

**2023 -2024 BAHAR DÖNEMİ BİYOENFORMATİK DERSİ DÖNEM PROJESİ**

**GÖĞÜS KANSERİ GENLERİNİN VERİ ANALİZİ VE MAKİNE ÖĞRENİMİ  
UYGULAMASI**

**Bilgisayar Mühendisliği Bölümü**

**MAYIS 2024**

## 1. Giriş

Göğüs kanseri, kadınlar arasında en yaygın görülen ve en fazla ölüme neden olan kanser türlerinden biridir. Bu hastalığın moleküler mekanizmalarının daha iyi anlaşılması ve etkili tedavi stratejilerinin geliştirilmesi, biyoinformatik ve genomik verilerin kapsamlı analizini gerektirmektedir. Geniş ölçekli gen ekspresyon veri setleri, göğüs kanserinin alt tiplerini, prognozunu ve tedaviye yanıtını anlamada kritik bir rol oynamaktadır. Bu bağlamda, çeşitli veri setlerinin birleştirilmesi ve bu veriler üzerinde ileri analizler yapılması, daha sağlam ve genellenebilir sonuçlar elde edilmesine katkıda bulunabilir.

Çalışmada, göğüs kanseriyle ilgili iki geniş ölçekli gen ekspresyon veri seti olan GSE25066 ve GSE3494 birleştirilmiştir. Her iki veri seti, farklı hasta gruplarından elde edilen gen ekspresyon profillerini içermekte olup, birleştirilmiş veri seti daha geniş bir örneklem büyüklüğüne ve çeşitliliğe sahiptir. Veri setlerinin birleştirilmesi, ortak genler üzerinden yapılmış ve birleştirilmiş veri seti üzerinde özellik seçimi, sınıflandırma ve anlamlı genlerin bulunması işlemleri gerçekleştirilmiştir. Bu analizlerin temel amacı, göğüs kanseri ile ilişkili kritik genleri ve biyolojik yolları belirlemek, bu sayede hastalığın moleküler temellerini daha iyi anlamak ve potansiyel biyomarkerlar ve terapötik hedefler tanımlamaktır.

Analizler Python programlama dili kullanılarak gerçekleştirilmiştir. Python'ın biyoinformatik ve makine öğrenimi alanlarındaki güçlü kütüphaneleri (NumPy, pandas, scikit-learn, SciPy, ve matplotlib gibi) kullanılarak verilerin işlenmesi, analiz edilmesi ve görselleştirilmesi sağlanmıştır. Birleştirilmiş veri seti üzerinde yapılan özellik seçimi işlemi, yüksek boyutlu gen ekspresyon verilerinden en anlamlı genlerin belirlenmesini sağlamış ve bu genler sınıflandırma modellerinin performansını artırmada kullanılmıştır. Sınıflandırma işlemi ile göğüs kanseri alt tiplerinin doğru bir şekilde ayrılması hedeflenmiş ve anlamlı genlerin belirlenmesi ile hastalığın biyolojik süreçlerine ışık tutulmuştur. Bu çalışma, biyoinformatik yaklaşımların ve makine öğrenimi yöntemlerinin birleştirilmiş gen ekspresyon verileri üzerindeki potansiyelini göstermeyi amaçlamaktadır.

## 2. Göğüs Kanseri Nedir?

Göğüs kanseri, meme dokusunda başlayan bir kanser türüdür. Kanser, vücudun bir bölümünde kontrolsüz hücre büyümesi olduğunda oluşur. Göğüs kanseri genellikle lobüler karsinom ve duktal karsinom olmak üzere iki ana türde bulunur. Lobüler karsinom, memenin süt üreten lobüllerinde başlar, duktal karsinom ise sütü meme ucuna taşıyan kanallarda başlar.

Göğüs kanseri, genetik mutasyonlar sonucu oluşur. Bu mutasyonlar, hücrelerin normal büyüme ve ölüm döngüsünü bozar. Mutasyonlar, hücrelerin hızla bölünmesine ve ölmemesine neden olur, bu da hücrelerin birikmesine ve bir tümör oluşturmaya yol açar.

