile ilgili insan (*homo sapiens*) gen ifadeleri ve tanımları alınmıştır. Alınan gen ifadelerinin açıklamaları Tablo 4.1.1’de gösterilmiştir.

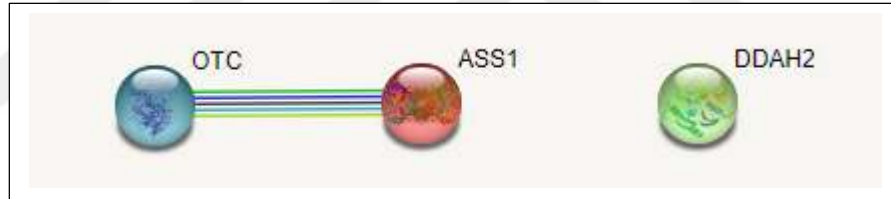
Tablo 4.1.1. Sitrülin İlişkili Genlerin (Homo Sapiens) Tanımları ve Görsel İfadeleri

● SLC25A13	Solute carrier family 25 (mitochondrial aspartate/glutamate transporter), member 12/13; Calcium-binding mitochondrial carrier protein Aralar2; Catalyzes the calcium-dependent exchange of cytoplasmic glutamate with mitochondrial aspartate across the mitochondrial inner membrane. May have a function in the urea cycle; EF-hand domain containing (676 aa)
● OTC	Ornithine carbamoyltransferase, mitochondrial; Ornithine carbamoyltransferase; Belongs to the ATCase/OTCase family (354 aa)
● ATP2B4	Plasma membrane calcium-transporting ATPase 4; Calcium/calmodulin-regulated and magnesium-dependent enzyme that catalyzes the hydrolysis of ATP coupled with the transport of calcium out of the cell. By regulating sperm cell calcium homeostasis, may play a role in sperm motility (By similarity); Belongs to the cation transport ATPase (P-type) (TC 3.A.3) family. Type IIB subfamily (1205 aa)
● ALDH18A1	Delta-1-pyrroline-5-carboxylate synthetase; Delta-1-pyrroline-5-carboxylate synthase; Bifunctional enzyme that converts glutamate to glutamate 5-semialdehyde, an intermediate in the biosynthesis of proline, ornithine and arginine (795 aa)
● CPS1	Carbamoyl-phosphate synthase [ammonia], mitochondrial; Involved in the urea cycle of ureotelic animals where the enzyme plays an important role in removing excess ammonia from the cell (1506 aa)
● DDAH1	N(G),N(G)-dimethylarginine dimethylaminohydrolase 1; Hydrolyzes N(G),N(G)-dimethyl-L-arginine (ADMA) and N(G)-monomethyl-L-arginine (MMA) which act as inhibitors of NOS. Has therefore a role in the regulation of nitric oxide generation (285 aa)
● NOS3	Nitric-oxide synthase, endothelial; Nitric oxide synthase, endothelial; Produces nitric oxide (NO) which is implicated in vascular smooth muscle relaxation through a cGMP-mediated signal transduction pathway. NO mediates vascular endothelial growth factor (VEGF)-induced angiogenesis in coronary vessels and promotes blood clotting through the activation of platelets (1203 aa)
● NOS1	Nitric oxide synthase, brain; Produces nitric oxide (NO) which is a messenger molecule with diverse functions throughout the body. In the brain and peripheral nervous system, NO displays many properties of a neurotransmitter. Probably has nitrosylase activity and mediates cysteine S-nitrosylation of cytoplasmic target proteins such SRR; Belongs to the NOS family (1468 aa)
● CAD	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase; CAD protein; This protein is a "fusion" protein encoding four enzymatic activities of the pyrimidine pathway (GATase, CPSase, ATCase and DHOase); In the central section; belongs to the metallo- dependent hydrolases superfamily. DHOase family. CAD subfamily (2225 aa)
● PADI2	Protein-arginine deiminase type-2; Catalyzes the deimination of arginine residues of proteins; Peptidyl arginine deiminases (665 aa)
● PADI4	Protein-arginine deiminase type-4; Catalyzes the citrullination/deimination of arginine residues of proteins such as histones, thereby playing a key role in histone code and regulation of stem cell maintenance. Citrullinates histone H1 at 'Arg-54' (to form H1R54ci), histone H3 at 'Arg-2', 'Arg-8', 'Arg-17' and/or 'Arg-26' (to form H3R2ci, H3R8ci, H3R17ci, H3R26ci, respectively) and histone H4 at 'Arg-3' (to form H4R3ci). Acts as a key regulator of stem cell maintenance by mediating citrullination of histone H1: citrullination of 'Arg- 54' of histone H1 (H1R54ci) results in H1 displacem [...] (663 aa)
● SLC25A15	Mitochondrial ornithine transporter 1; Ornithine transport across inner mitochondrial membrane, from the cytoplasm to the matrix; Solute carriers (301 aa)
● DDAH2	N(G),N(G)-dimethylarginine dimethylaminohydrolase 2; Hydrolyzes N(G),N(G)-dimethyl-L-arginine (ADMA) and N(G)-monomethyl-L-arginine (MMA) which act as inhibitors of NOS. Has therefore a role in the regulation of nitric oxide generation (285 aa)
● NOS2	Nitric-oxide synthase, inducible; Nitric oxide synthase, inducible; Produces nitric oxide (NO) which is a messenger molecule with diverse functions throughout the body. In macrophages, NO mediates tumoricidal and bactericidal actions. Also has nitrosylase activity and mediates cysteine S-nitrosylation of cytoplasmic target proteins such PTGS2/COX2 (By similarity). As component of the iNOS-S100A8/9 transnitrosylase complex involved in the selective inflammatory stimulus-dependent S-nitrosylation of GAPDH on 'Cys-247' implicated in regulation of the GAIT complex activity and probably mul [...] (1153 aa)
● SLC7A9	B(0,+)-type amino acid transporter 1; Involved in the high-affinity, sodium-independent transport of cystine and neutral and dibasic amino acids (system b(0,+)-like activity). Thought to be responsible for the high- affinity reabsorption of cystine in the kidney tubule; Belongs to the amino acid-polyamine-organocation (APC) superfamily (487 aa)
● ASS1	Argininosuccinate synthase; One of the enzymes of the urea cycle, the metabolic pathway transforming neurotoxic amonia produced by protein catabolism into innocuous urea in the liver of ureotelic animals. Catalyzes the formation of argininosuccinate from aspartate, citrulline and ATP and together with ASL it is responsible for the biosynthesis of arginine in most body tissues; Belongs to the argininosuccinate synthase family. Type 1 subfamily (412 aa)
● HIST2H3D	Histone cluster 2 H3 family member d; Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. DNA accessibility is regulated via a complex set of post-translational modifications of histones, also called histone code, and nucleosome remodeling (136 aa)
● SLC22A6	Solute carrier family 22 member 6; Involved in the renal elimination of endogenous and exogenous organic anions. Functions as organic anion exchanger when the uptake of one molecule of organic anion is coupled with an efflux of one molecule of endogenous dicarboxylic acid (glutarate, ketoglutarate, etc). Mediates the sodium-independent uptake of 2,3-dimercapto-1-propanesulfonic acid (DMPS) (By similarity). Mediates the sodium-independent uptake of p-aminohippurate (PAH), ochratoxin (OTA), acyclovir (ACV), 3'-azido- 3'-deoxythymidine (AZT), cimetidine (CMD), 2,4-dichloro- phenoxyacetat [...] (563 aa)
● CTLA4	Cytotoxic T-lymphocyte protein 4; Inhibitory receptor acting as a major negative regulator of T-cell responses. The affinity of CTLA4 for its natural B7 family ligands, CD80 and CD86, is considerably stronger than the affinity of their cognate stimulatory coreceptor CD28 (223 aa)

Sitrülinin biyosentez ve metabolik süreçleri gen düzeyinde incelendiğinde; CPS1, OTC, CAD, ALDH18A1 genleri biyosentez sürecinde etkili olduğu belirlenirken (Şekil 4.1.3); DDAH2 ve ASS1 genlerinin metabolik süreçlerde (Şekil 4.1.4) etkili olduğu STRING GENE analizi ile belirlenmiştir (Tablo 4.1.2).



Şekil 4.1.3 : Sitrülin *Biyosentez* Sürecinde Etkili Olan Genler Arasındaki İlişkinin Gösterilmesi



Şekil 4.1.4 : Sitrülin *Metabolik* Sürecinde Etkili Olan Genler Arasındaki İlişkinin Gösterilmesi

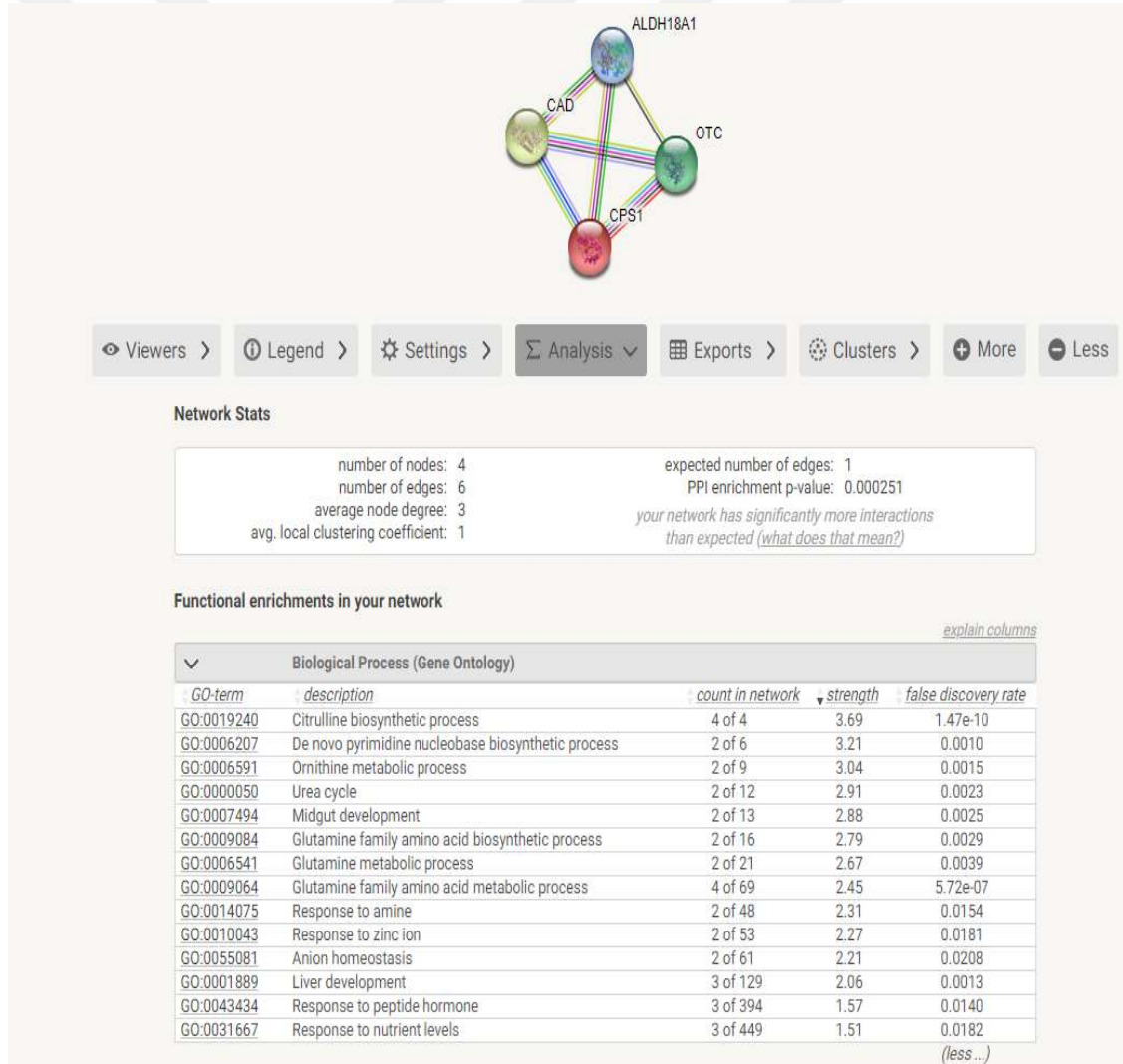
Veri bankasından gen setlerinin çekilmesinde GO_Central, BHF-UCL, Uni-Prot veri merkezleri değerlendirilmiştir. OTC gen verileri için GO_Central ve BHF-UCL; CPS1 için BHF-UCL; CAD için GO_Central; DDAH2 için Uni-Prot ve GO_Central; ALDH18A1 için Uni-Prot; ASS1 için BHF-UCL veri merkezleri kullanılmıştır.

