VERİ BİLİMİ DERGİSİ
www.dergipark.gov.tr/veri

İnsan Bağırsak Mikrobiyomunda Bakteri Kaynaklı İnsan Benzeri MikroRNA Tespiti

Ayşenur SOYTÜRK PATAT[1*3], Özkan Ufuk NALBANTOĞLU[2,3]

[1]*Necmettin Erbakan Üniversitesi, Fen Fakültesi, Moleküler Biyoloji ve Genetik, Konya*
[2]*Erciyes Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Kayseri*
[3]*Erciyes Üniversitesi Genom ve Kök Hücre Merkezi, Biyoinformatik Sistemler Biyolojisi, Kayseri*

Makale Bilgisi

Başvuru:
27/06/2021
Kabul:
29/11/2021

Özet

MikroRNA'lar 19-25 nükleotid aralığında küçük RNA parçalarıdır [1]. Gen ifadelerinde düzenleyici oldukları yürütülen deneysel çalışmalarca ispatlanmıştır [2]. MikroRNA'lar RNA'nın kodlanmayan bölgelerinden oluşur ve kodlanmayan bu bölgelerin büyüklüğü, ikincil yapı oluştururken oluşabilecek çoklu ihtimaller sebebiyle tespit edilmeyi zorlaşmıştır. Bu düzenleyici dizi içeriklerinin türler arasında korunmuş olabileceğine yönelik hesaplamalı çalışmalar mevcuttur [3]. İnsan bağırsak mikrobiyomu birçok mikroorganizmaya ev sahipliği yapmakla birlikte birçok hastalıkla ilişkili olduğu bilinmektedir [4]. Bu çalışmanın amacı insan mikrobiyomunda konakçıyı hedeflediği düşünülen bakteri kaynaklı moleküler mekanizmalardan insan mikroRNA'sına benzer dizilerin varlığının geliştirilen makine öğrenmesi modeli kullanılarak aranması, bu dizilerin karaciğer sirozu hastası (n=58) ve sağlıklı kontrolde (n=53) anlamlı bir fark oluşturup oluşturmadığının incelenmesidir. Bu gruplarda geliştirilen model ile tarama yapılıp, insan mikroRNA'sına benzer yapılar gösteren taksonlar belirlenmiştir. Belirlenen taksonlar karaciğer sirozu hastaları ve sağlıklı kontroller arasında anlamlı bir fark oluşturduğu istatistik testler ile doğrulanmıştır. Geliştirilen model ile insan bağırsak mikrobiyomunun, insan mikroRNA'sına benzeyen çok sayıda dizi içerdiği gözlenmiştir. Ayrıca hastalık açısından bazı türler barındırdıkları varsayılan mikroRNA dizilerinde çeşitlilik sergilediği gözlemlenmiştir. Bu yapıların varlığının belirlenmesi karaciğer sirozu hastalığında ilerleyen çalışmalara ışık tutacak niteliktedir.

Anahtar Kelimeler: *Metagenom, MikroRNA, Makine Öğrenmesi, Karaciğer Siroz*

Human-like MicroRNA Detection of Bacterial Origin in Human Gut Microbiome

Abstract

MicroRNAs are small pieces of RNA in the range of 19-25 nucleotides [1]. It has been proven by experimental studies that they are regulators of gene expression [2]. MicroRNAs consist of non-coding regions of RNA, and the size of these non-coding regions has made it difficult to detect due to the multiple possibilities that may occur when forming secondary structures. There are computational studies suggesting that the contents of these regulatory sequences may be conserved across species [3]. Although the human gut microbiome is home to many microorganisms, it is known to be

* İletişim e-posta:
asoyturk@erbakan.edu.tr

associated with many diseases [4]. The aim of this study is to search for the presence of human microRNA-like sequences from bacterial molecular mechanisms thought to target the host in the human microbiome with the developed machine learning model, and to examine whether these sequences make a significant difference in liver cirrhosis patients (n=58) and healthy controls (n=53). Scanning was performed with the model developed in these groups and taxa showing structures similar to human microRNA were determined. The determined taxa were examined using statistical tests to establish a significant difference between liver cirrhosis patients and healthy controls. With the developed model, a large number of sequence contents similar to human microRNA were observed in the human gut microbiome. In addition, it has been observed that there is diversity in the microRNA sequences that are assumed to host some species in terms of disease. Determining the presence of these structures will shed light on future studies in liver cirrhosis.