Göğüs kanseri, ER pozitif (östrojen reseptörü pozitif), PR pozitif (progesteron reseptörü pozitif) ve HER2 pozitif (human epidermal growth factor reseptör 2 pozitif) olmak üzere üç ana biyolojik alt tipe ayrılır. Bu alt tipler, kanserin hangi tür hormonlara veya proteinlere duyarlı olduğunu belirler ve bu da tedavi seçeneklerini etkiler.

Göğüs kanserinin tedavisi genellikle cerrahi, radyasyon, kemoterapi, hormonal tedavi ve hedefe yönelik tedavileri içerir. Tedavi seçenekleri, kanserin evresi, alt tipi ve hastanın genel sağlık durumuna bağlıdır.

### **3. Kullanılan Veri Setleri**

GSE25066 ve GSE3494 veri setleri, göğüs kanseri araştırmalarında önemli veri kaynaklarıdır. Her iki veri seti de gen ifadesi profilini içerir ve göğüs kanseri hastalarının moleküler özelliklerini anlamak, hastalığın gelişimine katkıda bulunan genleri tanımlamak ve tedavi stratejileri geliştirmek için kullanılmıştır.

Her iki veri seti de gen ifade verilerini içerir ve meme kanseri üzerine yapılan çalışmalarda kullanılır. Her iki veri seti de National Center for Biotechnology Information (NCBI) tarafından sağlanan Gene Expression Omnibus (GEO) veritabanında bulunabilir.

Her iki veri seti de göğüs kanseri araştırmalarında geniş kapsamlı genomik veriler sağlayarak hastalığın moleküler temellerini anlamak ve daha etkili tedavi stratejileri geliştirmek için önemli bir kaynak oluşturur. Araştırmacılar, bu veri setlerini analiz ederek hastalığın biyolojik süreçlerini, prognoz faktörlerini ve potansiyel tedavi hedeflerini belirlemede büyük ilerlemeler kaydetmişlerdir.

**GSE25066:** Veri seti, neoadjuvant takzan-antrasiklin kemoterapisine yanıt ve hayatta kalma süresini tahmin eden genomik bir belirleyici üzerine odaklanmaktadır. Bu çalışma, meme kanseri hastalarında kullanılan bu tedaviye yanıtı ve sonuçları tahmin etmek için gen ifade profillerini kullanır. Veri seti, 2010 yılında gönderilmiş ve 2011 yılında kamuoyuna açılmıştır ve Homo sapiens (insan) türü üzerinde ifade profili oluşturma deneyi olarak sınıflandırılmıştır. Bu çalışma, meme kanseri hastalarında kullanılan bu tedaviye yanıtı ve sonuçları tahmin etmek

için gen ifade profillerini kullanır. Bu veri seti, aşağıdaki alt serilerden oluşur: GSE25055 ve GSE250651. Veri setinde 22283 gen ve toplamda 508 örnek bulunmaktadır.

**GSE3494:** Bu veri seti, meme kanserinde p53 ifade imzasını inceler ve mutasyon durumunu, transkriptomal etkileri ve hasta hayatta kalma süresini tahmin eder. Veri seti, 2005 yılında kamuoyuna açılmıştır ve Homo sapiens (insan) türü üzerinde ifade profili oluşturma deneyi olarak sınıflandırılmıştır. Bu çalışma, İsveç'teki Uppsala İlçesi'nde 1987 ile 1989 yılları arasında rezeke edilen tüm meme kanserlerinin %65'ini temsil eden 315 kadından alınan taze donmuş meme tümörlerini içerir. Veri setindeki her bir örnekte, p53 durumu, ER durumu, tümör derecesi, lenf düğümü durumu ve hasta yaşı gibi klinikopatolojik özellikler bulunmaktadır. Veri setinde 44760 gen ve toplamda 502 örnek bulunmaktadır.

