**Rahim Ağzı Kanseri Teşhisinde Sınıflandırma Yöntemlerinin Etkisi**

Samet Zengin1\* (Orcid ID: 0009-0004-9235-7521)

*1Yazılım Mühendisliği / Fen Bilimleri Enstitüsü, Bandırma Onyedi Eylül Üniversitesi, Türkiye*

*\*(smt.zngn02@gmail.com)*

**Geliş Tarihi (Received):****Kabul Tarihi(Accepted):**

***Özet –*** Bu çalışmada, rahim ağzı kanseri teşhisinde sınıflandırma yöntemlerinin etkisini incelemek amacıyla üç farklı makine öğrenmesi modeli olan Destek Vektör Makineleri, Yapay Sinir Ağları ve Random Forest karşılaştırılmıştır. Çalışmanın veri seti, rahim ağzı kanserini etkileyebilecek klinik testler ve demografik faktörlerden oluşmuştur. Bu veri setinde Schiller, Hinselmann ve Sitoloji test sonuçları gibi özellikler, modelin performansını analiz etmek için kullanılmıştır. Öncelikle, modellerin performansı Biopsy, DxCancer ve Dx:HPV hedef değişkenleri üzerinde değerlendirilmiştir. Performans ölçütü olarak ROC eğrisi, Precision-Recall eğrisi ve Karışıklık Matrisleri kullanılmıştır. Sonuçlarına göre, Random Forest modeli, özellikle Biopsy hedef değişkeninde %96 doğruluk, 0.69 F1 skoru ve 0.98 AUC değeri ile en iyi performansı göstermiştir. Destek Vektör Makineleri ve Yapay Sinir Ağları modelleri, bazı hedef değişkenlerde kabul edilebilir performans sunsa da, Random Forest modelinin doğruluk ve genelleme yeteneği üstün bulunmuştur. Ayrıca, Grid Search yöntemi ile Random Forest modelinin hiperparametre optimizasyonu gerçekleştirilmiştir ve en iyi parametreler belirlenmiştir. Modelin sunduğu özellik önem sıralaması, Schiller Testi’nin sınıflandırmada en önemli özellik olduğunu, bunu Hinselmann Testi ve Sitoloji’nin takip ettiğini göstermiştir. Bu bulgular, klinik testlerin makine öğrenmesi yöntemleriyle entegre edilmesinin teşhis sürecindeki etkinliği artırabileceğini ortaya koymuştur. Sonuç olarak, makine öğrenmesi modellerinin rahim ağzı kanseri teşhisinde geleneksel yöntemlere kıyasla önemli avantajlar sağladığı görülmüştür. Özellikle Random Forest modeli, özelliklerin önem derecelerini belirlemesiyle klinik karar destek sistemlerinin geliştirilmesine katkı sunabilecek potansiyele sahiptir.

***Anahtar Kelimeler*** *–* Rahim ağzı kanseri, destek vektör makineleri, random forest, yapay sinir ağları, hiperparametre optimizasyonu, özellik çıkarımı

1. **GİRİŞ**

Rahim ağzı kanseri, dünya genelinde kadınlar arasında en yaygın görülen ve önemli bir halk sağlığı olan kanser türlerinden biridir. Bu kanser türü, çoğunlukla insan papilloma virüsü (HPV) enfeksiyonu ile ilişkili olup, düzgün tarama ve erken teşhis yöntemleri kullanılmadığı durumlarda ciddi mortalite oranlarına yol açmaktadır (Devi ve ark., 2016).

Dünya Sağlık Örgütü’ne (WHO) göre, rahim ağzı kanseri her yıl yaklaşık 570 bin yeni vaka ve 311 bin ölümle sonuçlanmakta, bu da kadınlar arasında dünyada dördüncü en sık görülen kanser türünü oluşturmaktadır (Bosch ve De Sanjosé, 2007). HPV, Rahim ağzı kanserinin ana etkeni olarak tanımlanmakta ve 16 ile 18 türleri bu hastalığın %70’inden sorumlu tutulmaktadır (Saslow ve ark., 2012).

Rahim ağzı kanserinin erken evrelerinde, Schiller testi, Sitoloji (Pap smear) ve Hinselmann gibi klinik değerlendirmelerle tespit edilebilmektedir. Ancak, bu geleneksel yöntemler; uygulamacının deneyimine bağlı hata riski, zaman alıcı olması ve maliyet etkinliğinin düşük olması gibi sorunları beraberinde getirmektedir (Devi ve ark., 2016). Bu durum, daha hassas ve otomatik karar destek sistemlerinin geliştirilmesine olan ihtiyacı göstermektedir.