Keywords: *Metagenome, MicroRNA, Machine Learning, Liver Cirrhosis*

1 Giriş

MikroRNA'lar gen düzenleyici mekanizmalar olup postranskripsiyonel olarak düzenleyici oldukları 1993 yılında Victor Ambros ve arkadaşları tarafından keşfedilmiştir [5]. Düzenleyici olarak görev yapmaları sebebiyle birçok hastalıktaki transkripsiyonel profille ilişkili oldukları düşünülmüştür. Nörodejeneratif bir hastalık olan Alzheimer hastalığı [6] Akhter'ın çalışmasında mikroRNA ile ilişkilendirilmeye çalışılmış ve bu hastalığa ilişkili tanıda kullanılabilecek muhtemel mikroRNA biyobelirteçleri belirtilmiştir. Benzer şekilde kanser ile mikroRNA'ların ilişkisini Cascione L. ve arkadaşları omik verilerini kullanarak belirlemeye çalışmış ve kanserle ilişkilendirilmiş mikroRNA'ların hedeflediği mRNA'ları [7] belirlemeye çalışmışlardır. Bir diğer çalışma ise akciğer hastalıklarıyla ilişkili olup KOAH hastalığındaki mikroRNA'nın düzenleyiciyi etkileri belirlenmeye çalışılmıştır. Melina M. Musri ve arkadaşlarının yürüttüğü bu çalışmada KOAH hastalarından alınan örneklerde miR-197'nin mikroRNA ekspresyonunda değişiklikler bulunmuştur [8]. Bununla birlikte kardiyovasküler hastalıklar [9], epigenetik faktörlü hastalıklar [10], hipertansiyon [11], inflamatuar bağırsak hastalığı [12] gibi birçok hastalıkta mikroRNA'nın önemi ortaya konmuştur. İnsan genomunda ise günümüze değin deneysel olarak doğrulanmış 1917 adet mikroRNA kodlayan dizi bulunmaktadır. Düzenleyici olan bu küçük RNA parçalarının ise sadece mesajcı RNA ile eksprese fonksiyonu bilinebilmekte olup, olgun mikroRNA'ların ikincil yapılarının oluşması için çoklu ihtimallerin olması, keşfinin laboratuvar koşulları için yoğun emekli, maliyetli ve karmaşık kılmıştır. Laboratuvar koşullarında belirlenmesi zor olan mikroRNA'lar son zamanlarda *in silico* yöntemler kullanılarak tahmin edilmeye çalışılmaktadır [3]. Birçok canlıya ait mikroRNA yapılarını içeren ve sekanslama verilerinden muhtemel mikroRNA tahmin edebilen bir araç olan Rfam[13] , mikroRNA hedef genlerini ve metabolik yolaklarını içeren miRNApath [14], insan hastalıkları ile ilişkilendirilmiş verileri barındıran miR2Disease veri tabanı gibi MikroRNA tespiti ve anlamlandırılması için birçok *in silico* araç geliştirilmiştir. Yeni Nesil Dizileme (YND) ile birlikte mikroRNA çalışmaları yanında mikrobiyal toplulukların genetik materyalini kültür bağımsız ortaya çıkarabilen metagenom çalışmaları da büyük bir ivme kazanmıştır.