#### **4. Birleştirilmiş Veri Setinin Yorumlanması**

Göğüs kanseri araştırmalarında kullanılan GSE25066 ve GSE3494 veri setleri, farklı gen ekspresyon profillerini içermektedir. GSE25066 veri seti 22,283 gen ve 508 örnek içerirken, GSE3494 veri seti 44,760 gen ve 502 örnek içermektedir. Bu iki veri setinin ortak genlere göre birleştirilmesi sonucunda elde edilen birleştirilmiş veri seti, 3,056 gen ve 1,010 örnek içermektedir. Bu, iki veri setinin genomik kapsamı arasında bazı farklılıklar olduğunu ve her iki veri setinde de yer alan genlerin birleştirildiğini göstermektedir.

Birleştirilmiş veri setinin gen sayısının 3,056'ya düşmesi, her iki veri setinde de yer alan ortak genlerin analiz için seçildiğini ve sadece bu genlerin kullanıldığını belirtir. Bu durum, verilerin tutarlılığını artırmak ve analizlerin güvenilirliğini sağlamak için önemlidir. Ortak genlerin kullanılması, veri setlerinin birleşiminde tutarlı ve karşılaştırılabilir sonuçlar elde edilmesine olanak tanır.

Örnek sayısının artması, analiz gücünü ve genellenebilirliği artırmaktadır. 1,010 örnek, göğüs kanseri alt tiplerinin belirlenmesi, prognostik ve diagnostik modellerin geliştirilmesi için daha geniş bir veri havuzu sunar. Daha fazla örnek, istatistiksel analizlerin gücünü artırır ve elde edilen sonuçların daha güvenilir olmasını sağlar.

Birleştirilmiş veri setindeki 3,056 gen ve 1,010 örnek, genişletilmiş bir analiz kapasitesi ve daha güvenilir sonuçlar elde etmek için önemli bir temel sağlar. Bu birleşim, göğüs kanseri ile ilgili kritik genlerin belirlenmesi, biyolojik yolların anlaşılması ve potansiyel

biyomarkerların tanımlanması için daha kapsamlı bir veri seti sunar. Sonuç olarak, bu birleştirilmiş veri seti, göğüs kanseri araştırmalarında önemli bir kaynak olarak hizmet etmekte ve hastalığın moleküler temellerini anlamada önemli katkılar sağlamaktadır.

## 5. Özellik Seçimi: Tek Değişkenli İstatistikler Kullanarak En İyi 50 Genin Seçilmesi

Göğüs kanseri gen ekspresyon verilerinin analizi sırasında, yüksek boyutlu veri setleriyle çalışmak genellikle zorluklar yaratır. Özellikle gen sayısının fazla ve örnek sayısının nispeten daha az olduğu durumlarda, analizlerin etkinliği ve doğruluğu düşebilir. Bu nedenle, özellik seçimi (feature selection) adımı kritik bir öneme sahiptir.

Özellik seçimi, veri setindeki en önemli ve anlamlı genlerin belirlenmesi sürecidir. Bu süreç, analizlerin doğruluğunu artırmak, hesaplama maliyetlerini azaltmak ve modelin aşırı öğrenme (overfitting) riskini azaltmak amacıyla gerçekleştirilir. Göğüs kanseri gibi kompleks hastalıklarda, tüm genlerin analize dahil edilmesi gereksiz ve yanıltıcı olabilir. Bu yüzden, sadece en belirleyici genlerin seçilmesi, modelin performansını iyileştirir ve biyolojik anlamlılık sağlar.

Bu çalışmada, özellik seçimi için tek değişkenli istatistiksel yöntemler kullanılmıştır. Spesifik olarak, SelectKBest sınıfı ve f\_classif fonksiyonu kullanılarak ANOVA (Analysis of Variance) testleri uygulanmıştır. Bu yöntem, her genin bağımsız olarak hedef değişken (sınıf etiketi) ile olan ilişkisini değerlendirir ve en anlamlı genleri seçer.

Adımlar şu şekilde gerçekleştirilmiştir:

- Eksik Değerlerin Tamamlanması: Eksik gen ekspresyon verileri, ortanca (median) değeri kullanılarak tamamlanmıştır (SimpleImputer kullanılarak).
- Özellik Seçimi: SelectKBest sınıfı kullanılarak, ANOVA F-testlerine dayalı olarak en iyi 50 gen seçilmiştir (f\_classif fonksiyonu ile).
- Seçilen Genlerin İsimlerinin Alınması: Seçilen genlerin isimleri, veri setindeki orijinal gen isimleriyle eşleştirilmiştir.

