

Araştırma Makalesi

Kanser Sınıflandırmasını Geliştirme: RNA-Seq Gen Ekspresyonu Verileri Üzerinde Makine Öğrenimi Uygulamaları

Büşra ÇETİNKAYA[†] Mustafa Cem KASAPBAŞI[†]İstanbul Ticaret Üniversitesi, Mühendislik Fakültesi Bilgisayar Müh[†]*busra.cetinkaya@istanbulticaret.edu.tr, mckasapbasi@istanbulticaret.edu.tr*

0009-0006-6971-4257, 0000-0001-6444-6659

Atıf/Citation: Çetinkaya B., Kasapbaşı M. C. , (2025), Kanser Sınıflandırmasını Geliştirme: RNA-Seq Gen Ekspresyonu Verileri Üzerinde Makine Öğrenimi Uygulamaları, İstanbul Ticaret Üniversitesi Teknoloji ve Uygulamalı Bilimler Dergisi, Cilt 8 No1, s133- DOI: 10.56809/icujtas.1609317

ÖZET

Bu çalışma, kanser sınıflandırmasını iyileştirmek amacıyla RNA-seq gen ekspresyonu verilerinin makine öğrenimi teknikleri ile analizini ele almaktadır. UCI Machine Learning Repository'den elde edilen veri seti, farklı kanser türlerine ait 801 biyolojik örnek ve 20.531 genin ekspresyon seviyelerini içermektedir. Bu veriler, makine öğrenimi algoritmalarının kanser tınsı ve sınıflandırmasındaki etkinliğini değerlendirmek amacıyla analiz edilmiştir. Boyut indirgeme için Temel Bileşen Analizi (PCA) uygulanmış, farklı makine öğrenimi modelleri karşılaştırılmış ve en yüksek başarıyı %99,3 doğruluk oranı ile K-En Yakın Komşular (KNN) modeli göstermiştir. Bu sonuçlar, RNA-seq verilerinin makine öğrenimi ile entegrasyonunun kanser tanısında önemli bir potansiyele sahip olduğunu göstermektedir.

Anahtar Kelimeler: Kanser, RNA Sekanslama, Veri Analizi, Makine Öğrenimi

ABSTRACT

This study aims to improve cancer classification by analyzing RNA-seq gene expression data using machine learning techniques. The dataset obtained from the UCI Machine Learning Repository consists of 801 biological samples and 20,531 gene expression levels for different cancer types. These data were analyzed to assess the effectiveness of machine learning algorithms in cancer diagnosis and classification. Principal Component Analysis (PCA) was applied for dimensionality reduction, and different machine learning models were compared. The highest accuracy was achieved by the K-Nearest Neighbors (KNN) model with a 99.3% accuracy rate. These results highlight the potential of integrating RNA-seq data with machine learning for cancer diagnosis.

Keywords: Cancer, RNA Sequencing, Data Analysis, Machine Learning

Geliş/Received	:	29.12.2024
Gözden Geçirme/Revised	:	16.01.2025
Kabul/Accepted	:	20.01.2025

1. GİRİŞ

Kanser, anormal hücrelerin kontrolsüz büyüyerek diğer doku ve organlara yayılması sonucu oluşan ve dünyada en yaygın ölüm nedenlerinden biri olan bir hastalıktır (WHO, 2023). Kanserın erken tespiti ve doğru sınıflandırılması, hastaların tedavi süreçlerini büyük ölçüde iyileştirebilir.