İnsan vücudu çok sayıda bakteri, mantar, virüs gibi birçok mikroorganizmaya ev sahipliği yapar. Bu mikroorganizmalar insanlar ile komensal bir ilişki içine girmektedir. Mikrobiyomumuz, ürettiği metabolitler ve genetik materyal ile insan metabolizması ve bağışıklığı için yardımcı ve düzenleyici rol oynarken, bu homeostatik dengenin sekteye uğradığı durumlarda ise patolojik fenotiplerle ilişkilendirilmektedir. İnsan genomu yaklaşık 20.000 protein kodlayan gen içerir, mikrobiyomumuzda bulunan mikrobiyal genomlar ise kendi genomumuzunkinden çok daha yüksek sayıda protein kodlama potansiyeline sahiptir. Bu açıdan ikinci genomumuz olarak da adlandırılmaktadırlar [15]. Bağırsak mikrobiyomundaki çeşitlilik ve barındırdığı varyasyonlar, insanlarda sağlık ve hastalıklarla ilişkilendirilmektedir ve burada konağı etkileyen bir dizi moleküler mekanizma bulunmaktadır [16]. Mikrobiyoma özgü genlerin, insan konakçı tarafından ihtiyaç duyulan metabolitleri oluşturmak için kullanıldığı bilinmektedir [17]. Bu mikrobiyoma özgü genlerin düzenlenmesi için bilinen birçok gen düzenleme mekanizmaları vardır. Gen düzenleme aşamalarından birini oluşturan bakteriyel mikroRNA'ların tahmini ve

insan transkriptomunda olası hedeflerini araştıran öncü çalışmalar mevcuttur [18] [19].

Makine öğrenmesi matematiksel işlemlerle mevcut veriler üzerinden çıkarımlar yaparak, bu çıkarımlarla bilinmeyen veriler üzerinde tahminler yapabilen bilgisayar algoritmalarıdır. Biyolojik verilerin giderek artması ile makine öğrenmesi yöntemleri biyolojik veriyi anlamlandırmak üzere kullanılmaya başlanmıştır. Makine öğrenmesi, biyolojik verinin matematiksel temsiliyetinden faydalanarak modelleme yapar ve mikroRNA tespiti için de kullanılan *in silico* bir yöntemdir [20]. İnsan hücresi transkriptomunda bakteriyel mikroRNA'ların hedeflerinin araştırıldığı ve 37 farklı bakteri genomuna ait olduğu tahmin edilen, 68 bakteri RNA'sının, 47 farklı hastalıkta etkili olabileceği sonucuna varılan öncü çalışmadan hareketle oluşturulan [18], bu çalışmamızın amacı insan mikrobiyomunda konakçıyı hedeflediği düşünülen moleküler mekanizmalardan insan mikroRNA'sına benzer dizilerin varlığını araştırmak ve bu dizilerin varlığının siroz hastaları ve sağlıklı kontroller arasında anlamlı farklılıklar oluşturup oluşturmadığının incelenmesidir. Bu çalışmada konakçıyı hedeflediği düşünülen, muhtemel insan mikroRNA'sını taklit eden bakteriyel mikroRNA tespiti için makine öğrenmesi yöntemlerine dayalı *in silico* bir tespit yaklaşımı geliştirilmiştir.

2 Materyal Method

2.1 Veri Setleri

Bu çalışmada biri hesaplamalı mikroRNA keşfini sağlayan modellerin eğitimi amaçlı, diğeri ise bu modellerin uygulanması için iki ayrı veri seti kullanılmıştır. İlk veri seti deneysel olarak bulunmuş öncü ve olgun olarak sınıflandırılmış birçok türe ait mikroRNA dizilerini içeren veri seti olan MiRBase olup, insan mikroRNA'sı ile diğer türlerin mikroRNA'larını ayırmak için kullanılan sınıflandırıcının oluşturulmasında eğitim ve test verisini oluşturmuştur [21]. Kullanılan ikinci veri seti karaciğer sirozu hastaları ve sağlıklı kontrollerin bağırsak mikrobiyomunu içeren veri setidir [22]. Bu veri seti siroz hastalarında ve sağlıklı kontrollerde insan mikroRNA'sına benzer yapı gösteren dizileri belirlemede kullanılmıştır. Veri setinden bir alt küme oluşacak şekilde 58 kişisi siroz hastası ve 53 sağlıklı kontrol, toplamda 111 kişiden oluşan okumaları birleştirilmiş veri seti sınıflandırma için kullanılmıştır.