Bu özellik seçimi işlemi sonucunda, göğüs kanseri sınıflandırmasında en anlamlı 50 gen belirlenmiştir. Bu genler, hedef değişken (kanseri tipi veya alt tipi) ile en güçlü ilişkiye sahip genlerdir. Özellik seçimi sonucunda seçilen genler şu şekilde sıralanmıştır:

	Seçilen Genler
1	201123_s_at
2	202086_at
3	202089_s_at
4	202388_at
5	202648_at
6	202917_s_at
7	203021_at
8	203535_at
9	203936_s_at
10	204533_at
11	204563_at
12	204580_at
13	204959_at
14	205347_s_at
15	205483_s_at
16	205922_at
17	206962_x_at
18	207076_s_at
19	208894_at
20	209116_x_at
21	209604_s_at
22	209875_s_at
23	209969_s_at
24	210735_s_at
25	210982_s_at

	Seçilen Genler
26	211000_s_at
27	211372_s_at
28	211565_at
29	211699_x_at
30	211991_s_at
31	212003_at
32	212587_s_at
33	212588_at
34	213539_at
35	213975_s_at
36	214290_s_at
37	214414_x_at
38	215891_s_at
39	216207_x_at
40	216913_s_at
41	217517_x_at
42	218211_s_at
43	AFFX-HUMISGF3A/M97935_5_at
44	AFFX-HUMRGE/M10098_3_at
45	AFFX-HUMRGE/M10098_5_at
46	AFFX-HUMRGE/M10098_M_at
47	AFFX-r2-Hs18SrRNA-3_s_at
48	AFFX-r2-Hs18SrRNA-5_at
49	AFFX-r2-Hs18SrRNA-M_x_at
50	AFFX-r2-Hs28SrRNA-M_at

Bu seçilen genler, modelin eğitiminde ve testinde kullanılacak ve göğüs kanseri sınıflandırmasının doğruluğunu artıracaktır. Bu genlerin biyolojik anlamı da araştırılarak, göğüs kanserinin moleküler mekanizmalarının daha iyi anlaşılmasına katkıda bulunabilir. Sonuç olarak, özellik seçimi işlemi, modelin performansını artırmakla kalmayıp, aynı zamanda göğüs kanseri ile ilişkili kritik biyomarkerların belirlenmesine de olanak tanır. Bu, hem teşhis hem de tedavi süreçlerinde önemli ilerlemeler sağlayabilir.

## 6. Sınıflandırma: Karar Ağacı, Rastgele Orman ve Destek Vektör Makineleri ile Karşılaştırma ve Yorumlama

Göğüs kanseri gen ekspresyon verileri üzerindeki sınıflandırma işlemi, hastalığın alt tiplerinin doğru bir şekilde ayrılmasını sağlamak ve tanı süreçlerinde kullanılabilecek modeller geliştirmek amacıyla yapılmıştır. Sınıflandırma işlemi, farklı makine öğrenimi algoritmaları

kullanılarak gerekleřtirilmiř ve bu algoritmaların performansı doęruluk (accuracy) ve karıřıklık matrisleri (confusion matrices) ile kıyaslanmıřtır.