Sitrülin ilişkili genlerin biyosentez ve metabolik olmak üzere görev aldıkları yollar değerlendirildiğinde; biyosentez yolağında en fazla görev alan gen CPS1 iken, metabolik yolakta en fazla görev alan genin ASS1 olduğu görüldü (Tablo 4.1.2).

Tablo 3.1.2. Sitrülin İlişkili Genlerin Görev Aldıkları Yolaklar

Gen Adı	İlişkili Olduğu Yolak Sayısı (n)	Yolak Türü
OTC	45	Sitrülin Biyosentez
CAD	48	Sitrülin Biyosentez
ALDH18A1	25	Sitrülin Biyosentez
CPS1	79	Sitrülin Biyosentez
DDAH2	27	Sitrülin Metabolik
ASS1	101	Sitrülin Metabolik

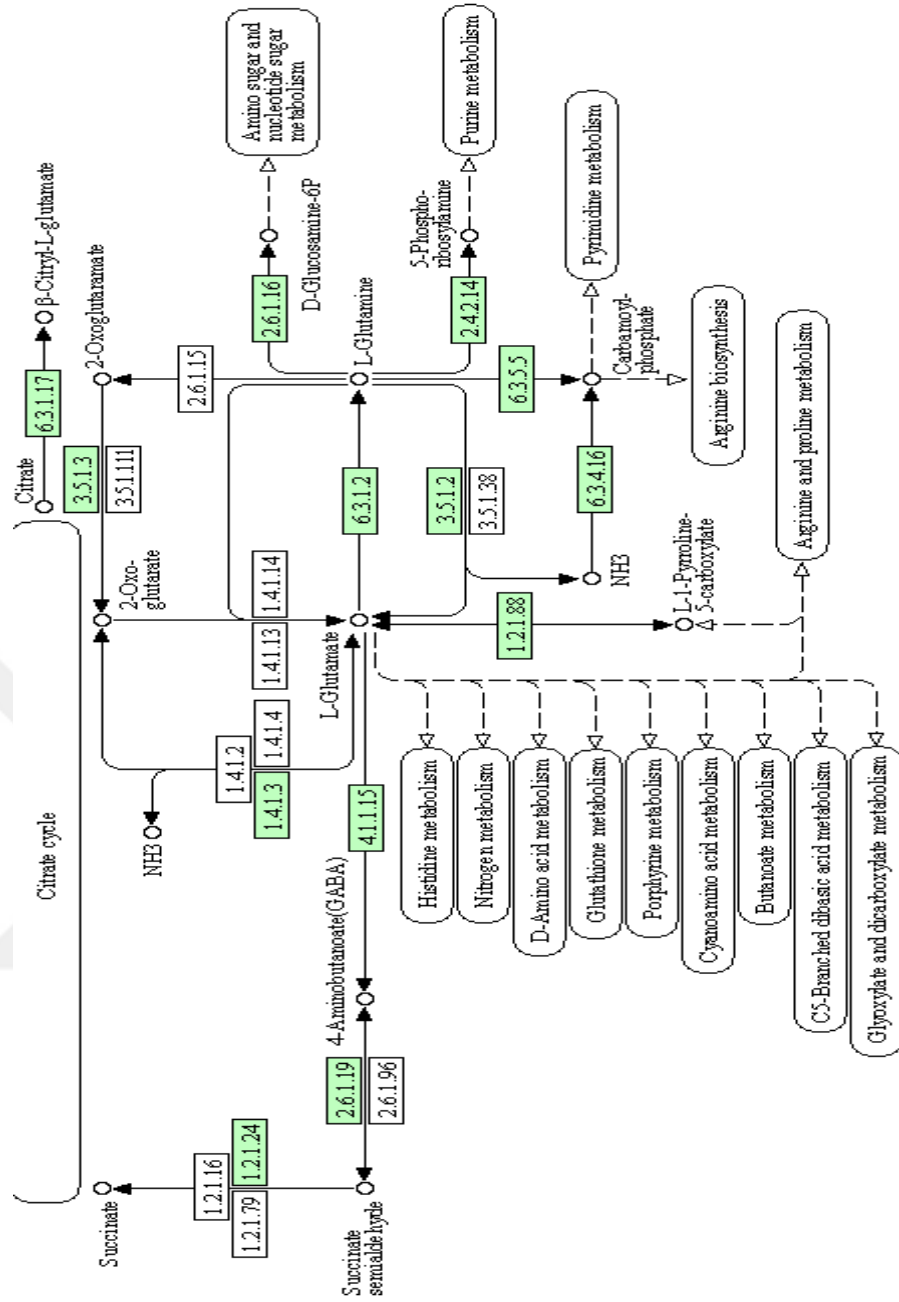
Sitrülin ile yakın ilişkili genlerin ontolojik analizi yapılmıştır. Analize göre yolak analizlerine geçilmiştir (Şekil 4.1.5).



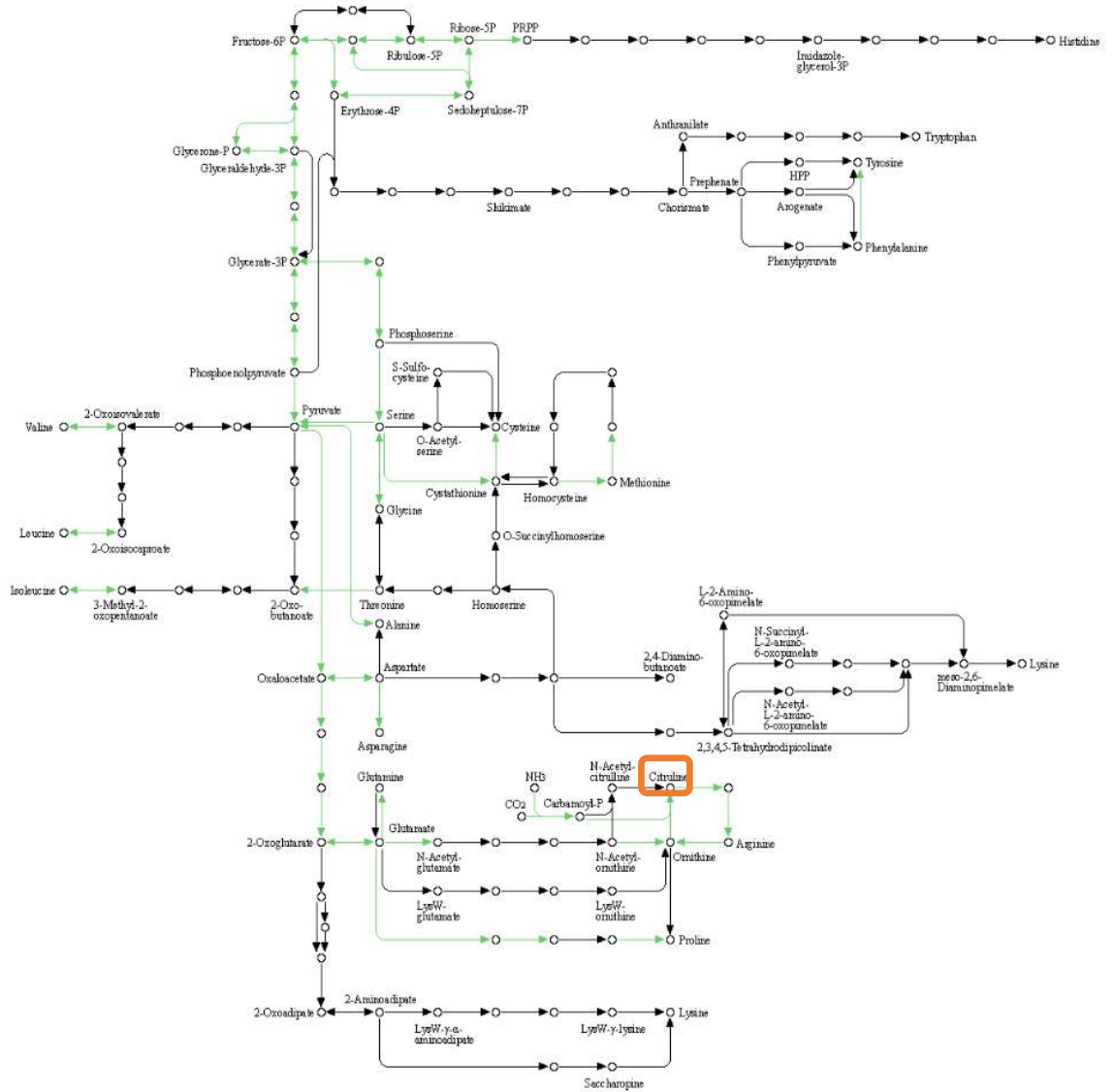
Şekil 4.1.5: Sitrülin Biyosentez Sürecinde Etkili Olan Genlerin Analizi

Sitr lin ile iliřkili genlerin g rev aldıęı yolaklar, yapay zeka araları olan Reactome, KEGG, STRING GENE ve Wiki analiz sistem ve yazılımları ile deęerlendirildi. Genlerin en etkili olduęu yolakların; arginin biyosentez, (KEGG pathways, řekil 4.1.6), alanin, aspartat ve glutamat metabolizması (KEGG pathways, řekil 4.1.7), amino asitlerin biyosentezi (KEGG pathways, řekil 4.1.8),  re d ng s  (Wiki Pathway, řekil 4.1.10) algoritması elde edildi. Wiki Pathway ile aynı zamanda  re d ng s  ile ilgili hastalıkların analizi (řekil 4.1.9) verilmiřtir.