2.2 Veri Seti Erişim Kodu

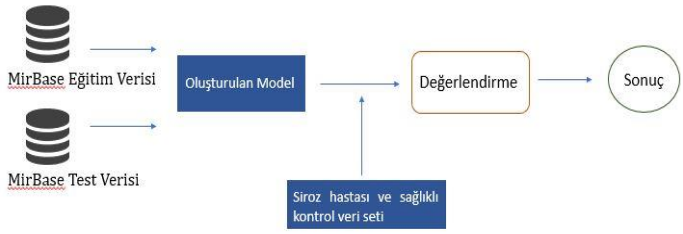
Bu çalışmada ilk veri seti olan ve hesaplamalı mikroRNA keşfi için kullanılan eğitim, test setini oluşturan veriye https://www.mirbase.org/ adresinden ulaşılabilir. Kullanılan ikinci veri seti karaciğer sirozu hastaları ve sağlıklı kontrollerin bağırsak mikrobiyomunu içeren ve oluşturulan modelle insan benzeri mikroRNA tespiti amacıyla kullanılan veri setidir. Tüm örnekler için ham Illumina okuma verileri Avrupa Biyoinformatik Enstitüsü, Avrupa Nükleotid Arşivinde (ENA) saklanmaktadır. Veriye https://www.ebi.ac.uk adresinden ERP005860 erişim numarasıyla ulaşılabilir.

2.3 MikroRNA Dizilerinin Spektral Temsili

Nükleotid dizilerinin çeşitli örüntülerinin karakterize edilmesi amacıyla içerdiği oligonükleotid frekanslarının spektral olarak k-mer kümeleri halinde temsil edilmesi biyolojik dizi analizinde sıklıkla kullanılan bir yöntemdir. Yürütülen çalışmada öncü-mikroRNA tespiti için kullanılmıştır. Sekanslama sonucu elde edilen okumaların parçalara ayrılmasına dayanır ve bu küçük parçalar k-mers ya da n-grams olarak adlandırılır [23]. Örneğin 1-mer için üretilecek temsil kümesi {A,U,C,G}, 2-mer içinse {AA,AU,AC,AG,UA,UU,UC,UG,CA,CU,CC,CG,GA,GU,GC, GG} şeklindedir. MikroRNA dizilerinin kısa olması ve nispeten uzun oligonükleotidlerin frekanslarını kestirecek örneklemi oluşturamaması sebebiyle bu çalışmada k-mer'ler k=4 olacak şekilde belirlenmiştir. Bu şekilde her bir okumanın içerdiği tüm 4-mer'i kanonik sırayla içerecek şekilde mikroRNA temsiliyet vektörleri oluşturulmuştur. Sekanslama sonucu oluşmuş olan hangi nükleotid olduğu tam belli olmayan örneğin N,K,R… gibi k-mer'ler filtrelenmiştir.

2.4 Sınıflandırma Yaklaşımları

İlk olarak Rastgele orman (RF) [24], Naive Bayes(NB) [25], Güçlendirilmiş Karar ağaçları(XGBoost) [26], K-en yakın komşu(KNN) [27] sınıflandırma algoritmaları uygulanmıştır. En iyi performans verdiğinden XGBoost algoritması sınıflandırma işlemi için kullanılmıştır. Model eğitilirken MirBase'den alınan verinin %80'i eğitim için %20'side test verisini oluşturacak şekilde ayrılmıştır. Burada veri ayırma işleminde test verisi eğitimde kullanılmayacak şekilde ayrılmıştır. Oluşturulan makine öğrenmesine ait yaklaşım Şekil 1'de gösterilmiştir.



Şekil 1. Oluşturulan Makine Öğrenmesi Yapısı

2.5 Performans Ölçütleri

Bir modelin kullanılabilir olması performans değerlerine bağlıdır ve bizim çalışmamız için kullandığımız performans ölçütleri doğruluk değeri (accuracy), hassaslık(precision), geri çağırma(recall) ve F1 değeri(f1-score) şeklindedir. Bu parametreler için gerekli olan değişkenler ve bu parametrelerin hesaplama fonksiyonları ise şu şekildedir:

Gerçek Pozitifler(GP): Doğru olduğu öngörülen pozitif değerlerdir. Gerçek sınıf değerinin pozitif olduğu ve öngörülen sınıf değerinin de pozitif olduğu durumların sayısını ifade eder.