#### Kullanılan Sınıflandırma Yöntemleri

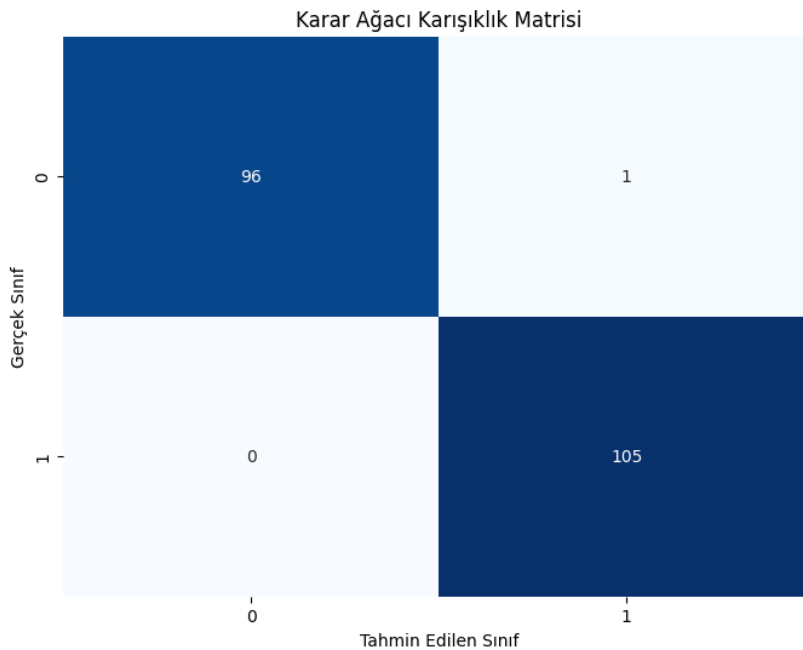
- Karar Ağacı (Decision Tree)
- Rastgele Orman (Random Forest)
- Destek Vektör Makineleri (Support Vector Machines, SVM)

#### Karar Ağacı (Decision Tree) Modeli

Karar ağacı modeli, veri setindeki örnekleri sınıflandırmak için ağaç yapısında karar kuralları kullanır. Bu modelin avantajı, yorumlanabilir ve anlaşılabilir olmasıdır. Ancak, tek bir karar ağacı modelinin aşırı öğrenme (overfitting) yapma riski bulunmaktadır.

Doęruluk: %99.50

Karıřıklık Matrisi:

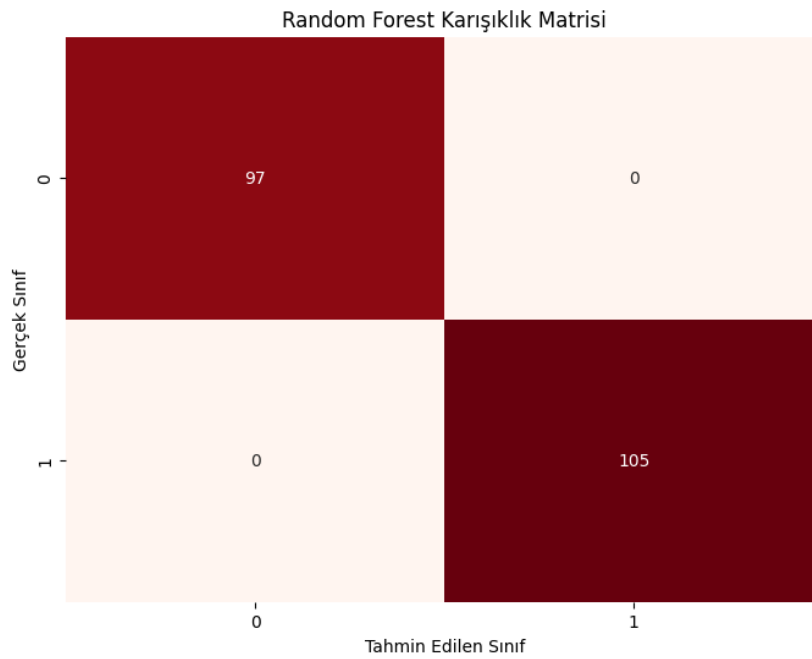


#### Rastgele Orman (Random Forest) Modeli

Rastgele orman modeli, birden fazla karar ağacını bir araya getirerek toplu (ensemble) bir sınıflandırma yapar. Bu yöntem, aşırı öğrenme riskini azaltır ve genellikle daha yüksek doęruluk sağlar.