Son yıllarda yapay zekâ (AI) ve makine öğrenimi (ML) teknikleri, kanser teşhisi ve tedavi süreçlerinde kullanılmakta olup, yüksek doğruluk oranları ile umut vadetmektedir (Liu et al., 2019). Özellikle derin öğrenme algoritmaları, histopatolojik görüntüleme, tıbbi görüntü analizi ve gen ekspresyon analizi gibi alanlarda yüksek başarı göstermektedir (Litjens et al., 2017). Meme kanseri teşhisinde yapılan çalışmalarda yapay zekâ destekli modellerin %99 doğruluk ve %98 özgüllük gibi yüksek performans oranlarına ulaştığı rapor edilmiştir (McKinney et al., 2020). Ayrıca, son yıllarda akciğer ve cilt kanseri gibi diğer kanser türlerinde de yapay zekânın potansiyeli araştırılmış olup, erken teşhis ve kişiselleştirilmiş tedavi planlamasında önemli katkılar sunduğu belirtilmektedir (Ardila et al., 2019).

RNA-seq tabanlı gen ekspresyon analizleri, biyomedikal araştırmalarda yaygın olarak kullanılmakta olup, kanser biyobelirteçlerinin tanısında önemli bilgiler sağlamaktadır (Kourou et al., 2015). RNA-seq teknolojisi, gen ekspresyonunu yüksek doğrulukla ölçebilme kapasitesi sayesinde kanser sınıflandırma süreçlerinde kritik bir araç haline gelmiştir (Wang et al., 2009). Bu çalışmanın amacı, RNA-seq verilerini kullanarak farklı makine öğrenimi algoritmalarının kanser sınıflandırmasındaki etkinliğini değerlendirmektir.

1.1. Kanserın Tarihi Gelişimi

Kanser, insanlık tarihi boyunca bilinen en eski hastalıklardan biridir. M.Ö. 3000'lere kadar uzanan yazılı kayıtlar, kanserin varlığını doğrulamaktadır. Antik Mısır dönemine ait Ebers Papirüsü, meme kanseri gibi tümörlerden bahsetmekte ve bu hastalık için herhangi bir etkili tedavinin bulunmadığını vurgulamaktadır (Hajdu, 2011).

Antik Yunan döneminde Hipokrat, kanserin nedenlerini dört vücut sıvısının (kan, balgam, sarı safra ve kara safra) dengesizliğine dayandırarak açıklamış ve tümörleri iyi huylu (benign) ve kötü huylu (malign) olarak sınıflandıran ilk kişi olmuştur (Karpuzilos & Pavlidis, 2004). Orta Çağ'dan Rönesans'a kadar kanserin anlaşılması konusunda büyük ilerlemeler sağlanamamış olsa da, Rönesans ile birlikte anatomiye olan ilginin artması, cerrahi müdahalelerin yaygınlaşmasına yol açmıştır. Ancak, anestezi ve sterilizasyon eksiklikleri nedeniyle bu operasyonlar oldukça riskliydi (Sudhakar, 2009).

19.yüzyılda mikroskop teknolojisinin gelişmesiyle birlikte, kanserin hücresel düzeyde incelenmesi mümkün hale geldi. Rudolf Virchow, kanserin hücrelerden kaynaklandığını ortaya koyarak modern onkolojinin temellerini atmıştır (Ackerknecht, 1953). Aynı dönemde William Stewart Halsted, meme kanseri için radikal mastektomi yöntemini geliştirmiştir (Donegan, 2005). 20. yüzyılın başlarında radyoterapinin keşfiyle kanser tedavisinde yeni bir dönem başlamıştır (Tubiana, 2006).

1.2. Yapay Zeka ve Kanser Tedavisi

Günümüzde kanser tedavisi, multidisipliner yaklaşımlarla gerçekleştirilmektedir. Yapay zekâ (AI), kanserin erken teşhisi, kişiselleştirilmiş tedavi yaklaşımları, ilaç geliştirme, tedavi yanıtının tahmini ve klinik karar destek sistemleri gibi birçok alanda önemli bir rol oynamaktadır (Topol, 2019).