Gerçek Negatifler(GN): Doğru olduğu öngörülen negatif değerlerdir. Gerçek sınıf değerinin negatif olduğu ve öngörülen sınıf değerinin de negatif olduğu durumların sayısını ifade eder.

Yanlış Pozitifler(YP): Gerçek sınıfın negatif ve öngörülen sınıfın pozitif olduğu durumların sayısıdır.

Yanlış Negatifler(YN): Gerçek sınıfın pozitif ve öngörülen sınıfın negatif olduğu durumların sayısıdır.

$$hassaslık = \frac{GP}{GP + YP} \quad (1)$$

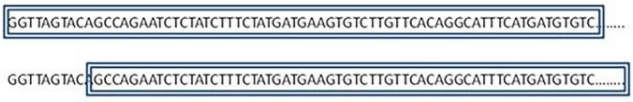
$$doğruluk = \frac{GP + GN}{GP + YP + YN + GN} \quad (2)$$

$$geri\ çağırma = \frac{GP}{GP + GN} \quad (3)$$

$$F1\ değeri = \frac{2 * (geri\ çağırma * hassaslık)}{geri\ çağırma + hassaslık} \quad (4)$$

2.6 İnsan Bağırsak Mikrobiyomunda Sınıflandırma

Performans ölçütlerine göre en iyi değerleri veren XGBoost algoritması ile geliştirilen model ve olgun mikroRNA uzunlukları dikkate alınarak 70 bazlık parçalar halinde sınıflandırıcıda çalıştırıldı. Bir sonraki 70'lik parça ilk parçanın son 60 bazını içerecek şekilde kaydırılarak sonraki pencere yapısı oluşturulmuş oldu. Oluşturulan pencere yapısı Şekil 2'de gösterilmiştir. Bu şekilde her bir birey için metagenom verisi taranmış oldu. Bu pencerelerde sınıflandırıcı puanlama yaparken 0.95 sınıflandırma tahmin sınır değerini geçen parçaların muhtemel insan mikroRNA yapısı tespiti ile belirlendi. Belirlenen lokasyonlar için tür ataması yapıldı ve tür ataması yüksek doğruluk değeri ve hızlı sınıflandırma sonuçları elde etmek için k-mer eşleşmeleri kullanan taksonomik bir sınıflandırma sistemi olan Kraken2 [28] aracı kullanılarak yapıldı.



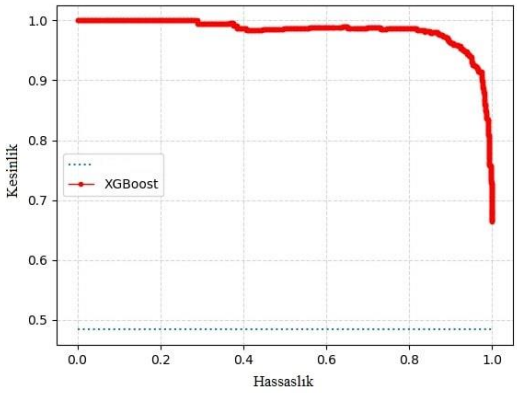
Şekil 2. Oluşturulan pencere yapısı

2.7 İstatistiki testler

Her bir örnekte insan mikroRNA'sı olduğu varsayılan lokasyonlarda tür ataması yapıldıktan sonra karaciğer sirozu hastası ve sağlıklı kontrol grubu arasında t-test uygulanarak insan mikroRNA'sı yapısı taşıyan türlerin görülme oranı olarak farklılıklar belirlenmiştir. Burada birbirinden bağımsız iki grup olduğu için t-test uygulanmış ve en düşük p değeri alan türler bu test ile belirlenmiştir.