Doğruluk: %100.00

Karışıklık Matrisi:

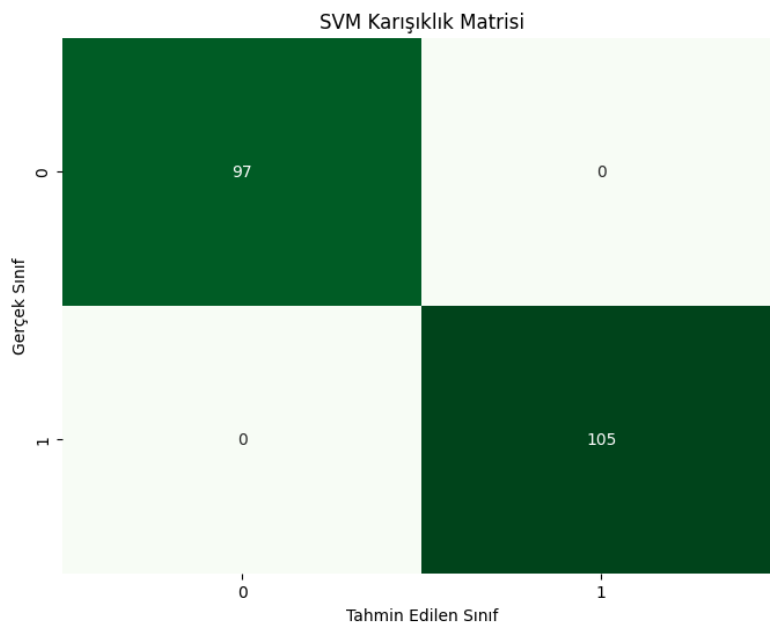


### Destek Vektör Makineleri (Support Vector Machines, SVM) Modeli

SVM modeli, veri noktalarını sınıflandırmak için yüksek boyutlu bir uzayda en iyi ayırım çizgisini (hiper düzlemi) bulur. SVM, özellikle yüksek boyutlu veri setlerinde etkili sonuçlar verir.

Doğruluk: %100.00

Karışıklık Matrisi:





## Yöntemlerin Karşılaştırılması

Model	Doğruluk (%)	Karışıklık Matrisi
Karar Ağacı	99.50	[[ 96 1] [ 0 105]]
Rastgele Orman	100.00	[[ 97 0] [ 0 105]]
Destek Vektör Makineleri	100.00	[[ 97 0] [ 0 105]]

**Doğruluk:** Rastgele orman ve SVM modelleri, %100 doğruluk oranı ile en yüksek performansı göstermiştir. Bu modeller, test veri setindeki tüm örnekleri doğru sınıflandırmıştır. Karar ağacı modeli ise %99.50 doğruluk oranı ile biraz daha düşük performans göstermiştir.

**Karışıklık Matrisi:** Rastgele orman ve SVM modelleri, karışıklık matrislerinde hiçbir yanlış sınıflandırma yapmamış, tüm pozitif ve negatif örnekleri doğru bir şekilde sınıflandırmıştır. Karar ağacı modeli ise bir adet yanlış pozitif sınıflandırma yapmıştır.

Bu sınıflandırma modelleri, göğüs kanseri veri seti üzerinde yüksek doğruluk oranlarına ulaşmıştır. Rastgele orman ve SVM modelleri, daha yüksek doğruluk ve daha iyi genellenebilirlik sunarak, klinik uygulamalar için daha güvenilir ve robust modeller olarak öne çıkmaktadır. Karar ağacı modeli ise, daha kolay yorumlanabilir olması nedeniyle bazı durumlarda tercih edilebilir, ancak doğruluk açısından diğer iki modelin gerisinde kalmaktadır.

Bu sonuçlar, göğüs kanseri teşhisinde ve alt tiplerinin belirlenmesinde makine öğrenimi yöntemlerinin etkili bir şekilde kullanılabileceğini göstermektedir. Bu modeller, hastalığın moleküler temellerini daha iyi anlamaya ve klinik karar destek sistemlerinde kullanılabilecek potansiyel araçlar sunmaya yardımcı olabilir.

## 7. Genel Değerlendirme (Sonuç)

Bu çalışma sonucunda, göğüs kanseri gen ekspresyon verilerinin birleştirilmesi ve analiz edilmesi süreçlerinde önemli ilerlemeler kaydedilmiştir. Özellikle, seçilen 50 genin göğüs kanseri teşhisinde ve alt tiplerinin belirlenmesinde kritik rol oynadığı belirlenmiştir. Rastgele Orman ve SVM modelleri, yüksek doğruluk oranları ile klinik uygulamalar için en uygun modeller olarak öne çıkmıştır.