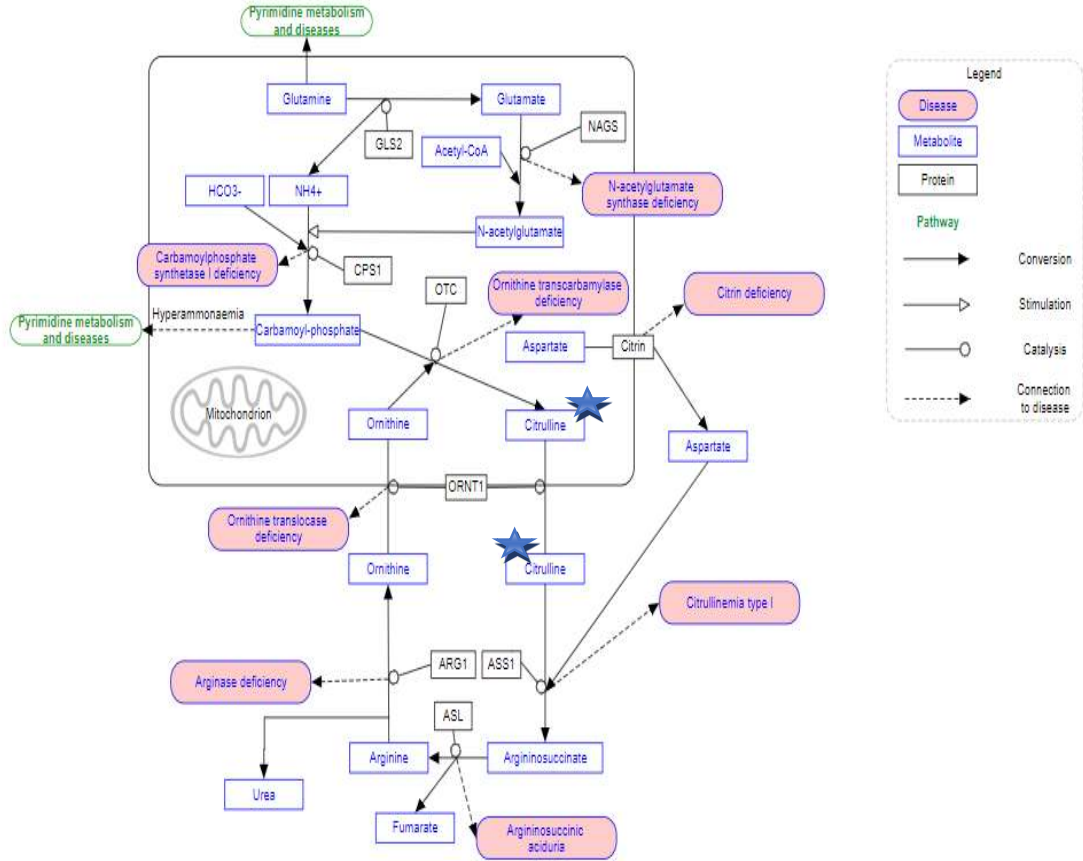




Şekil 4.1.7 : KEGG Pathway Analizi ile Alanin, Aspartat ve Glutamat Metabolizması

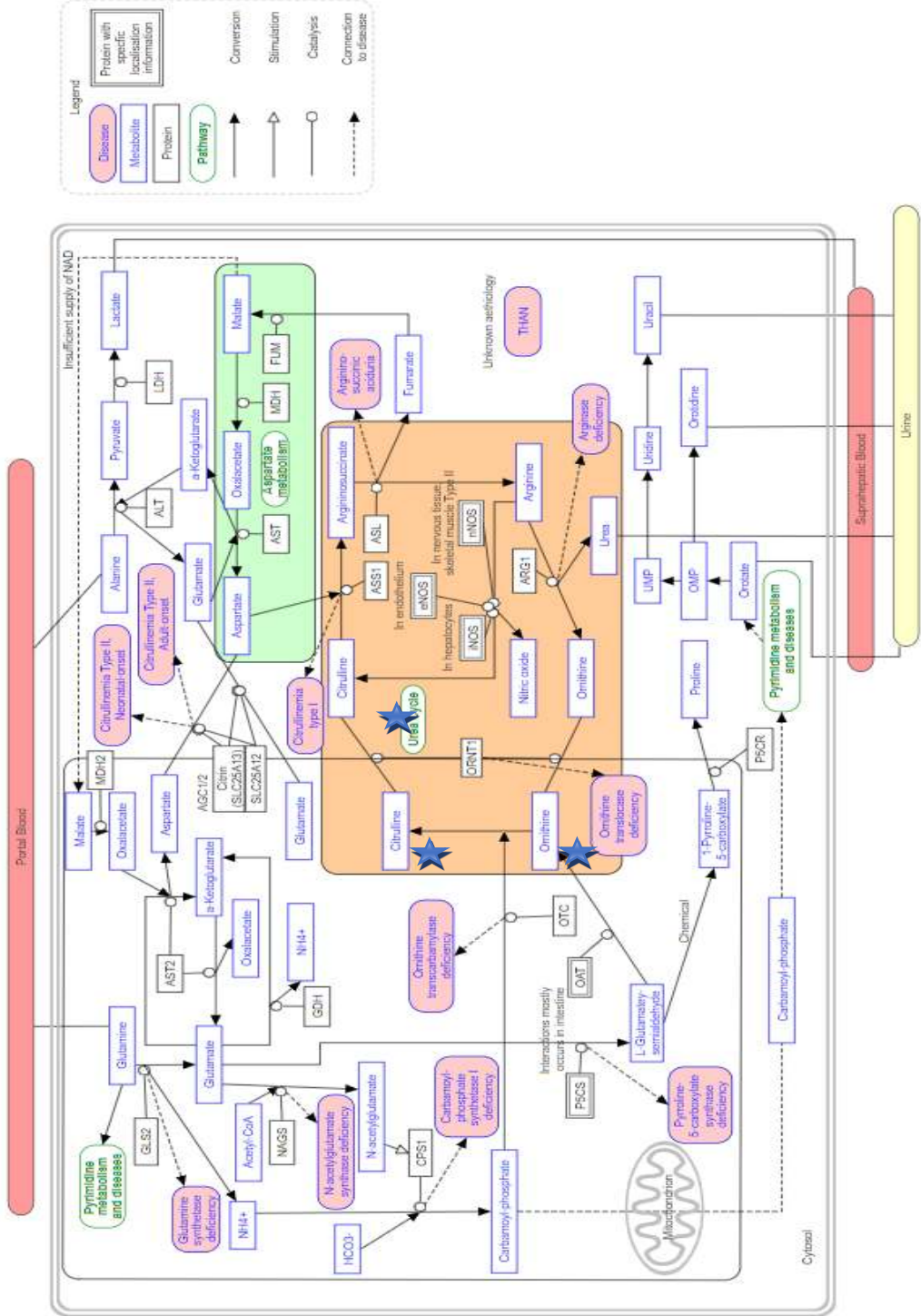


Şekil 4.1.8 : KEGG Pathway Analizi ile Amino Asitlerin Biyosentezi



Şekil 4.1.9 : Üre Döngüsü ve İlgili Hastalıklar (Wiki Pathway)

(https://search.crossref.org/?q=10.3180/REACT_847.1) (Erişim tarihi: 05.09.2022)



Şekil 4.1.10 : Üre Döngüsü ve İlişkili Yollar (Wiki Pathway)
https://search.crossref.org/?q=10.3180/REACT_847.1 (Erişim tarihi:
 05.09.2022)

- **İstatistiksel analiz**

Sitrülin ile ilişkili genler SLC25A13, NAGS, OTC, ASS1, ASL, CPS1, ARG1, SLC25A15, ARG2, GYPB 'dir. Genlerin z skor değerleri Tablo 4.1.3'de verilmiştir. Yapılan analizler sonrasında üre bozukluğunda görev alan OTC, ASS1 ve CPS1 genleri de belirtilmiştir. Sırasıyla bu genlerin z skoru değerleri 6,9 , 6,8 ve 6,1 olarak tespit edilmiştir (Tablo 4.1.3).

Tablo 4.1.3. Üre Döngüsü Bozukluğunda Sitrülin İlişkili Genlerin Z Skorları

İsim	z-skoru
SLC25A13	7,8
NAGS	7,2
OTC	6,9
ASS1	6,8
ASL	6,3
CPS1	6,1
ARG1	6,0
SLC25A15	5,8
ARG2	4,9

Sitrülin ile ilişkili genlerin dahil olduğu pathway isimleri gösterilmiştir. Reactome software programın da biyosentez ve metabolik süreçlerde aktif rol oynayan yolların istatistiksel analizi verilmiştir. Tablo 4.1.4' de sitrülin ile ilişkili en yakından en uzak ilişkiliye doğru yolak analizleri sıralanmıştır. Elde edilen sonuçlara göre en yakın yolak olan üre döngüsünde rol alan sitrülin ile ilişkili genlerin önemli farklılığı olduğu tespit edilmiştir ($p<0,001$) (Tablo 4.1.4).

Tablo 4.1.4. Sitrülin İlişkili Genlerin Dahil Olduğu Biyosentez ve Metabolik Yolların İstatistiksel Analizi (Reactome Software)

Yolak adı	Değer				Reaksiyon	
	oran	oran	p-value	FDR*	oran	oran
Üre döngüsü	3/10	8,61e-04	7,63e-08	1,60e-06	3/8	5,74e-04
Amino asitlerinin ve türevlerinin metabolizması	5/375	0,032	7,72e-06	7,72e-05	4/285	0,02
Metabolizma	8/2,143	0,185	4,23e-05	2,96e-04	10/2,268	0,163
Glutamat ve glutamin metabolizması	2/15	0,001	7,46e-05	3,73e-04	1/14	0,001
Oleoyl-phe metabolizması	1/1	8,61e-05	8,61e-04	0,003	2/2	1,44e-04
Pirimidin biyosentezi	1/3	2,58e-04	0,003	0,008	3/7	5,02e-04
Mitokondriyal ayrışma	1/6	5,17e-04	0,005	0,013	2/9	6,46e-04
eNOS aktivitesi	1/11	9,47e-04	0,009	0,013	1/20	0,001
Apoptoz kaynaklı DNA parçalanması	1/13	0,001	0,011	0,013	10/12	8,61e-04
Nükleotid biyosentezi	1/14	0,001	0,012	0,013	3/24	0,002
Nitrik oksit metabolizması: NOS3 aktivasyonu ve regülasyonu	1/15	0,001	0,013	0,013	1/26	0,002
Düz kas kasılması	1/44	0,004	0,037	0,037	4/16	0,001
Apoptotik yürütme aşaması	1/52	0,004	0,044	0,044	10/57	0,004
Mitokondriyal protein taşınması	1/65	0,006	0,055	0,055	4/14	0,001
Nükleotidlerin metabolizması	1/101	0,009	0,084	0,084	3/142	0,01
Solunum elektron taşınımı, kemiozmotik eşleşme ile ATP sentezi ve proteinleri ayırarak ısı üretimi	1/127	0,011	0,104	0,104	2/31	0,002
Protein lokalizasyonu	1/166	0,014	0,134	0,134	4/53	0,004
Sitrik asit döngüsü ve solunum elektron taşınması	1/178	0,015	0,143	0,143	2/67	0,005
Apoptoz	1/182	0,016	0,146	0,146	10/142	0,01
Kas kasılması	1/204	0,018	0,162	0,162	4/53	0,004
Programlanmış hücre ölümü	1/217	0,019	0,172	0,172	10/198	0,014

R programında en çok ilişkisi tespit edilen ASS1 geni GSE1563 data seti için okutuldu. Çıkan ekspresyon verileri 62 hasta üzerinde böbrek transplantasyon ve doku

hasarı bulunan klinik hasta datası üre döngüsü ilişkili 100 gen için değerlendirildi. ASS1 geni için Log Fc değeri 3,36 olarak hesaplandı ($p<0000,1$) (Tablo 4.1.5).