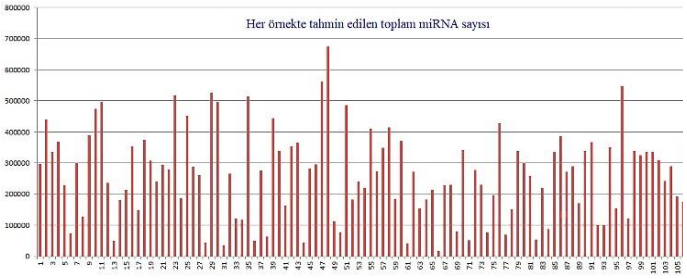
3 Sonuçlar

Sınıflandırma işlemi için denenen algoritmalar ve performans ölçütü denemeleri Tablo 1'de gösterilmiştir. Bu sonuçlara göre Rastgele Orman algoritması'nın doğruluk değeri yaklaşık %87 olarak bulunmuştur, K-en Yakın Komşu algoritması için doğruluk değeri yaklaşık %81, Lojistik Regresyon algoritması için doğruluk değeri yaklaşık %93, XGBoost algoritması için doğruluk değeri yaklaşık %93 ve Naive Bayes algoritması için doğruluk değeri yaklaşık %91 olarak bulunmuştur. Bu değerlendirmeler sonucunda en iyi performansı veren algoritma olan XGBoost algoritması sınıflandırma işlemi kullanılmıştır. XGBoost algoritmasına ait Kesinlik-Hassaslık grafiği Şekil 3'te gösterilmiştir.



Şekil 3. XGBoost'a ait Kesinlik-Hassaslık grafiği

Bu grafiğe göre eşik değeri 0.95 belirlenen sınıflandırıcı her birey için taramalar yapıp muhtemel insan mikroRNA'larını belirlemiştir. Belirlenen bu lokasyonlarda tür ataması yapılmıştır. İlk 58 birey siroz hastası ve sonraki 53 birey sağlıklı kontrol olmak üzere toplam 111 bireyde insan mikroRNA'sı olduğu tahmin edilen tür sayıları Şekil 4'de gösterilmiştir.



Şekil 4. Her bireyde insan mikroRNA'sı olduğu tahmin edilen tür sayıları

Bağımsız iki grup olan siroz hasları ve sağlıklı kontroller için t-test uygulandı ve kabul edilen p değeri literatürde belirtildiği gibi 0.05 olarak alındı. P değeri 0.05 in altında olan 729 adet farklı tür ve cins seviyesi bulunmuştur. Bulunan bu seviyelerde en iyi p değerini veren cins 8.5e-12 değeri ile *Veillonella*'a aittir. Bir diğer en yüksek p değeri alan tür 1.9e-8 ile *Veillonella rodentium'a* aittir. En düşük 15 p değeri veren türler ve cinsleri gösteren Tablo 2'de gösterilmiştir.

4 Tartışma

Yapılan çalışma insan benzeri mikroRNA'ların bağırsak mikrobiyomu içerisinde *in silico* olarak tespit edilmesi yönünden öncül bir uygulama olup bir hastalık kohortunda hasta ve sağlıklı bireylerde tahminlenen mikroRNA dizilerinin istatistiksel olarak anlamlı bir şekilde farklı bolluklarda olduğunu göstermesi açısından önem arzetmektedir. İstatistiki test sonucu önemli p değeri gösteren taksonomik seviyeler daha önce siroz hastalığı ile ilişkilendirilmiş türler olması nedeniyle de önemli bir sonuç oluşturmaktadır [14]. YND teknolojilerinin gelişmesi ile birlikte artan veriler sonuçların güvenilirliğini arttırmış ve diğer çalışmalara ışık tutmuştur. Bu anlamda yapılan çalışmada genel algoritmalar yerine çalışma için geliştirilmiş ve daha büyük çalışma grubu için eğitilmiş makine öğrenmesi yöntemleri, sonuçların güvenirliliğini artırmak için gerekli olabilir. Aynı zaman da insan MikroRNA'sına benzer yapı gösteren lokasyonların belirlenmesi ileriki çalışmalarda önemli hastalık biyobelirteçlerin bulunabileceğini göstermiştir. Moleküler mimikri, yani mikroorganizmalar tarafından üretilen ve insan organizması tarafından üretilen biyomoleküllere benzemesi dolayısıyla insan metabolizmasında doğrudan rol alan dış kaynaklı ürünler çeşitli hastalık kontekslerinde önem taşımaktadır. Yürütülen *in silico* çalışmanın öne sürdüğü en önemli hipotez, insan bağırsağındaki mikroorganizmaların ürettiği mikroRNA'ların moleküler mimikri ile insan mikroRNA'ları şeklinde davranıp insan genlerini regüle ederek çeşitli hastalıkların patolojisinde rol oynayabileceği görüşüdür. Çalışma sonucunda elde edilen siroz hastaları ve sağlıklı bireylerin bağırsak bakterilerinde tespit edilen insan mikroRNA'sı benzeri moleküllerin miktarlarının istatistiki olarak anlamlı bir şekilde farklılık gösterdiği bulgusu bu görüşe destek vermektedir.