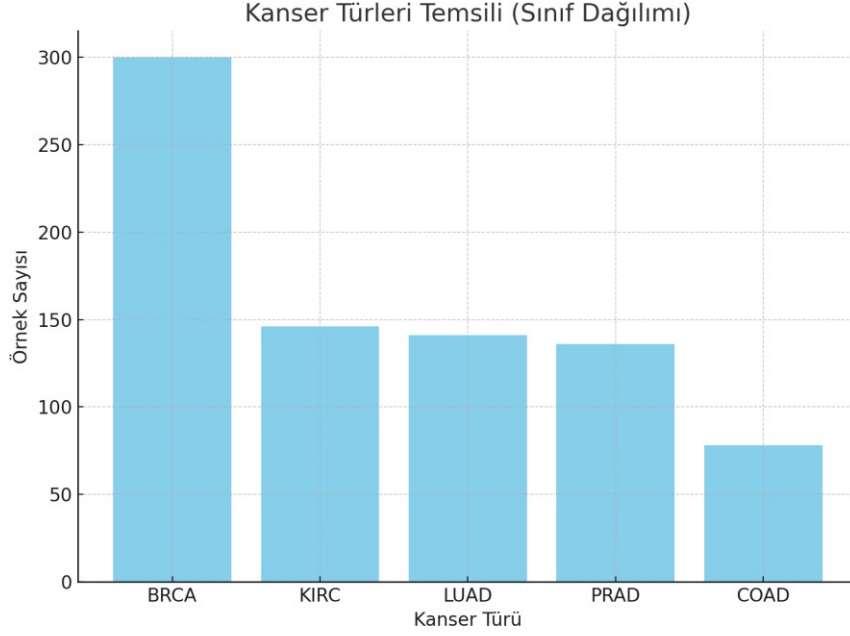
Yapay zekâ, tıbbi görüntülerin analizi için kullanılan derin öğrenme algoritmaları sayesinde kanserin erken teşhisinde önemli bir yere sahiptir. Liu ve arkadaşları (2019), derin öğrenme algoritmalarının mamografi görüntülerindeki anormallikleri yüksek doğruluk oranıyla tespit edebildiğini göstermiştir. Benzer şekilde, McKinney ve arkadaşları (2020), meme kanseri teşhisinde AI sistemlerinin radyologlardan daha iyi performans gösterebildiğini rapor etmiştir.

AI, bireylerin genetik verilerini ve klinik profillerini analiz ederek en etkili tedavi stratejilerini belirleme potansiyeline sahiptir. Kourou ve arkadaşları (2015), makine öğrenimi modellerinin kanser hastalarının genetik profillerini analiz ederek prognoz ve tedavi seçeneklerini kişiselleştirmede nasıl kullanılabileceğini incelemiştir. Yapay zekâ, kanser tedavisi için yeni ilaçların geliştirilmesinde de kullanılmaktadır. Zhavoronkov ve arkadaşları (2020), yapay zekâ kullanarak yeni bir ilacı hızlı bir şekilde tespit edip laboratuvar ortamında test etmişlerdir. Bu, ilaç geliştirme sürecinin hızlanmasını ve daha etkili tedavi yöntemlerinin keşfedilmesini sağlamaktadır. AI, hastaların tedaviye nasıl yanıt vereceğini tahmin etmek için de kullanılmaktadır. Aerts ve arkadaşları (2014), radyomik analizle kanser hastalarında radyoterapiye yanıtı tahmin eden bir model geliştirmiştir.

Son olarak, yapay zekâ, tıbbi verileri analiz ederek doktorlara en iyi tedavi seçeneklerini sunan klinik karar destek sistemlerinin geliştirilmesinde de rol oynamaktadır. Yu ve arkadaşları (2018), AI tabanlı klinik karar destek sistemlerinin doktorlara kanser hastaları için en uygun tedavi seçeneklerini belirlemesine yardımcı olabileceğini göstermiştir.