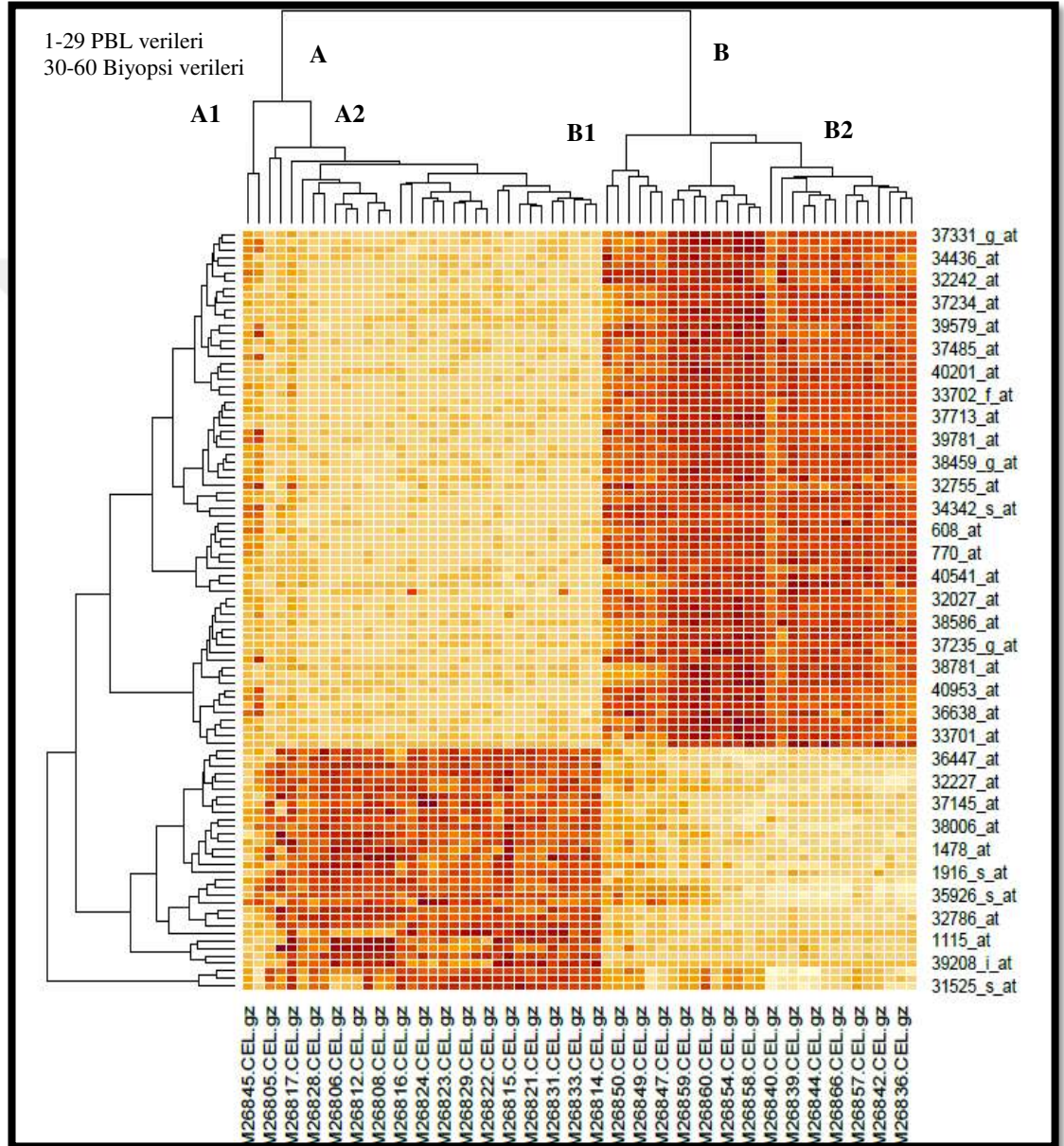
Tablo 4.1.5. Sitrülin İlişkili Genlerin Log Fc Değerleri

Gen Kodu	Log Fc Değeri	P.Value	Gen Adı
34342_s_at	6,15	$p<0000,1$	<u>SPP1</u>
2092_s_at	5,75	$p<0000,1$	<u>SPP1</u>
41377_f_at	5,53	$p<0000,1$	<u>UGT2B7</u>
33702_f_at	5,18	$p<0000,1$	<u>PCK1</u>
40541_at	3,36	$p<0000,1$	*ASS1
1916_s_at	-4,89	$p<0000,1$	<u>FOS</u>
41096_at	-4,66	$p<0000,1$	<u>S100A8</u>
31506_s_at	-4,58	$p<0000,1$	<u>DEFA1</u>
Ekspresyon Azalması		Ekspresyon Artması	

Kümeleme analizi (Cluster-Heatmap) bir ısı haritasıdır. Küme analizinde Log Fc değerlerine göre hangi gen aralığında önemli çıktı ise o gen aralığı analiz edilmiştir. Şekil 4.1.11’ de gösterilen gen kodunun çalıştırıldığı uygun makale olan böbrek transplantasyon ve doku hasarı bulunan klinik hasta datasının üre döngüsü ile ilişkili 100 genin küme analizini göstermektedir (Şekil 4.1.11).

Kümeleme analizinde böbrek transplantasyonu gerçekleştirilmesinde periferik kan lenfositleri (PBL) ve biyopsi gen profil verileri karşılaştırılmıştır. Hücre verilerini hiyerarşik küme yöntemi kullanılarak elde edilen dendrogram ile çözümleyerek iki ana küme elde edilmiştir. Bunlardan birinci ana küme PBL; ikinci ana küme biyopsi verileridir. Veri setinden 62 örnek çekilerek PBL ve biyopsi verileri karşılaştırıldığında: GSM26819, GSM26822, GSM26863, GSM26852, GSM26806, GSM26818, GSM26831, GSM26810, GSM26843, GSM26814, GSM26812, GSM26805, GSM26865, GSM26811, GSM26844, GSM26836, GSM26853, GSM26854, GSM26809, GSM26824, GSM26817, GSM26846, GSM26838 veri seti PBL ait sol clusterda gösterilmiştir. GSM26828, GSM26850, GSM26837, GSM26807, GSM26840, GSM26857, GSM26848, GSM26849, GSM26821, GSM26808, GSM26861, GSM26866, GSM26816, GSM26839, GSM26832, GSM26830, GSM26855, GSM26842, GSM26815, GSM26823, GSM26845, GSM26820, GSM26834, GSM26835, GSM26827, GSM26862, GSM26860, GSM26859, GSM26829, GSM26833, GSM26856, GSM26851, GSM26864,

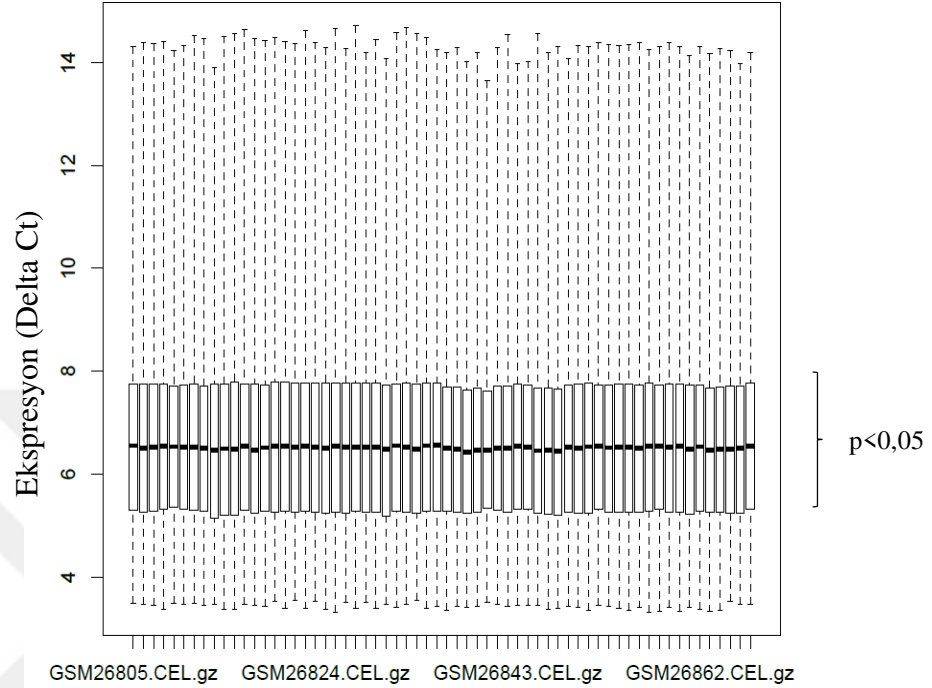
GSM26825, GSM26826, GSM26847, GSM26841, GSM26813, GSM26858 veriseti ise biyopsi ile ilişkili sağ clusterda yer almaktadır. Ana kümelerin kendi içinde oluşturduğu alt kümeler Şekil 4.1.11’de verilmiştir. Yaklaşık olarak 100 gen analizinde 33 genin ekspresyonu azalır iken 67 genin ekspresyonu artmıştır. Alt kümelerde grup içi ekspresyon değişimi 30 gen için A1 de artarken A2 de azalmıştır. Bu değişim B1 ve B2 alt kümelerinde görülmemiştir.



A: Ana küme B: Ana küme A1: Alt küme A2: Alt küme B1: Alt küme B2: Alt küme
CEL: hasta _ at: gen

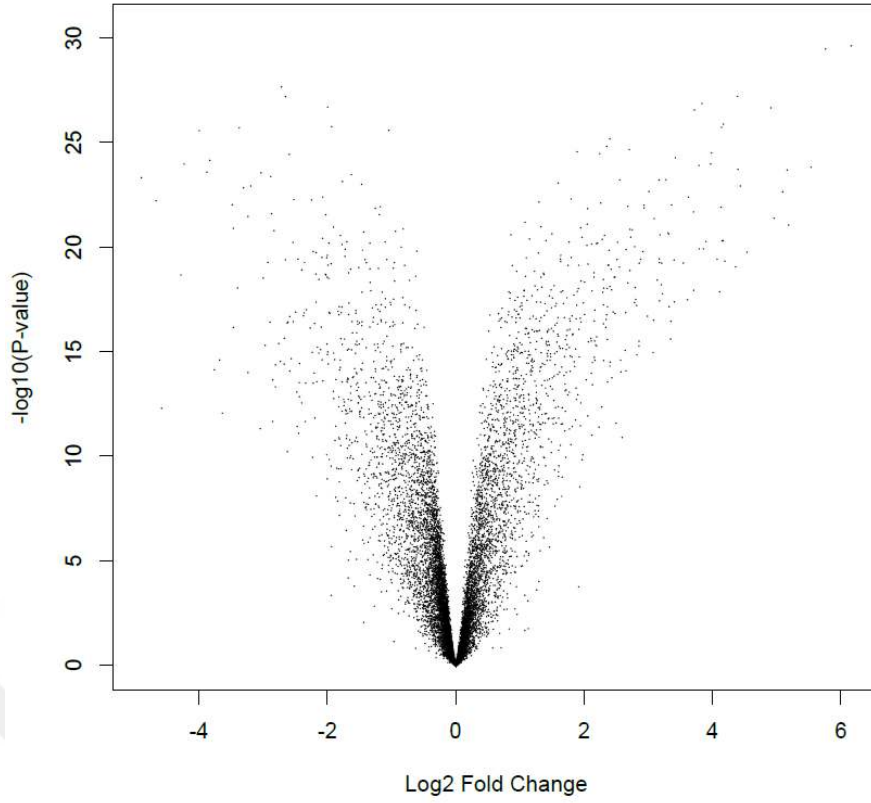
Şekil 4.1.11: Böbrek Transplantasyon ve Doku Hasarı Bulunan Klinik Hasta Datası Üre Döngüsü İlişkili 100 Genin Küme (Cluster) Analizi (Heat-Map Analizi)

Boxplot istatistik yöntemi ile sitrölin ilişkili genlerin ekspresyon seviyeleri yönünden değerlendirildiğinde yakın ilişkili (5,6-7,9) olduğu tespit edilmiştir ($p<0,05$) (Şekil 4.1.12).