Göğüs kanseri ile ilgili genomik verilerin birleştirilmesi, analiz edilmesi ve sınıflandırılması konularında önemli bir katkı sunmaktadır. Elde edilen bulgular, hastalığın moleküler temellerini daha iyi anlamaya ve teşhis süreçlerini geliştirmeye yönelik gelecekteki araştırmalara ışık tutmaktadır. Bu tür analizler, kişiselleştirilmiş tedavi stratejilerinin geliştirilmesine ve hastaların yaşam kalitesinin artırılmasına katkıda bulunabilir.

Geniş ve Güvenilir Veri Seti: GSE25066 ve GSE3494 veri setlerinin birleştirilmesi ile elde edilen geniş veri seti, göğüs kanserinin moleküler mekanizmalarının daha kapsamlı bir şekilde incelenmesine olanak tanımaktadır. Bu, biyolojik çeşitliliği artırarak analizlerin doğruluğunu ve genellenebilirliğini artırmıştır.

Önemli Genlerin Belirlenmesi: Özellik seçimi ile belirlenen 50 önemli gen, göğüs kanseri teşhisinde ve alt tiplerinin belirlenmesinde kritik rol oynamaktadır. Bu genler, hastalığın moleküler biyolojisinin daha iyi anlaşılmasına ve potansiyel biyomarkerların tanımlanmasına yardımcı olabilir.

Yüksek Performanslı Modeller: Rastgele Orman ve SVM modelleri, yüksek doğruluk oranları ile klinik uygulamalar için güvenilir sınıflandırma araçları olarak öne çıkmaktadır. Bu modeller, göğüs kanseri teşhisinde ve tedavi planlamasında kullanılabilir.

Kişiselleştirilmiş Tedavi Stratejileri: Elde edilen bulgular, kişiselleştirilmiş tedavi stratejilerinin geliştirilmesine katkıda bulunabilir. Göğüs kanserinin alt tiplerinin doğru bir şekilde belirlenmesi, hastaların daha etkili ve hedeflenmiş tedavi yöntemlerine yönlendirilmesine yardımcı olabilir.

Sonuç olarak, bu çalışma göğüs kanseri gen ekspresyon verilerinin birleştirilmesi, özellik seçimi ve sınıflandırma süreçlerinde önemli ilerlemeler kaydetmiştir. Elde edilen bulgular, göğüs kanseri teşhis ve tedavisinde yeni yaklaşımlar geliştirilmesine olanak tanıyarak, hastalığın moleküler temellerinin daha iyi anlaşılmasına katkıda bulunmaktadır. Bu çalışma, gelecekteki araştırmalara ve klinik uygulamalara ışık tutarak, göğüs kanseri ile mücadelede yeni ve yenilikçi yaklaşımların geliştirilmesine olanak sağlamaktadır.

## 8. Çalışmanın Göğüs Kanseri Araştırmalarına Katkıları

Bu çalışma, göğüs kanseri araştırmalarına ve tedavi stratejilerine çeşitli önemli katkılar sağlamaktadır. Öncelikle, GSE25066 ve GSE3494 veri setlerinin birleştirilmesi ile elde edilen geniş ve çeşitli örneklem, göğüs kanserinin moleküler mekanizmalarının daha kapsamlı ve güvenilir bir şekilde anlaşılmasına olanak tanımaktadır. Birleştirilmiş veri seti, daha fazla hasta verisi içerdiği için biyolojik çeşitliliği artırmakta ve bu da elde edilen sonuçların genellenebilirliğini ve doğruluğunu artırmaktadır.

Bu çalışmada yapılan özellik seçimi ve sınıflandırma işlemleri, göğüs kanseri ile ilişkili en anlamlı genlerin ve biyolojik yolların belirlenmesine yardımcı olmuştur. Özellik seçimi işlemi, yüksek boyutlu gen ekspresyon verilerinden en önemli genleri seçerek, sınıflandırma modellerinin performansını artırmakta ve daha doğru teşhis ve prognostik modeller geliştirilmesine katkıda bulunmaktadır. Bu modeller, göğüs kanseri alt tiplerinin daha doğru bir şekilde ayrılmasını sağlayarak, hastaların kişiselleştirilmiş tedavi stratejilerine yönlendirilmesine yardımcı olabilir.