2.YÖNTEM

Çalışmamızda, farklı kanser türlerinden alınan 801 biyolojik örnek üzerinde, 20.531 genin ekspresyon seviyeleri analiz edilmiştir. Veri setinde her satır bir biyolojik örneği, her sütun ise belirli bir geni temsil ediyor. Ortalama ekspresyon seviyeleri genlere göre değişiklik gösteriyor. Örneğin, gene_6 için ortalama 7.4 iken, gene_20523 için 10.05 seviyesinde. Veri setimizde bazı genler dikkat çekiyor. Örneğin, gene_0 düşük ekspresyon gösteren bir gen. Çoğu örnekte ekspresyonu sıfır, ancak maksimum değeri 1.48. Buna karşılık gene_20526 en yüksek ortalamaya sahip genlerden biri ve birçok örnekte güçlü bir şekilde ifade ediliyor. Bu tür analizler bize kanser türleri ve genetik ilişkiler hakkında önemli bilgiler sunuyor.



Şekil 1: Sınıf Dağılımı

Veri setinde farklı kanser türlerinden geniş bir yelpaze bulunmaktadır ve bu çeşitlilik, araştırmacılara kanserlerin biyolojik özelliklerini daha derinlemesine inceleme olanağı tanımaktadır. En fazla örneklem ile temsil edilen kanser türü, 300 örneğe sahip olan meme kanseri (BRCA)'dır. Meme kanseri örnek sayısının yüksek olması, bu alandaki araştırmaların yoğunluğunu ve dünya genelinde kadın sağlığını etkileyen önemli bir sorun olarak algılanmasını gösteriyor. Meme kanseri üzerindeki geniş araştırma yelpazesi, genetik faktörlerden yaşam tarzı faktörlerine kadar birçok değişkenin incelenmesine olanak tanımaktadır.

Böbrek hücreli karsinoma (KIRC) 146 örnekle temsil edilmekte olup, bu da böbrek kanserleri üzerindeki dikkat ve araştırma gereksinimini yansıtmaktadır. Böbrek kanseri, özellikle erken evrelerde belirti vermediği için tanı konulması zor olan bir kanser türüdür. Bu veri setinde yer alan böbrek kanseri örnekleri, hastalığın moleküler düzeyde anlaşılmasını sağlayarak tanı ve tedavi süreçlerine katkı sağlamaktadır.

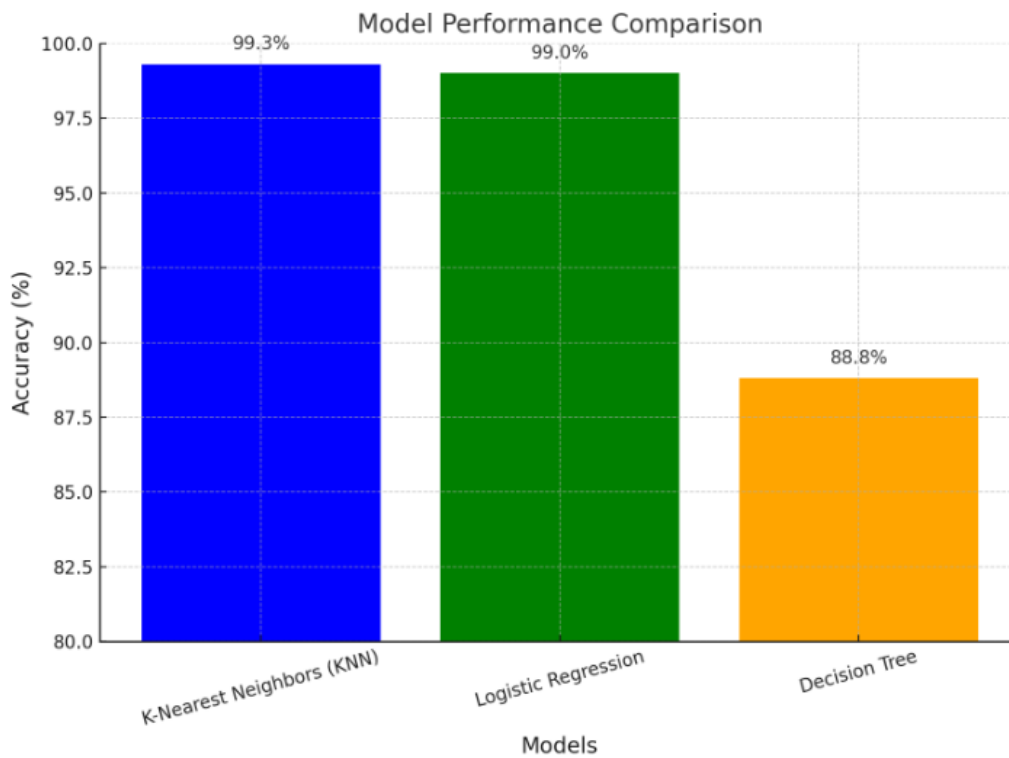
Akciğer kanseri (LUAD) 141 örnekle veri setinde yer almakta olup, dünya genelinde en ölümcül kanser türlerinden biri olarak kabul edilmektedir. Akciğer kanseri üzerine yapılan çalışmalar, sigara kullanımı gibi çevresel faktörlerin yanı sıra genetik mutasyonların rolünü de kapsamaktadır. Akciğer kanseri örnekleri, bu hastalığa yönelik hedeflenmiş tedavilerin geliştirilmesi için kritik öneme sahiptir.