Tablo 1. Denenen algoritmalar ve performans değerleri

Sınıflandırma	Kesinlik	Hassaslık	F-1 Skoru	Doğruluk
Rastgele Orman	0 0.87	0 0.89	0 0.88	0.876
Algoritması	1 0.89	1 0.86	1 0.87	
K-en Yakın Komşu	0 1.00	0 0.63	0 0.77	0.812
Algoritması	1 0.72	1 1.00	1 0.84	
Logistic Regression	0 0.95	0 0.92	0 0.93	0.932
Algoritması	1 0.92	1 0.95	1 0.93	
XGBoost	**0 0.97**	**0 0.90**	**0 0.93**	**0.933**
Algoritması	**1 0.90**	**1 0.97**	**1 0.93**	
Naive Bayes	0 0.90	0 0.93	0 0.91	0.911
Algoritması	1 0.92	1 0.90	1 0.91	

Tablo 2. En düşük 15 p değeri veren türler ve cinsler

Atanmış Türler ve Cinsler	Hesaplanan p Değerleri
Lactobacillus	9.82e-05
Veillonella	8.51e-12
Bacteroides heparinolyticus	7.61e-05
Streptococcus salivarius	7.34e-06
Bacteroides helcogenes P 36-108	5.20e-05
Paludibacter propionicigenes WB4	4.78e-05
Streptococcaceae	4.14e-07
Streptococcus sp. FDAARGOS_192	3.34e-05
Streptococcus	2.99e-06
Streptococcus salivarius JIM8777	2.20e-05
Veillonella rodentium	1.91e-08
Lactobacillaceae	1.54e-05

5 Kaynaklar

[1] A. M. Mohr and J. L. Mott, 'Overview of microRNA biology', *Semin. Liver Dis.*, vol. 35, no. 1, pp. 3–11, Feb. 2015.

[2] J. Krol, I. Loedige, and W. Filipowicz, 'The widespread regulation of microRNA biogenesis, function and decay', *Nat. Rev. Genet.*, vol. 11, no. 9, pp. 597–610, Sep. 2010.

[3] M. Yousef, W. Khalifa, İ. E. Acar, and J. Allmer, 'MicroRNA categorization using sequence motifs and k-mers', *BMC Bioinformatics*, vol. 18, no. 1, p. 170, Mar. 2017.

[4] V. Caputi and M. C. Giron, 'Microbiome-Gut-BrainAxis and Toll-Like Receptors in Parkinson's Disease', *Int. J. Mol. Sci.*, vol. 19, no. 6, p. 1689, Jun. 2018.

[5] R. C. Lee, R. L. Feinbaum, and V. Ambros, 'The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.', *Cell*, vol. 75, no. 5, pp. 843–854, Dec. 1993.

[6] R. Akhter, 'Circular RNA and Alzheimer's Disease.', *Adv. Exp. Med. Biol.*, vol. 1087, pp. 239–243, 2018.

[7] L. Cascione, 'Integration of Omics Data to Identify Cancer-Related MicroRNA.', *Methods Mol. Biol.*, vol. 1970, pp. 85–99, 2019.

[8] M. M. Musri *et al.*, 'MicroRNA Dysregulation in Pulmonary Arteries from Chronic Obstructive Pulmonary Disease. Relationships with Vascular Remodeling.', *Am. J. Respir. Cell Mol. Biol.*, vol. 59, no. 4, pp. 490–499, Oct. 2018.