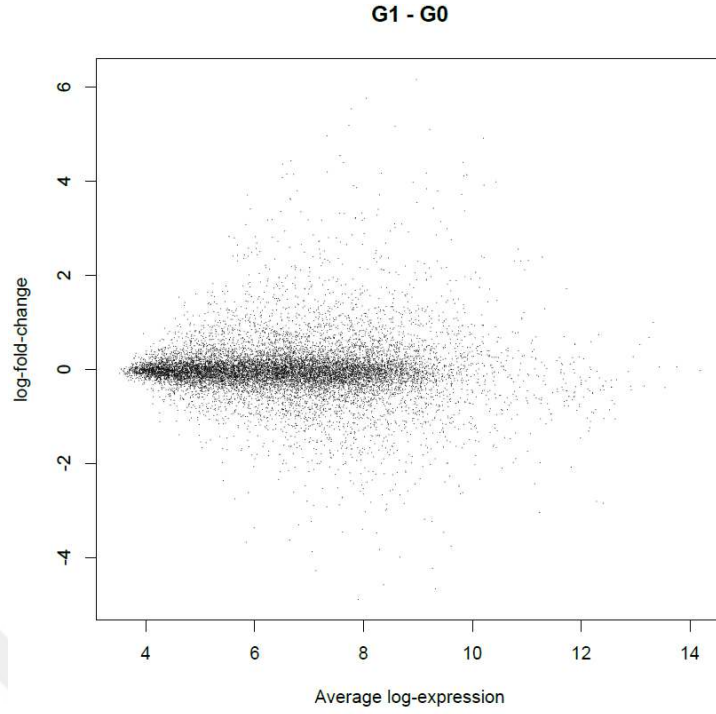
Şekil 4.1.12: Boxplot Analizi

Veri madenciliği yöntemi verilerde VolcanoPlot analizini gerçekleştirerek Şekil 4.1.13’ de gösterilmiştir. Şekil 4.1.13 LogFc değerlerine göre değişimlerinin istatistiksel anlamı olarak grafiksel dağılımını göstermektedir. Görüldüğü gibi logFc’deki kat farklarını VolcanoPlot ile genlerin ayrımı iyi bir şekilde yapılmıştır. LogFc değerlerinin yüksek çıktığı düzenlenmiş genler sağa doğru, düşük değerdeki düzenlenmiş genler sola doğru ve istatistiksel olarak önemli farklara sahip genler yukarı doğru dağıldığı bulunmuştur.



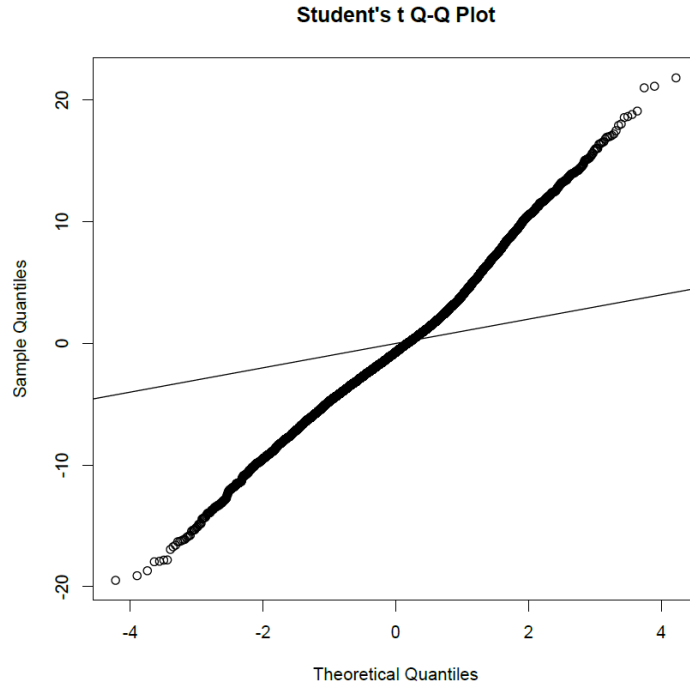
Şekil 4.1.13 : VolcanoPlot Analizi

PlotMD istatiksel analiz gerçekleştiren bir grafikdir. PlotMD grafiği veri madenciliği yönteminde ortalama yoğunluğu gösteren bir grafik olarak bilinmektedir. Verilerimizde iki farklı veri nesnesini hesaplayarak Şekil 4.1.14’ de göstermiştir. Grafikte LogFc değerlerinin ortalama renk yoğunluğu verilmiştir. PlotMD grafiği ve küme ekspresyon değerlerini ifade etmek için lcpm değerleri kullanılmıştır. Verilerin iyi bir ayrımı yapılmış ve büyük farkları ortaya koyulmuştur.



Şekil 4.1.14 : PlotMD Analizi

Şekil 4.1.15’ de gösterilen grafik iki veri grubunun dağılımını karşılaştırmaktadır. İki grup verinin dağılımını karşılaştıran grafik düz çizgi üzerinde dağıldığında, iki grubun aynı dağılışa sahip olduğunu göstermektedir.



Şekil 4.1.15 : QQ-Plot Analizi

5.TARTIŞMA

5.1. Tartışma

Günümüzde özellikle sağlık alanında yapay zeka giderek büyük önem kazanmaya başlamıştır. Başta yapay zeka olmak üzere bilgisayar bilimi, makine öğrenmesi, veri görselleştirme, veri tabanı yönetimi ve algoritmaları, istatistiği birleştiren veri madenciliği için disiplinler arası bir alanı oluşturmaktadır.

Yapay zeka ve makine öğrenmesinde kullanılan yöntemler istatistik bilim dalında geliştirilen yöntemlerdir. Nitekim bu yöntemler problem formülasyonu, hipotez geliştirme, veri analizi, kanıta dayalı yöntem seçimi, yorumlama, karar verme ve sonuçların uygulanmasında çözümler sağlar (Celik, 2022). Büyük veri (Big data) veri madenciliğinin bilgi deposunu oluşturur. Bu büyük veriyi indirgeyen, yorum çıkaran ve sonuçlandıran veri madenciliği dahil diğer istatistik yöntemlerdir. Günümüzde büyük sağlık problemleri sözü edilen bu yöntemlerle çözüme kavuşmuştur (Celik, 2022).

Sandeep ve arkadaşları, veri madenciliğinin diyabet, boğaz kanseri, meme kanseri, kalp hastalıkları, böbrek hastalıkları ve karaciğer hastalıklarının teşhisinde önemli bir yöntem olduğunu rapor etmişlerdir (Kautish and Ahmed, 2016).

Özellikle sağlık alanında son yıllarda veri madenciliği daha çok popüler olmaya devam etmektedir. Hastanelerde büyük, kompleks ve çok boyutlu verilerin artması yeni problemlerin gelişmesine neden olmaktadır. Bu verilerin çözümü için veri madenciliği yeni modellerin gelişmesini sağlamak durumunda kalmaktadır (Jagadish et al., 2014). Görüldüğü gibi birçok faktör sağlık alanında veri madenciliğinin uygulamasını motive etmektedir. Veri madenciliği ile geliştirilen yöntemler sağlık profesyonellerine karar verme sürecine yeni perspektifler kazandırmaktadır (Firouzi et al., 2018).

Veri madenciliği yoğun manuel ve hesaplamalı görevleri içeren karmaşık bir süreçtir. Bir sağlık kuruluşunun asıl mesleği, madenciliğe uygun veri toplamak değil, bireyselleştirilmiş hasta bakımı sağlamaktır. Bu nedenle birikmiş klinik verilerden ilginç ve geçerli bilgiler bulmak amacıyla klinik veri madenciliğini düzene sokmak için birçok zorluk vardır (Iavindrasana et al., 2009).

Amin ve arkadaşlarına göre hastalık tahmini, klinik veri analizi bölümünde en önemli konulardan biri olarak kabul edilmektedir. Sağlık sektöründe çok büyük veri

miktarı bulunmaktadır. Veri madenciliği, geniş ham sağlık verileri koleksiyonunu, bilinçli karar ve tahmin vermeye yardımcı olabilecek bilgilere dönüştürür. Bazı hastalıkların tahmininde veri madenciliği tekniklerini uygulayan bazı mevcut çalışmalar vardır. Tahmin modellerinin performansını artırabilecek önemli özelliklerin doğru kombinasyonunu seçmek çok önemlidir (Amin et al., 2019).

Açık erişimle ulaşılan NCBI- GEO kütüphanesinden sitrülün ilişkili gen bilgisini içeren tüm insan verilerini inceleyen bu araştırmanın sonuçları, daha önce yapılmış olan araştırma sonuçları ile tartışmalı olarak verilmiştir.

String veri tabanı farklı genlerin birbirleriyle olan ilişkilerini analiz etme yöntemi olarak kullanıldığından, genlerin “*STRING CONSORTIUM*” veri tabanı ile gen-protein ilişkisi incelenmiştir. Buna göre belirtilen CTLA4, PADI4, DDAH1, NOS2, DDAH2, NOS1, NOS3, ATP2B4, OTC, ASS1, SLC25A15, CPS1, SLC7A9, SLC25A13 CAD ve ALDH18A1 kodlu genlere ulaşılmıştır.

Bu çalışmada CPS1 geninin biyosentez sürecinde en fazla görev aldığı belirlenmiştir. CPS1’in sağlık ve hastalıktaki rolünü inceleyen bir çalışmada, enzimlerin veya taşıyıcıların herhangi birinin işlevlerinin bozulmasının, yüksek amonyak ve nörolojik hasara yol açtığını buna göre CPS1 geninin, amonyanın üre döngüsü ara ürünlerine doğrudan dahil edilmesinden sorumlu ilk ve hız sınırlayıcı görev alan enzim olduğunu saptamıştır. CPS1’in temel ve translasyonel araştırmalarındaki son gelişmelerle birlikte, bu enzimin üre döngüsündeki temel rolünün takdir edilmesi gerekliliği vurgulanmıştır (Nitzahn and Lipshutz, 2020).