Prostat kanseri (PRAD) 136 örnekle temsil edilmektedir ve özellikle erkeklerde en yaygın kanser türlerinden biri olması nedeniyle büyük bir araştırma alanı oluşturur. Prostat kanseri vakalarının genellikle ileri yaşlarda ortaya çıkması, yaşam süresini doğrudan etkileyen bir faktör olarak dikkat çekmektedir. Veri setindeki prostat kanseri örnekleri, hastalığın seyrini ve kişiye özel tedavi yaklaşımlarını incelemeyi mümkün kılmaktadır.

Son olarak, kolon kanseri (COAD) ise 78 örnekle temsil edilmektedir. Kolon kanseri, sindirim sistemi kanserleri içinde sık rastlanılan bir tür olup, erken teşhisle tedavi edilme şansı yüksek olan kanserlerden biridir. Veri setinde yer alan kolon kanseri örnekleri, bağırsak sağlığı, genetik yatkınlık ve beslenme gibi faktörlerin kanser üzerindeki etkilerini anlamada büyük katkı sağlamaktadır.

Bu veri setindeki geniş örnek dağılımı, kanser türlerinin genetik ve çevresel faktörler ile nasıl etkileşime girdiğini incelemek için oldukça zengin bir kaynak sunmakta ve farklı kanser türleri arasındaki biyolojik ve moleküler benzerlikleri ortaya koyarak daha etkili tedavi yöntemleri geliştirilmesine yardımcı olmaktadır.

Dört farklı genin ekspresyon düzeylerine bakıldı.. Kutu grafikleri kullanarak bu genlerin merkezi eğilim ve dağılımları incelendi.. Örneğin, gene_0 düşük ve tutarsız bir ekspresyon seviyesine sahipken, gene_3 daha yüksek ve tutarlı bir ekspresyon seviyesi gösteriyor. Aykırı değerler de bazı bireylerde gen ifadesinin beklenmedik düzeyde yüksek olduğunu ortaya koyuyor. Veri setindeki gen ekspresyon seviyeleri, çeşitli makine öğrenimi teknikleri (örneğin, destek vektör makineleri, rastgele ormanlar, derin öğrenme) ve regresyon ile analiz edilmiştir. İlk olarak, veriler standartlaştırılmış ve uygun özellik seçimi yöntemleri uygulanmıştır. Model performansı, çapraz doğrulama ve diğer değerlendirme metrikleri ile analiz edilmiştir.