[9] A. Wojciechowska, A. Braniewska, and K. Kozar-Kamińska, 'MicroRNA in cardiovascular biology and disease.', *Adv. Clin. Exp. Med. Off. organ Wroclaw Med. Univ.*, vol. 26, no. 5, pp. 865–874, Aug. 2017.

[10] L. Zhang, Q. Lu, and C. Chang, 'Epigenetics in Health and Disease.', *Adv. Exp. Med. Biol.*, vol. 1253, pp. 3–55, 2020.

[11] X. Li, Y. Wei, and Z. Wang, 'microRNA-21 and hypertension.', *Hypertens. Res.*, vol. 41, no. 9, pp. 649–661, Sep. 2018.

[12] K. Fisher and J. Lin, 'MicroRNA in inflammatory bowel disease: Translational research and clinical implication.', *World J. Gastroenterol.*, vol. 21, no. 43, pp. 12274–12282, Nov. 2015.

[13] S. Griffiths-Jones, A. Bateman, M. Marshall, A.

Khanna, and S. R. Eddy, 'Rfam: an RNA family database.', *Nucleic Acids Res.*, vol. 31, no. 1, pp. 439–441, Jan. 2003.

[14] A. O. Chiromatzo *et al.*, 'miRNApath: a database of miRNAs, target genes and metabolic pathways.', *Genet. Mol. Res.*, vol. 6, no. 4, pp. 859–865, Oct. 2007.

[15] E. A. Grice and J. A. Segre, 'The human microbiome: our second genome.', *Annu. Rev. Genomics Hum. Genet.*, vol. 13, pp. 151–170, 2012.

[16] A. L. Richards *et al.*, 'Gut Microbiota Has a Widespread and Modifiable Effect on Host Gene Regulation.', *mSystems*, vol. 4, no. 5, Sep. 2019.

[17] L. V Hooper, T. Midtvedt, and J. I. Gordon, 'How host-microbial interactions shape the nutrient environment of the mammalian intestine.', *Annu. Rev. Nutr.*, vol. 22, pp. 283–307, 2002.

[18] A. Shmaryahu, M. Carrasco, and P. D. T. Valenzuela, 'Prediction of bacterial microRNAs and possible targets in human cell transcriptome.', *J. Microbiol.*, vol. 52, no. 6, pp. 482–489, Jun. 2014.

[19] T. Yu *et al.*, 'Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy.', *Cell*, vol. 170, no. 3, pp. 548-563.e16, Jul. 2017.

[20] G. Stegmayer *et al.*, 'Predicting novel microRNA: a comprehensive comparison of machine learning approaches.', *Brief. Bioinform.*, vol. 20, no. 5, pp. 1607–1620, Sep. 2019.

[21] S. Griffiths-Jones, R. J. Grocock, S. van Dongen, A. Bateman, and A. J. Enright, 'miRBase: microRNA sequences, targets and gene nomenclature', *Nucleic Acids Res.*, vol. 34, no. Database issue, pp. D140–D144, Jan. 2006.

[22] N. Qin *et al.*, 'Alterations of the human gut microbiome in liver cirrhosis', *Nature*, vol. 513, no. 7516, pp. 59–64, Sep. 2014.

[23] M. Abdallah, A. Mahgoub, H. Ahmed, and S. Chaterji, 'Athena: Automated Tuning of k-mer based Genomic Error Correction Algorithms using Language Models', *Sci. Rep.*, vol. 9, no. 1, p. 16157, Nov. 2019.

[24] T. K. Ho, 'The random subspace method for constructing decision forests', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 832–844, 1998.

[25] D. J. Hand and K. Yu, 'Idiot's Bayes: Not So Stupid after All?', *Int. Stat. Rev. / Rev. Int. Stat.*, vol. 69, no. 3, pp. 385–398, Apr. 2001.

[26] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[27] N. S. Altman, 'An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression', *Am. Stat.*, vol. 46, no. 3, pp. 175–185, Apr. 1992.

[28] D. E. Wood and S. L. Salzberg, 'Kraken: ultrafast metagenomic sequence classification using exact alignments', *Genome Biol.*, vol. 15, no. 3, pp. R46–R46, Mar. 2014.