Bu çalışmada OTC geninin biyosentez sürecinde görev aldığı belirlenmiştir. Üre döngüsünde proteinler üzerindeki fonksiyonel çalışmayı kolaylaştırmak için gerçekleştirilen bir çalışmada, üre döngüsü bozukluklarının hiperamonyemi ve geniş bir fenotipik şiddet yelpazesi ile karakterize edilen bir grup nadir metabolik durum olarak değerlendirilmesi doğrultusunda, beş üre döngüsü bozukluğu olan vaka incelenmiştir. Yüksek verimli tekniklerin bir kombinasyonu ile klinik, biyokimyasal ve genetik analize tabi tutulmuştur. Ayrıca, tespit edilen yanlış anlamlı genetik varyantlar üzerinde in silico analizi çalışılmıştır. Üre döngüsü bozukluğu olan çeşitli klinik ve biyokimyasal endikasyonlar, beş proband ile kendini göstermiş ve sekans analizi sonucu, bir vakada argininosüksinik asidüriye, iki vakada CPS1, bir vakada Ornitin transkarbamilaz eksikliğine (OTCD) ve bir vakada sitrin eksikliğine neden olan üçü yeni olmak üzere dokuz tanısal varyant saptanmıştır (Liu et al., 2021).

Bu arařtırmada CAD geninin biyosentez sürecinde görev aldığı belirlenmiřtir. CAD geni aktivitesindeki kusurların neden olduđu yeni ve nadir bir nörometabolik bozukluđuın son bulgusunu özetleyen bir alıřmada, her biri pirimidin nükleotitlerinin novo biyosentezi için bařlangı reaksiyonlarından birini katalize eden 250 kDa'lık bir proteinin heksamerik birleřmesiyle oluřturulan ve 1.5 MDa partikülü olan CAD geni hakkındaki bilgilerin kıt ve paralı olduđu belirtilmiřtir. CAD'i anlamannın, yalnızca üç boyutlu yapıları belirlemeyi ve farklı enzimatik alanların katalitik ve düzenleyici mekanizmalarını tanımlamayı değıl, aynı zamanda bu alanların nasıl birbirine karıřtıđını ve koordineli, düzenlenmiř bir řekilde alıřtıđını kavramanın önemi vurgulanmıřtır (del Caño-Ochoa and Ramón-Maiques, 2021).

Bu arařtırmada ALDH18A1 geninin biyosentez sürecinde görev aldığı belirlenmiřtir. Bifonksiyonel homooligomerik enzim Δ 1-pirolin-5-karboksilat sentetaz (P5CS) ve onu kodlayan ALDH18A1 geni, 1998'de hastalıkla iliřkilendirildiğı belirtilmiřtir. alıřmada nörokutanöz sendromla rapor edilen 32 hastadan 21'inde ailevi olanlar homozigot veya bileřik heterozigot ALDH18A1 gen mutasyonlarına ev sahipliğı yaparken, 11 sporadik de novo heterozigot ALDH18A1 gen mutasyonları tařıdığı saptanmıřtır. Klinik verilere (önceki hasta raporları dahil), ALDH18A1 mutasyon spektrumlarına ve P5CS proteini hakkındaki bilgimize dayanarak, dört sendromun azalmıř P5CS fonksiyonuna dayalı olarak aynı patojenik mekanizmaları paylařtığı sonucuna varılmıřtır (Marco-Marin et al., 2020).

Bu arařtırmada dimetilarginin dimetilaminohidrolaz 2 (DDAH2) geninin metabolik süreçte görev aldığı belirlenmiřtir. DDAH2 ile ilgili yapılan alıřmada, minyatür domuzun fetal akciğerine karřı fare monoklonal antikorlar geliřtirilmiřtir ve akciğer adenokarsinomuna odaklanarak malignitenin erken evresinde üretilen bir embriyonal protein bulmak için taranmıřtır. Akciğer adenokarsinomunun stromasını spesifik olarak boyayan bir antikor klonu bulunmuřtur. LC-MS/MS, bu klon tarafından tanınan proteini, antiaterosklerotik aktivite ile bilinen bir enzim olan DDAH2 olarak tanımlanmıřtır. In vitro analiz, DDAH2'nin endotelial nitrik oksit sentazın (eNOS) ekspresyonunu arttırdığı, vasküler endotel hücrelerinin proliferasyonunu ve kapiler benzeri tüp oluřumunu indüklediğı gösterilmiřtir. Verilerin, DDAH2 ekspresyonunun, tümör anjiyogenezi yoluyla akciğer adenokarsinomunun invazivliğı ile iliřkili olduđu gösterilmiřtir. DDAH2 ekspresyonu, akciğer adenokarsinomunda prognostik bir faktör olabileceğı belirtilmiřtir (Shiozawa et al., 2016).

Bu arařtırmada ASS1 geninin metabolik süreçte en fazla görev aldığı belirlenmiştir. Japonya’da yapılan bir çalışmada, üre döngüsü bozukluğu olan hastaların başlangıç zamanını, gen varyantlarını, klinik belirtilerini ve tedavisini araştırılmıştır. Ocak 2000 ile Mart 2018 arasında üre döngüsü bozukluğu tanısı alan ve/veya tedavi edilen 229 hastadan, 102 hastanın tanımlanmış gen varyantları ve klinik bilgileri alınmıştır. Buna göre 62 hastanın OTC, 18 CPS1, 16 hastanın ASS1, altı hastanın ise ASL eksikliği olduğu tespit edilmiştir. Üre döngüsü bozukluğu olan hastalarda başlangıç zamanı ve şiddeti, ülke çapındaki bu çalışma ve önceki çalışmalardan her hastada tanımlanan gen varyantlarına dayanarak tahmin edilebilmektedir. Bu genetik bilgi, üre döngüsü bozukluğu olan hastalarda uzun vadeli sonucu tahmin etmede ve karaciğer transplantasyonu gibi spesifik tedavi stratejilerini belirlemede yardımcı olabileceği belirtilmiştir (Kido et al., 2021).

Bu arařtırmadan elde edilen sonuçlar literatür bilgileri ile paralellik göstermektedir. Genlerin en etkili olduğu yollar; arginin biyosentez, alanin, aspartat ve glutamat metabolizması, amino asitlerin biyosentezi, üre döngüsü algoritması olarak tespit edilmiştir. Wiki Pathway ile aynı zamanda üre döngüsü ile ilgili hastalıkların analizi verilmiştir. Elde edilen sonuçlara göre en yakın yolak olan üre döngüsünde rol alan sitrölin ile ilişkili genlerin önemli farklılığı olduğu belirlenmiştir ($p<0,001$).

Çalışma için metabolik yolda en fazla ilişkili olduğu belirlenen üre döngüsü değerlendirildiğinde, klinik tablo örneği olarak seçilen böbrek transplantasyon ve doku hasarı bulunan hasta datası üzerinden sitrölin ilişkili 100 gen için kümeleme analizi (cluster) yapılmıştır. Yapılan kümeleme analizine göre seçilen sitrölin ilişkili genlerin fonksiyonel olduğu belirlenmiştir.

Çalışma ile sitrölinin metabolik yolları ile ilgili olası bazı mekanizmalar nitrit oksit, laktik asit, biyoenerji, krebs döngüsü, üre döngüsü ve arginin biyosentez yolağındaki ve bu yollar üzerinden metabolizmadaki önemi ve yeri anlaşılmaktadır. Yapılan analizler sonucunda sitrölin ile ilişkili genlerin ekspresyonunun üre döngüsü hastalıkları ile ilişkili olduğu gösterilmiştir.

Fletcher ve arkadaşlarının yapmış olduğu çalışmada, transplant hastalarında böbrek biyopsileri ve PBL’lerin gen ekspresyon profillerini belirlemek için DNA mikrodizileri (HG-U95Av2 GeneChips, Affymetrix) kullanılmıştır. Çalışmaya dahil edilme kriterleri; normal donör böbrekleri, reddedilmeyen iyi işleyen nakiller, akut reddedilen böbrekler ve reddedilmeyen böbrek fonksiyon bozukluğu olan nakillerini

kapsamaktadır. Hem biyopsiler hem de PBL'ler için farklı transplant hasta sınıflarının her biri ile önemli ölçüde ilişkili olan farklı gen ekspresyon imzaları belirlenmiştir. Bu sonuçların önemi, eğer çok merkezli ileriye dönük bir çalışmada doğrulanması halinde, nakledilen hastaların bağışıklık durumunu izlemek için gen ifadesi imzalarına dayalı bir ölçünün oluşturulması hedeflemektedir (Flechner et al., 2004).

5.2. Sonuç ve Öneriler

Başta yapay zeka olmak üzere bilgisayar bilimi, makine öğrenmesi, veri görselleştirme ve veri tabanı yönetimi ile matematik algoritmalarını, istatistiği birleştiren veri madenciliği disiplinler arası bir alanı oluşturmaktadır.

Büyük hacimli verileri verimli bir şekilde analiz etmek ve bilgiyi keşfetmek için veri madenciliğini geliştirmeye ihtiyaç vardır. Dünyada gerçekleştirilen teknolojik gelişmelere paralel olarak ülkemizde de veri madenciliği uygulamaları her geçen gün yaygınlaşmaktadır. Veri madenciliği istatistik biliminden köken almaktadır. "istatistik" veya "istatistiksel teknik" terimleri, veri madenciliği için kullanılmaktadır. İstatistik uygulamaları veriye dayalı olduğu için kalıpları keşfetmek ve tahmine dayalı istatistiklerde (regresyon olarak da adlandırılan) modeller oluşturmak için kullanılır. Başka bir ifadeyle, veri madenciliği istatistikle doğal bir bağlantıdadır. Veri madenciliği için istatistiksel analizler; bayes ağı, korelasyon analizi, faktör analizi, diskriminant analizi, küme analizi, regresyon analizi vb. gibi yaygın olarak kullanılmaktadır. Veri madenciliği görevlerinin çoğu, bir veya daha fazla istatistiksel yaklaşım ile gerçekleştirilir.

Çok sayıda alanda kullanılabilen veri madenciliği yöntemi sağlık alanı uygulamaları için de önem kazanmıştır. Bu çalışma, sitrülün ile ilişkili muhtemel metabolik yolların aydınlatılması için istatistik temele dayalı veri madenciliği yöntemi uygulaması ile birlikte verilmiştir. Sağlık alanında gerçekleştirilen uygulamaların birçoğu hastalık belirtilerini ve var olan belirtilerden bir şablon ortaya çıkarma amacıyla yapıldığı görülmüştür. Ülkemizde veri madenciliği çalışmalarının tıp alanında çok daha çeşitli ve etkili kullanılması gerektiği, bu konuda eksikliğin olduğu tespit edilmiştir.

ARG metaboliti olan sitrülün, GLN metabolizmasının son ürünü olan bir amino asittir. Sitrülinin ARG'e göre depo oranı daha fazla olması dolayısıyla biyoyararlanımının da daha yüksek oluşu ilişkili oldukları metabolik ve biyosentez yollarında dikkatleri üzerine çekmektedir. Bu çalışmada sitrülünün metabolik ve

biyosentez yolları ile ilgili muhtemel bazı mekanizmalar açıklanmıştır. Sitrülinin metabolizma ve biyosentezdeki önemi ve yeri araştırılmıştır. Son derece karmaşık olan metabolik ilişkileri açıklayabilmek için GenBank Overview-NCBI-GEO veri tabanı kullanılarak sitrülin ile ilişkili detaylı gen verilerine ulaşılmıştır ve veri madenciliği modelleri geliştirilmiştir. Çalışma için GSE1563 veri seti kullanılmıştır. Metabolik yolda en fazla ilişkili olduğu belirlenen üre döngüsü değerlendirildiğinde klinik tablo örneği olarak seçilen böbrek transplantasyon ve doku hasarı bulunan hasta verisi üzerinden sitrülin ilişkili 100 gen için kümeleme analizi (cluster) yapılmıştır. Kümeleme algoritması ile R programlama dili kullanılarak arayüz geliştirilmiştir.