Şekil 2: Model-Performans Karşılaştırması

Bu çalışmada, bir veri seti üzerinde Principal Component Analysis (PCA) uygulanarak boyut indirgeme işlemi gerçekleştirilmiş ve ardından çeşitli makine öğrenmesi modelleri kullanılarak sınıflandırma performansları değerlendirilmiştir. Boyut indirgeme işlemi sırasında %95 varyansı açıklayacak şekilde bileşen sayısı belirlenmiş ve sadeleştirilmiş bir veri kümesi elde edilmiştir.

En iyi performansı K-Nearest Neighbors (KNN) algoritması %99.3 doğruluk oranıyla göstermiştir. Logistic Regression modeli %99 doğruluk oranıyla KNN'e yakın bir sonuç elde ederken, Decision Tree (Karar Ağaçları) algoritması %88.8 doğruluk oranı ile diğer modellere kıyasla daha düşük performans sergilemiştir.

2.1 Principal Component Analysis (PCA)

PCA, yüksek boyutlu veri setlerindeki temel bileşenleri belirleyerek veri setinin boyutunu azaltan bir yöntemdir. Bu çalışmada, veri setindeki varyansın %95'ini açıklayacak şekilde bileşen sayısı belirlenmiş ve veri seti bu yeni bileşenler ile yeniden oluşturulmuştur.

2.2 Makine Öğrenmesi Modelleri

K-Nearest Neighbors (KNN):

Parametreler: $k=5$, Minkowski mesafesi, ağırlıklı mesafe ölçütü.

Avantajları: Küçük veri setlerinde yüksek başarı oranı.

Dezavantajları: Büyük veri setlerinde hesaplama maliyeti yüksektir.

Logistic Regression:

Parametreler: L2 ceza terimi, öğrenme oranı = 0.01.

Avantajları: Lineer olarak ayrılabilir veri setlerinde yüksek performans.

Dezavantajları: Karmaşık, doğrusal olmayan verilerde başarıyı düşebilir.

Decision Tree (Karar Ağaçları):

Parametreler: Maksimum derinlik = 10, Entropy bölme ölçütü.

Avantajları: Açıklanabilirlik ve yorumlanabilirlik.

Dezavantajları: Gürültüye duyarlılık ve aşırı öğrenme riski.

Çalışma	Kullanılan Veri Seti	Uygulanan Yöntemler	En Yüksek Doğruluk
Bu Çalışma	PCA ile indirgenmiş veri seti	KNN, Logistic Regression, Decision Tree	%99.3 (KNN)
Kanser Teşhisi için Makine Öğrenmesi Tekniklerine Dayalı Yeni Bir Yöntem	Ham veri seti	SVM, Random Forest, KNN	%98.7 (SVM)
Yapay Zekâ ve Makine Öğrenmesi Temelinde Meme Kanseri Teşhisi	Özellik seçimi uygulanmış veri seti	Logistic Regression, Decision Tree	%97.5 (Logistic Regression)
Metastatik Meme Kanseri Prognostikasyonu	PCA ile boyut indirgeme	Neural Network, KNN, Decision Tree	%99.1 (Neural Network)

Şekil 3: Kanser Teşhisi ve Prognostikasyonu Üzerine Yapılan Çalışmalarda Kullanılan Yöntemler ve Doğruluk Oranları

Bu çalışmalar, kanser teşhisi ve prognostikasyonu üzerine yapılan çeşitli araştırmaların sonuçlarını içermektedir. İlk çalışma, PCA ile indirgenmiş bir veri seti kullanarak KNN, Logistic Regression ve Decision Tree