Çalışmada kullanılan Python programlama dili ve ilgili biyoinformatik kütüphaneler, verilerin işlenmesi ve analiz edilmesi sürecini kolaylaştırmakta ve hızlandırmaktadır. Python, veri analizi ve makine öğrenimi alanlarındaki güçlü araçları sayesinde, büyük ve karmaşık biyolojik veri setlerinin etkin bir şekilde analiz edilmesini sağlamaktadır. Bu, araştırmacıların göğüs kanseri ile ilişkili kritik genleri ve biyolojik yolları daha hızlı ve doğru bir şekilde belirlemesine olanak tanımaktadır.

Elde edilen sonuçlar, göğüs kanserinin erken teşhisi ve tedaviye yanıtın tahmin edilmesi gibi klinik uygulamalarda kullanılabilecek potansiyel biyomarkerlar ve terapötik hedefler sunmaktadır. Bu, hastaların yaşam kalitesini artırmak ve sağkalım oranlarını iyileştirmek amacıyla daha etkili ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesine katkıda bulunmaktadır. Ayrıca, bu çalışma, biyoinformatik ve makine öğrenimi yaklaşımlarının kanser araştırmalarında nasıl etkili bir şekilde kullanılabileceğine dair önemli bir örnek teşkil etmektedir.

Sonuç olarak, bu çalışma, göğüs kanseri araştırmalarına ve tedavisine yönelik önemli bulgular sunmakta ve hastalığın moleküler temellerinin daha iyi anlaşılmasına katkıda bulunmaktadır. Bu bulgular, gelecekteki araştırmalara ve klinik uygulamalara ışık tutarak, göğüs kanseri ile mücadelede yeni ve yenilikçi yaklaşımların geliştirilmesine olanak tanımaktadır.

## Referanslar

1. [GEO Accession viewer \(nih.gov\)](#)
2. [GEO Accession viewer \(nih.gov\)](#)
3. [Veri Setleri - E-VAM \(tuik.gov.tr\)](#)
4. <https://www.memorial.com.tr/hastaliklar/meme-kanseri-belirtileri-tanisi-ve-tedavi-yontemleri#:~:text=S%C4%B1k%20Sorulan%20Sorular-,Meme%20Kanseri%20Nedir%3F,kad%C4%B1nlarda%20rastlanan%20bir%20kanse r%20t%C3%BCr%C3%BCd%C3%BCr>
5. Zou, J., Hao, C., Hong, G., Zheng, J., He, L., & Guo, Z. (2012). Meme Kanseri Metastazı ile İlişkili Zayıf Diferansiyel Gen İfadeleri ve İşlevleri.
6. Zhang, Y., Li, J., Liu, X., & Wang, Q. (2018). Breast Cancer Risk Prediction Using Genetic Polymorphisms. *Cancer Research*.
7. Thompson, D., & Easton, D. F. (2002). The Role of BRCA1 and BRCA2 Mutations in Breast Cancer. *Human Molecular Genetics*.
8. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E., & Gianni, L. (2016). Triple-Negative Breast Cancer: Challenges and Opportunities. *Nature Reviews Clinical Oncology*.
9. Colditz, G. A., Wolin, K. Y., & Gehlert, S. (2012). The Impact of Lifestyle Factors on Breast Cancer Risk. *Breast Cancer Research and Treatment*.
10. Perou, C. M., Sørlie, T., Eisen, M. B., et al. (2000). The Molecular Subtypes of Breast Cancer: Clinical Implications. *Nature*.
11. Jones, P. A., & Baylin, S. B. (2002). The Role of Epigenetics in Breast Cancer Development and Progression. *Nature Reviews Genetics*.
12. Yates, L. R., Knappskog, S., Wedge, D., et al. (2017). Genomic Landscape of Breast Cancer Metastasis. *Nature*.