Veri madenciliğinin kullanıldığı alanların çeşitlendirilmesi, gerek ülkemiz sağlık alanları gerekse bu sağlık alanlarının ilişkili olduğu kurum ve kuruluşlardan hizmet alan ülkemiz insanlarına büyük fayda sağlayacağı kanısındayız. Yine bu tür metabolik ve biyosentez yollarının aydınlatılması da veri madenciliği dolayısıyla da istatistik gibi temel bilim dallarından faydalanılması hastalıkların kaynağının ve seyrinin belirlenmesinde önemli bir nitelik taşımaktadır.

6. KAYNAKLAR

- Akashi, K., Miyake, C., & Yokota, A. (2001). Citrulline, a novel compatible solute in drought-tolerant wild watermelon leaves, is an efficient hydroxyl radical scavenger. *Febs Letters*, 508(3), 438-442.
- Akpınar, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29(1), 1-22.
- Albayrak, A. S., & Yılmaz, Ş. K. (2009). Data mining: decision tree algorithms and an application on ISE data. *Suleyman Demirel University the Journal of Faculty of Economics and Administrative Sciences*, 14(1), 31-52.
- Allerton, T. D., Proctor, D. N., Stephens, J. M., Dugas, T. R., Spielmann, G., & Irving, B. A. (2018). l-Citrulline supplementation: impact on cardiometabolic health. *Nutrients*, 10(7), 921.
- Alpaydın, E. (2000). Zeki veri madenciliği: Ham veriden altın bilgiye ulaşma yöntemleri. *Bilişim 2000 eğitim semineri*, 1-10.
- Altınışik, M. <https://www.mustafaaltinisik.org.uk/89-1-19.pdf>, Protein ve Aminoasit
- Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, 82-93.
- Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, 82-93.
- Anyanwu, MN., & Shiva, SG. (2009). Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer cience and Security*. Jun;3(3):230-40.
- Aşık, F. (2019). Böbrek nakli yapılan hastalarda veri madenciliği yöntemleri ile akut rejeksiyon durumunun incelenmesi. *Yüksek Lisans Tezi*.
- Atasoy, N. A., & Demiröz, A. (2021). Makine Öğrenmesi Algoritmaları Kullanılarak Prostat Kanseri Tümör Oluşumunun İncelenmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (29), 87-92.
- Batal, M. S. (2016). Yapay Zeka Uygulamaları ve Yapay Zekanın Geleceği. *Uluslararası Sunhill Üniversitesi*, 1-9.
- Batal, M. S. (2016). Yapay Zeka Uygulamaları ve Yapay Zekanın Geleceği. *Uluslararası Sunhill Üniversitesi*, 1-9.
- Bharati, M., & Ramageri, M. (2010). Data mining techniques and applications.
- Boon, M. A., Einmahl, J. H., & McKeague, I. W. (2013). Visualizing multiple quantile plots. *Journal of Computational and Graphical Statistics*, 22(1), 69-78.

Brosnan, M. E., & Brosnan, J. T. (2004). Renal arginine metabolism. *The Journal of nutrition*, 134(10), 2791S-2795S.

Celik, Y. (2022). Health Informatics Artificial Intelligence Machine Learning and R Programming. *Lord Matbaacılık ve Kağıtçılık*, ISBN 978-605-62409-4-2.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 1-27.

Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with applications*, 34(1), 280-290.

Cifci, M. A., & Hussain, S. (2018). Data mining usage and applications in health services. *JOIV: International Journal on Informatics Visualization*, 2(4), 225-231.

Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (2012). *Data mining methods for knowledge discovery* (Vol. 458). Springer Science & Business Media.

Crenn, P., Coudray-Lucas, C., Thuillier, F., Cynober, L., & Messing, B. (2000). Postabsorptive plasma citrulline concentration is a marker of absorptive enterocyte mass and intestinal failure in humans. *Gastroenterology*, 119(6), 1496-1505.

Curis, E., Nicolis, I., Moinard, C., Osowska, S., Zerrouk, N., Bénazeth, S., & Cynober, L. (2005). Almost all about citrulline in mammals. *Amino acids*, 29(3), 177-205.

Cynober, L., Le Boucher, J., & Vasson, M. P. (1995). Arginine metabolism in mammals. *The Journal of Nutritional Biochemistry*, 6(8), 402-413.

Cynober, L., Moinard, C., & De Bandt, J. P. (2010). The 2009 ESPEN Sir David Cuthbertson. Citrulline: a new major signaling molecule or just another player in the pharmaconutrition game?. *Clinical Nutrition*, 29(5), 545-551.

Çavuşoğlu, Ü., & Kaçar, S. (2019). Anormal trafik tespiti için veri madenciliği algoritmalarının performans analizi. *Academic Platform-Journal of Engineering and Science*, 7(2), 205-216.

Dabas, P., & Singh, B. (2021, September). Analysis of Data Mining Classification Techniques. In *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-3). IEEE

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. John Wiley & Sons.

del Caño-Ochoa, F., & Ramón-Maiques, S. (2021). Deciphering CAD: Structure and function of a mega-enzymatic pyrimidine factory in health and disease. *Protein Science*, 30(10), 1995-2008.

Diler, S. (2016). Veri madenciliği süreçleri ve karar ağaçları algoritmaları ile bir uygulama. In Yüksek Lisans Tezi. Van, Türkiye: <https://tez.yok.gov.tr/UlusalTezMerkezi/>.

Diwani, S., Mishol, S., Kayange, D. S., Machuve, D., & Sam, A. (2013). Overview applications of data mining in health care: The case study of Arusha region. *International journal of computational engineering research*, 3(8), 73-77.

Erdem, S., Özdağoğlu, G. (2008). Analyzing of Emergency Data of a Training and Research Hospital in Aegean Region Using Data Mining”, *Anadolu University Journal of Science and Technology*, 9(2), 261-270.

Eyileten, C., Wicik, Z., De Rosa, S., Mirowska-Guzel, D., Soplinska, A., Indolfi, C., ... & Postula, M. (2018). MicroRNAs as diagnostic and prognostic biomarkers in ischemic stroke—a comprehensive review and bioinformatic analysis. *Cells*, 7(12), 249.

Fastabiqi, I., Asnar, E., & Harlina, H. (2020). The Effect of Citrulline Supplementation on Lactic Acid Blood Level and Oxygen Saturation (SaO₂) in Submaximal Physical Activity. *Folia Medica Indonesiana*, 56(4), 245-247.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.

Fearon, W. R. (1939). The carbamido diacetyl reaction: a test for citrulline. *Biochemical Journal*, 33(6), 902.

Firouzi, F., Rahmani, A. M., Mankodiya, K., Badaroglu, M., Merrett, G. V., Wong, P., & Farahani, B. (2018). Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics. *Future Generation Computer Systems*, 78, 583-586.

Flechner, S. M., Kurian, S. M., Head, S. R., Sharp, S. M., Whisenant, T. C., Zhang, J., ... & Salomon, D. R. (2004). Kidney transplant rejection and tissue injury by gene profiling of biopsies and peripheral blood lymphocytes. *American Journal of Transplantation*, 4(9), 1475-1489.

Forrest, S. (1993). Genetic algorithms: principles of natural selection applied to computation. *Science*, 261(5123), 872-878.

Gambardella, J., Khondkar, W., Morelli, M. B., Wang, X., Santulli, G., & Trimarco, V. (2020). Arginine and endothelial function. *Biomedicines*, 8(8), 277.

Gershman, A., Meisels, A., Lüke, K. H., Rokach, L., Schclar, A., & Sturm, A. (2010). A decision tree based recommender system. *10th International Conference on Innovative Internet Community Systems (I2CS)–Jubilee Edition 2010*–.

Goto, S., Nishioka, T., & Kanehisa, M. (2000). LIGAND: chemical database of enzyme reactions. *Nucleic acids research*, 28(1), 380-382.

Gündoğdu, Ö. E. (2007). Veri madenciliğinde genetik algoritmalar. In Yüksek Lisans Tezi. Kocaeli, Türkiye: <https://tez.yok.gov.tr/UlusalTezMerkezi/>.

Gustafson D.E., Kessel W.C., (1978), „Fuzzy Clustering with a Fuzzy Covariance Matrix“, Conference Paper in Proceedings of the IEEE Conference on Decision and Control, Researchgate, ss.761-766

Han, J., & Kamber, M. (2006). Data Mining Concepts and Techniques, published by Morgan Kauffman.

Han, J., & Kamber, M. (2006). Data mining: concepts and techniques, 2nd. *University of Illinois at Urbana Champaign: Morgan Kaufmann*.

Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. *Morgan kaufmann*.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press.

Hussain, S., Muhammad, L. J., Ishaq, F. S., Yakubu, A., & Mohammed, I. A. (2019). Performance evaluation of various data mining algorithms on road traffic accident dataset. In *Information and Communication Technology for Intelligent Systems* (pp. 67-78). Springer, Singapore.

Javindrasana, J., Cohen, G., Depeursinge, A., Müller, H., Meyer, R., & Geissbuhler, A. (2009). Clinical data mining: a review. *Yearbook of medical informatics*, 18(01), 121-133.

Jadhav, S. D., & Channe, H. P. (2016). Efficient recommendation system using decision tree classifier and collaborative filtering. *Int. Res. J. Eng. Technol*, 3(8), 2113-2118.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.

Joopudi S. vd(2013), „A New Cluster Validity Index for Fuzzy Clustering“, 10 IFACInternational Symposium on Dynamics and Control of Process Systems; ss.325-330

Kamel, H., Abdulah, D., & Al-Tuwaijari, J. M. (2019, June). Cancer classification using gaussian naive bayes algorithm. In *2019 International Engineering Conference (IEC)* (pp. 165-170). IEEE.

Kandemir, A. Ş. (2018). Bulanık Kümeleme Analizi ile Türkiye’deki İllerin Konaklama İstatistiklerine Göre Sınıflandırılması. *Seyahat ve Otel İşletmeciliği Dergisi*, 15(3), 657-668.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27-30.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, 45(D1), D353-D361.

Kaore, S. N., & Kaore, N. M. (2019). Citrulline: Pharmacological Perspectives and Role as a Biomarker in Diseases and Toxicities. In *Biomarkers in Toxicology* (pp. 1073-1093). Academic Press.

Kaore, S. N., Amane, H. S., & Kaore, N. M. (2013). Citrulline: pharmacological perspectives and its role as an emerging biomarker in future. *Fundamental & clinical pharmacology*, 27(1), 35-50.

Kapucu, M. S. (2014). Bilim ve Teknoloji Haberlerinin Ortaokul Öğrencilerinin Daimi Bilim Öğrenme Motivasyonu Üzerine Etkisi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 446-467.

Karabulut, D. (2021). *Hastane Bilgi Yönetim Sistemlerinde Veri Madenciliği: Hasta Profil Tahmini* (Doctoral dissertation).

Karaduman, T. (2019). Yapay zekâ uygulama alanları. *Gazi Üniversitesi, Bilisim Enstitüsü, Adli Bilisim. ABD*.

Kartal, C. (2020). Destek Vektör Makineleri ile Borsa Endekslerinin Tahmini. *Itobiad: Journal of The Human & Social Science Researches*, 9(2).