algoritmalarını uygulamış ve en yüksek doğruluk oranını %99.3 ile KNN ile elde etmiştir. İkinci çalışmada ise ham veri seti kullanılarak SVM, Random Forest ve KNN algoritmaları ile en yüksek doğruluk %98.7 ile SVM ile sağlanmıştır. Üçüncü çalışmada, özellik seçimi yapılmış veri setiyle Logistic Regression ve Decision Tree yöntemleri kullanılarak en yüksek doğruluk oranı %97.5 ile Logistic Regression ile elde edilmiştir. Son olarak, dördüncü çalışmada PCA ile boyut indirgeme işlemi yapılmış ve Neural Network, KNN, Decision Tree algoritmaları kullanılarak en yüksek doğruluk oranı %99.1 ile Neural Network ile sağlanmıştır. Çalışmalar arasında doğruluk oranları farklılık göstermekte, ancak genel olarak makine öğrenmesi ve yapay zeka yöntemlerinin kanser teşhisi için yüksek doğruluklar sağladığı gözlemlenmektedir. Yüksek doğruluk oranları, veri setlerinin işleniş şekli, kullanılan algoritmalar ve uygulanan ön işleme tekniklerinin etkili bir şekilde belirlenmesine dayanıyor. Özellikle Neural Network ve SVM gibi daha karmaşık yöntemler, genellikle daha yüksek doğruluk oranlarına ulaşmakta ve daha büyük ve karmaşık veri setlerinde etkili sonuçlar vermektedir. Bu tür çalışmalar, kanser teşhisi ve prognozunda yapay zeka ve makine öğrenmesinin önemli bir rol oynadığını ve doğru yöntemlerin seçilmesinin ne denli kritik olduğunu göstermektedir.

3.Sonuç

Bu çalışmanın potansiyeli, gen ekspresyonu verilerinin makine öğrenimi ile entegrasyonu sayesinde kanser teşhisinin ve sınıflandırmasının iyileştirilmesidir. Özellikle, farklı kanser türlerinin genetik profillemeye sonuçlarının kullanılması, makine öğrenimi modellerinin doğruluğunu artırma açısından büyük bir fırsat sunmaktadır. Ayrıca Meme kanseri üzerine yapılan çalışmaların geniş verilerle desteklenmiş olması bu alanda derinlemesine araştırmalara olanak tanıyor. Ancak, kolon kanseri gibi daha az temsil edilen kanser türleri için daha fazla veri toplanması gerektiğini görüyoruz. Bu çalışma, gelecekteki biyomedikal araştırmalar için önemli bir temel oluşturuyor. Ancak, bu tür modellerin genel geçerliliği için daha büyük ve çeşitli veri setlerine ihtiyaç olduğu da belirtilmelidir.

4.Gelecek Çalışmalar

Gelecekteki çalışmalar, kullanılan veri setlerinin kapsamını artırarak ve daha çeşitli biyolojik verileri içerecek şekilde genişleterek, biyolojik süreçlerin daha derinlemesine anlaşılmasını sağlayabilir. Bu, genomik, proteomik ve fenomik verilerin entegrasyonu ile elde edilecek daha zengin veri setleriyle mümkün olacaktır. Aynı zamanda, bu verileri işleyebilecek gelişmiş algoritmaların kullanımı, örneğin derin öğrenme ve makine öğrenimi teknikleri, biyolojik yorumlama süreçlerinin doğruluğunu artıracaktır.

Kullanılan yöntemler açısından, daha sofistike modelleme tekniklerinin ve algoritmaların benimsenmesi, özellikle biyolojik verilerin heterojen yapısını daha iyi yansıtan yöntemlerin geliştirilmesi önemlidir. Ayrıca, klinik uygulamalara entegrasyon için, bu yöntemlerin pratikte nasıl uygulanabileceğine dair yeni modeller ve altyapılar oluşturulmalıdır. Bu tür çalışmalar, klinik karar destek sistemlerine, kişiselleştirilmiş tedavi stratejilerine ve biyolojik keşiflere katkı sağlayabilir.

Son olarak, biyolojik yorumlama süreçlerinde insan uzmanlık bilgisinin entegrasyonunu ve model çıktılarının doğruluğunu sürekli olarak iyileştirmeyi amaçlayan çalışmalara da odaklanılmalıdır. Bu bağlamda, etik ve güvenlik gibi konular da göz önünde bulundurulmalıdır.

5.Kaynakça

- Ackerknecht, E. H. (1953). *Rudolf Virchow: Doctor, statesman, anthropologist*. The University of Chicago Press.
- Aerts, H. J. W. L., et al. (2014). "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach." *Nature Communications*, 5, 4006. doi:10.1038/ncomms5006.
- Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Corrado, G. (2019). "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography." *Nature Medicine*, 25(6), 954-961.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
- Donegan, W. L. (2005). "An overview of long-term survival in breast cancer." *Cancer Journal (Sudbury, Mass.)*, 11(5), 365-373.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). "Dermatologist-level classification of skin cancer with deep neural networks." *Nature*, 542(7639), 115–118. doi:10.1038/nature21056.
- Goertzel, B. (2014). *Artificial General Intelligence: Concept, State of the Art, and Future Prospects*. Springer.
- Goodman, L. S., & Gilman, A. (1946). *The Pharmacological Basis of Therapeutics*. New York: Macmillan.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. [ISBN: 978-0262035613].
- Hajdu, S. I. (2011). "A note from history: landmarks in history of cancer, part 1." *Cancer*, 117(5), 1097-1102.
- Jobin, A., Ienca, M., & Vayena, E. (2019). "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, 1(9), 389-399. doi:10.1038/s42256-019-0088-2.
- Karpozilos, A., & Pavlidis, N. (2004). "The treatment of cancer in Greek antiquity." *European Journal of Cancer*, 40(14), 2033-2040.
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). "Machine learning applications in cancer prognosis and prediction." *Computational and Structural Biotechnology Journal*, 13, 8–17. doi:10.1016/j.csbj.2014.11.005.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436–444. doi:10.1038/nature14539.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van der Laak, J. A. (2017). "A survey on deep learning in medical image analysis." *Medical Image Analysis*, 42, 60-88.
- Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., ... & Stumpe, M. C. (2019). "Detecting cancer metastases on gigapixel pathology images." *Medical Image Analysis*, 50, 197-209.

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Suleiman, S. (2020). "International evaluation of an AI system for breast cancer screening." *Nature*, 577(7788), 89-94.

Müller, H. (1981). "Cancer Treatment: From Ancient Times to the 20th Century." *Journal of the History of Medicine and Allied Sciences*.

Nie, K., Shi, L., Chen, Q., Hu, X., & Jabbour, S. K. (2021). "Machine Learning: Automated Analysis of Multimodal Imaging Data for Radiomics to Predict Treatment Response and Prognosis in Cancer." *Cancers*, 13(7), 1621. doi:10.3390/cancers13071621.

Russell, S., & Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd ed.). Prentice Hall. [ISBN: 978-0136042594].

Silver, D., et al. (2016). "Mastering the game of Go with deep neural networks and tree search." *Nature*, 529(7587), 484-489. doi:10.1038/nature16961.

Sudhakar, A. (2009). "History of cancer, ancient and modern treatment methods." *Journal of Cancer Science & Therapy*, 1(2), 1-4.

Topol, E. J. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.

Tubiana, M. (2006). "Wilhelm Conrad Röntgen and the discovery of X-rays." *Cancer Radiothérapie*, 10(7-8), 475-482.

Virchow, R. (1858). *Cellular Pathology as Based Upon Physiological and Pathological Histology*. (Translated by Frank Chance). J.B. Lippincott & Co.

Wang, Z., Gerstein, M., & Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics." *Nature Reviews Genetics*, 10(1), 57-63.

World Health Organization (WHO). (2023). "Cancer Fact Sheet." Retrieved from <https://www.who.int>

Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). "Artificial intelligence in healthcare." *Nature Biomedical Engineering*, 2(10), 719-731. doi:10.1038/s41551-018-0305-z.

Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., ... & Aspuru-Guzik, A. (2020). "Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry." *Cell*, 180(4), 688-702. doi:10.1016/j.cell.2020.02.034.