Kautish, S., & Ahmed, R. K. A. (2016). A comprehensive review of current and future applications of data mining in medicine & healthcare. *algorithms*, 6(8), 9.

Kawasaki, S., Miyake, C., Kohchi, T., Fujii, S., Uchida, M., & Yokota, A. (2000). Responses of wild watermelon to drought stress: accumulation of an ArgE homologue and citrulline in leaves during water deficits. *Plant and Cell Physiology*, 41(7), 864-873.

Kayaalp, F., & Başarslan, M. S. (2018). Open source data mining programs: a case study on R. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 6(2), 455-468.

Kızılkaya, Y. M., & Oğuzlar, A. (2018). Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili İle Kıyaslanması. *Karadeniz Uluslararası Bilimsel Dergi*, 37(37), 90-98.

Kido, J., Matsumoto, S., Sugawara, K., Sawada, T., & Nakamura, K. (2021). Variants associated with urea cycle disorders in Japanese patients: Nationwide study and literature review. *American Journal of Medical Genetics Part A*, 185(7), 2026-2036.

Kiyici, F., Eroğlu, H., Kishali, N. F., & Burmaoglu, G. (2017). The effect of citrulline/malate on blood lactate levels in intensive exercise. *Biochemical genetics*, 55(5), 387-394.

Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of healthcare information management*, 19(2), 65.

Koyuncugil, A., & Özgülbaş, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2).

Kumar, N., Hoque, M., & Sugimoto, M. (2018). Robust volcano plot: identification of differential metabolites in the presence of outliers. *BMC bioinformatics*, 19(1), 1-11.

Leiman, P. G., & Taylor, N. M. (2019). Reference Module in Life Sciences.

Liao, S. H. (2003). Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert systems with applications*, 25(2), 155-164.

Liao, Y., & Vemuri, V. R. (2002). Use of k-nearest neighbor classifier for intrusion detection. *Computers & security*, 21(5), 439-448.

Liu, F., Bao, L. S., Liang, R. J., Zhao, X. Y., Li, Z., Du, Z. F., & Lv, S. G. (2021). Identification of rare variants causing urea cycle disorders: A clinical, genetic, and biophysical study. *Journal of cellular and molecular medicine*, 25(8), 4099-4109.

Marco-Marín, C., Escamilla-Honrubia, J. M., Llácer, J. L., Seri, M., Panza, E., & Rubio, V. (2020). Δ 1-Pyrroline-5-carboxylate synthetase deficiency: An emergent multifaceted urea cycle-related disorder. *Journal of Inherited Metabolic Disease*, 43(4), 657-670.

Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D. N., Hanspers, K & Kutmon, M. (2021). WikiPathways: connecting communities. *Nucleic acids research*, 49(D1), D613-D621.

Mering, C. V., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., & Snel, B. (2003). STRING: a database of predicted functional associations between proteins. *Nucleic acids research*, 31(1), 258-261.

Michael, B., & Gordon, L. (2000). Mastering data mining: the art and science of customer relationship management. *New York: John Wiley&Sons*.

Mishra, A. K., & Ratha, B. K. (2016). Study of random tree and random forest data mining algorithms for microarray data analysis. *International Journal on Advanced Electrical and Computer Engineering*, 3(4), 5-7.

Muhammad, L. J., Sani, S., Yakubu, A., Yusuf, M. M., Elrufai, T. A., Mohammed, I. A., & Nuhu, A. M. (2017). Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano–Wudil highway. *Int J Database Theory Appl*, 10(11), 197-208.

Nadiammai, G. V., & Hemalatha, M. J. E. I. J. (2014). Effective approach toward Intrusion Detection System using data mining techniques. *Egyptian Informatics Journal*, 15(1), 37-50.

- Nagamani, S. C., Erez, A., & Lee, B. (2012). Argininosuccinate lyase deficiency. *Genetics in medicine*, 14(5), 501-507.
- Nicholson, J., Darzi, A., Holmes, E., & Lindon, J. C. (Eds.). (2016). *Metabolic phenotyping in personalized and public healthcare*. Academic Press.
- Nitzahn, M., & Lipshutz, G. S. (2020). CPS1: Looking at an ancient enzyme in a modern light. *Molecular genetics and metabolism*, 131(3), 289-298.
- Oğuz, M., & Akbaş, S. (1997). Genetik algoritmalar. *YTÜ Endüstri Mühendisliği*, Bitirme Tezi.
- Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21).
- Oğuzlar, A. (2004). Veri Madenciliğine Giriş, *Ekin Kitabevi*, Bursa.
- Papadia, C., Osowska, S., Cynober, L., & Forbes, A. (2018). Citrulline in health and disease. Review on human studies. *Clinical Nutrition*, 37(6), 1823-1828.
- Parker, D.G. (2004). Data Mining: Modules in emerging fields, CD-ROM. vol 7.
- Pei, M., Goodman, E. D., & Punch, W. F. (1998, October). Feature extraction using genetic algorithms. In *Proceedings of the 1st International Symposium on Intelligent Data Engineering and Learning, IDEAL* (Vol. 98, pp. 371-384).
- Peter, S. C., Dhanjal, J. K., Malik, V., Radhakrishnan, N., Jayakanthan, M., Sundar, D., ... & Jayakanthan, M. (2019). Encyclopedia of bioinformatics and computational biology. *Ranganathan, S., Grib-skov, M., Nakai, K., Schönbach, C., Eds*, 661-676.
- Pirim, A. G. H. (2006). Yapay zeka. *Yaşar Üniversitesi E-Dergisi*, 1(1), 81-93.
- Pradhan, R., Aggarwal, M., Maheshwari, D., Chaturvedi, A., & Sharma, D. K. (2020, February). Diabetes mellitus prediction and classifier comparative study. In *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)* (pp. 133-139). IEEE.
- Priyam, A., Abhijeeta, G. R., Rathee, A., & Srivastava, S. (2013). Comparative analysis of decision tree classification algorithms. *International Journal of current engineering and technology*, 3(2), 334-337.
- Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.
- Rimando, A. M., & Perkins-Veazie, P. M. (2005). Determination of citrulline in watermelon rind. *Journal of Chromatography A*, 1078(1-2), 196-200.
- Rosli, M. M., Tempero, E., & Luxton-Reilly, A. (2016, February). What is in our datasets? Describing a structure of datasets. In *Proceedings of the Australasian Computer Science Week Multiconference* (pp. 1-10).

Ruspini E.H.,(1970), Numerical Methods for Fuzzy Clustering*, Information Sciences,2, ss.319-350

Sağlık Bakanlığı, www.saglik.gov.tr. ErişimTarihi:05.12.2022.

Sanchez-Pinto, L. N., Luo, Y., & Churpek, M. M. (2018). Big data and data science in critical care. *Chest*, 154(5), 1239-1248.

Shiozawa, T., Iyama, S., Toshima, S., Sakata, A., Usui, S., Minami, Y., ... & Noguchi, M. (2016). Dimethylarginine dimethylaminohydrolase 2 promotes tumor angiogenesis in lung adenocarcinoma. *Virchows archiv*, 468(2), 179-190.

Shouman, M., Turner, T., & Stocker, R. (2013). Integrating clustering with different data mining techniques in the diagnosis of heart disease. *J. Comput. Sci. Eng*, 20(1), 1-10.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern recognition letters*, 10(5), 335-347.

Singh, M. (1997). Learning Bayesian networks from incomplete data. *AAAI/IAAI*, 995, 534-539.

Stolba, N., & Tjoa, A. M. (2006). The relevance of data warehousing and data mining in the field of evidence-based medicine to support healthcare decision making. *International Journal of Computer Systems Science and Engineering*, 3(3), 143-148.

Sureda, A., Córdova, A., Ferrer, M. D., Pérez, G., Tur, J. A., & Pons, A. (2010). L-citrulline-malate influence over branched chain amino acid utilization during exercise. *European journal of applied physiology*, 110(2), 341-351.

Şentürk, A. (2006). Veri Madenciliği Kavram ve Teknikler, *Ekin Kitabevi*, Bursa.

Tuğ, E. (2005). Genetik algoritmalar ile tıbbi veri madenciliği. *Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi*.

Tüzüntürk, S. (2010). Veri madenciliği ve istatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 29(1), 65-90.

Uğur, A., & Kınacı, A. C. (2006). Yapay zeka teknikleri ve yapay sinir ağları kullanılarak web sayfalarının sınıflandırılması. *XI. Türkiye'de İnternet Konferansı (inet-tr'06)*, Ankara, 1(4).

Ulutağay G.(2004), Bulanık C-Ortalamlar Kümeleme Analizi ve Uygulamaları, Yayınlanmamış Dr. Tezi, Ege Üniversitesi, Fen Bilimleri Enstitüsü

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Wang, H., & Druzdzel, M. J. (2013). User interface tools for navigation in conditional probability tables and elicitation of probabilities in Bayesian networks. *arXiv preprint arXiv:1301.4430*.

Wiki Pathway, https://search.crossref.org/?q=10.3180/REACT_847.1. Eriřim tarihi: 05.09.2022

Xu, Y., Liang, Z., Legrain, C., R ger, H. J., & Glansdorff, N. (2000). Evolution of arginine biosynthesis in the bacterial domain: novel gene-enzyme relationships from psychrophilic Moritella strains (Vibrionaceae) and evolutionary significance of N- -acetyl ornithinase. *Journal of bacteriology*, 182(6), 1609-1615.

Yal n, L. (2019). "Saėlık Sekt r nde Veri Madenciliėi", Y ksek Lisans Tezi, *Milli Savunma  niversitesi, Hezarfen Havacılık ve Uzay Teknolojileri Enstit s *, End stri M hendisliėi Programı, İstanbul.

Yal n,  . (2013). Veri Madenciliėi Y ntemleri, 2nd Ed., *Papatya Yayıncılık*.

Yavuz, H. U. (2006). Arjinin Ve Egzersiz. *Spor Bilimleri Dergisi*, 17(3), 143-157.

Yokota, A., Kawasaki, S., Iwano, M., Nakamura, C., Miyake, C., & Akashi, K. (2002). Citrulline and DRIP-1 protein (ArgE homologue) in drought tolerance of wild watermelon. *Annals of Botany*, 89(7), 825-832.

Y ksel, S., Eroėlu,  ., & Y ksel, S. (2018). Arařtırmalarda Karar S reci Ve Bayes Teorimi. *Ejons International Journal*, 2(6), 56-66.

7. EKLER

Ek-1 : Etik Onay



8. ÖZGEÇMİŞ



9. İNTİHAL RAPORU

YAPAY ZEKADA VERİ MADENCİLİĞİ YÖNTEMİ KULLANARAK
BİYOİNFORMATİK YAZILIM GELİŞTİRİLMESİ: SİTRÜLİN İLE BİR
UYGULAMA

ORJİNALLİK RAPORU

% 21	% 20	% 3	% 10
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	acikbilim.yok.gov.tr İnternet Kaynağı	% 5
2	dergipark.org.tr İnternet Kaynağı	% 3
3	docplayer.biz.tr İnternet Kaynağı	% 3
4	www.mehmetbatal.com İnternet Kaynağı	% 2
5	www.researchgate.net İnternet Kaynağı	% 2
6	issuu.com İnternet Kaynağı	<% 1
7	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	<% 1
8	discourse.mc-stan.org İnternet Kaynağı	<% 1