Makine öğrenimi, klinik verilerin hızlı ve doğru analiz edilmesi konusunda etkili bir yöntem olarak öne çıkmaktadır. Bu çalışmanın amacı, Rahim ağzı kanserinin erken teşhisinde Destek Vektör Makineleri (SVM), Yapay Sinir Ağları (ANN) ve Random Forest (RF) gibi makine öğrenimi algoritmalarını karşılaştırmak ve bu yöntemlerin etkinliğini değerlendirmektedir. RF algoritması, özellikle özellik önem sıralaması ve öngörü performansı konularında öne çıkmakta ve Schiller, Sitoloji ve Hinselmann gibi testlerin klinik değerini göstermektedir.

Bu çalışma, makine öğrenimi algoritmalarının Rahim ağzı kanseri teşhisindeki potansiyelini göstermeyi ve bu yöntemlerin geleneksel yöntemlere alternatif olarak klinik uygulamalarda nasıl kullanılabileceğini incelemeyi hedeflemektedir.

1. **LİTERATÜR TARAMASI**

Rahim ağzı kanseri teşhisinde makine öğrenimi yöntemlerinin etkili olduğunu çeşitli çalışmalar göstermiştir. Devi ve ark. [1], yapay sinir ağlarını kullanarak %94.5 doğruluk oranı ile rahim ağzı kanseri sınıflandırmasında yüksek performans elde etmişlerdir. Bosh ve De Sanjosé [2], HPV’nin Rahim ağzı kanseri üzerindeki nedensel etkisini epidemiyolojik verilerle açıklamış, HPV DNA’sının tanısal kesinliğini artırmada PCR tabanlı yöntemlerin önemini vurgulamışlardır. Saslow ve ark. [3], HPV’nin yüksek riskli türlerini tanımlayarak, tarama rehberlerini güncellemiş ve tarama ile mortalite oranlarının %50 oranında düşürülebileceğini belirtmişlerdir. Wu ve ark.[4], derin öğrenme tabanlı algoritmaların sitoloji görüntüleri üzerinde insan düzeyinde doğruluk sağlayarak kanser taramasını daha hızlı ve hassas hale getirdiğini göstermiştir (%96.8 doğruluk, %0.92 AUC). Khan ve ark [5]. sınıf dengeleme teknikleriyle desteklenen SVM ve KNN algoritmaları ile %95’in üzerinde AUC değerlerine ulaşmışlardır. Li ve ark. [6], STRIDE adlı büyük ölçekli bir yapay zeka sistemiyle 341,889 slaytta %98’in üzerinde doğruluk ve patalog düzeyinde açıklanabilirlik sağlamıştır. Rahaman ve ark. [7], DeepCervix adlı hibrit derin öğrenme tekniği ile iki sınıflı sınıflandırmada %99.85, beş sınıflı sınıflandırmada ise %99.14 doğruluk elde ederek mevcut yöntemleri geride bırakmıştır. Sarhangi ve ark.[8], kolposkopi görüntülerinde segmentasyon ve sınıflandırma için derin öğrenme kullanmış, %93 doğrulukla hassas bi tarama sağlamıştır. Poudel ve ark. [9], düşük kaynaklı ülkelerdeki görsel muayene testlerinde, akıllı telefon destekli yapay zeka sistemleri ile hassasiyetin artırılabileceğini göstermiştir. Son olarak, Rahimi ve ark. [10], rahim ağzı kanserinde hayatta kalma tahmini için makine öğrenimi algoritmalarının kullanıldığını ve AUC değerlerinin %0.40 ile %0.99 arasında değiştiğini rapor etmiştir.

Bu çalışmalar, rahim ağzı kanserinin erken teşhisinde makine öğrenimi vey apay zeka tabanlı sistemlerin potansiyelini ve bu yöntemlerin klinik uygulamalarda nasıl fayda sağlayabileceğini göstermektedir.

1. **MATERYAL VE YÖNTEM**

3.1 **Veri Seti**

Bu çalışmada kullanılan veri seti, rahim ağzı kanseri teşhisinde etkili olan kinik ve demografik bilgileri içermektedir. Veri seti, toplamda 36 özelliği ve 835 gözlemi içerir. Bu özellikler, rahim ağzı kanserine yönelik erken teşhis ve risk faktörlerinin analizine olanak tanımaktadır. Tablo 1’de veri setindeki seçilen önemli özellikler ve bunların açıklamaları verilmiştir:

|  |  |
| --- | --- |
| **Özellik** | **Açıklama** |
| **Age** | Hastanın yaşı. |
| **Number of sexual partners** | Hastanın cinsel partner sayısı. HPV enfeksiyonu riskini değerlendirmek için önemlidir. |
| **First sexual intercourse** | İlk cinsel ilişkinin gerçekleştiği yaş. Erken yaşta cinsel aktivite HPV enfeksiyonu riskini artırır. |
| **Num of pregnancies** | Hastanın geçirdiği toplam gebelik sayısı. Hormonal değişiklikler risk faktörlerini etkileyebilir. |
| **Smokes** | Sigara içme durumu. Sigara, bağışıklık sistemini baskılayarak HPV’nin ilerlemesini kolaylaştırabilir. |
| **Hormonal Contraceptives** | Hormonal doğum kontrol yöntemlerinin kullanım süresi. Uzun süreli kullanım bazı kanser türleriyle ilişkilidir. |
| **STDs:HPV** | Hastanın HPV ile ilişkili cinsel yolla bulaşan hastalık geçmişi. |
| **Dx:Cancer** | Kanser teşhisini gösteren bir özellik. |
| **Biopsy** | Hastanın biyopsi sonuçlarını ifade eden değişken. Rahim ağzı kanseri teşhisinin temel çıktısıdır. |

Tablo 1. Veri setinde seçilen önemli özellikler ve açıklamaları

Veri setinde eksik değerler medyan ile doldurulmuş ve tüm veriler analize uygun hale getirilmiştir. Ayrıca, aşağıda veri setinden 5 örnek gözlem sunulmaktadır.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Age** | **Number of sexual partners** | **First sexual intercourse** | **Num of pregnancies** | **Smokes** | **Hormonal Contraceptives** | **STDs:HPV** | **Dx:Cancer** | **Biopsy** |
| 25 | 1 | 18 | 0 | 0 | 0 | 0 | 0 | 0 |
| 30 | 2 | 16 | 1 | 1 | 1 | 1 | 0 | 1 |
| 35 | 3 | 21 | 2 | 0 | 1 | 0 | 1 | 0 |
| 40 | 2 | 19 | 3 | 1 | 0 | 1 | 0 | 1 |
| 28 | 1 | 18 | 0 | 0 | 0 | 0 | 0 | 0 |

Tablo 2. Veri seti örnekleri

Bu özelliklerin seçiminde, rahim ağzı kanseriyle ilgili literatürden elde edilen bilgiler ve model tarafından belirlenen özellik önem dereceleri dikkate alınmıştır. Ayrıca, bu özelliklerin dağılımları histogramlar yardımıyla görselleştirilmiş ve veri setinin genel yapısı incelenmiştir. Aşağıda özelliklerin histogram dağılımları listelenmektedir.

diyagram, ekran görüntüsü, metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 1: Yaş özelliğine göre dağılım grafiği.

metin, ekran görüntüsü, diyagram, ekran, görüntüleme içeren bir resim

AI-generated content may be incorrect.

Şekil 2: Hastanın cinsel partner sayısına göre dağılım grafiği.

metin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 3: İlk cinsel ilişkinin gerçekleştiği yaşın dağılım grafiği.

metin, ekran görüntüsü, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 4: Geçirilen toplam gebelik sayısının dağılım grafiği.

metin, ekran görüntüsü, ekran, görüntüleme, diyagram içeren bir resim

AI-generated content may be incorrect.

Şekil 5: Sigara içen sayısının dağılım grafiği.

metin, ekran görüntüsü, ekran, görüntüleme, çizgi içeren bir resim

AI-generated content may be incorrect.

Şekil 6: Doğum kontrol yöntemlerini kullananların dağılım grafiği.

metin, ekran görüntüsü, ekran, görüntüleme, sayı, numara içeren bir resim

AI-generated content may be incorrect.

Şekil 7:HPV ile ilişkili geçen hastalık geçmişinin dağılım grafiği.

metin, ekran görüntüsü, ekran, görüntüleme, sayı, numara içeren bir resim

AI-generated content may be incorrect.

Şekil 8:DxCancer özelliğinin dağılım grafiği.

metin, ekran görüntüsü, ekran, görüntüleme, çizgi içeren bir resim

AI-generated content may be incorrect.

Şekil 9: Biyopsi sonuçlarının dağılım grafiği.

3.2 **Hiperparametre Optimizasyonu ve Özellik Çıkarımı**

**Hiperparametre Optimizasyonu**: Hiperparametre optimizasyonu, bir makine öğrenmesi modelinin performansını en iyi hale getirmek için modelin eğitilmeden önce ayarlanması gereken parametrelerin (hiperparametrelerin) optimize edilme sürecidir [18]. Makine öğrenmesi algoritmalarının performansı, doğrudan doğru hiperparametre ayarına bağlıdır.

* Hiperparametreler, modelin yapısını ve eğitim sürecini kontrol eder. Örneğin, RF modelinde ağaç sayısı (n\_estimators) ve maksimum derinlik (max\_depth) gibi parametreler hiperparametrelerdir.
* Hiperparametre optimizasyonunda, yaygın olarak Grid Search ve Random Search gibi yöntemler kullanılır [18] :
  + **Grid Search:** Tüm olası hiperparametre kombinasyonlarını dener ve en iyi sonucu veren kombinasyonu seçer.
  + **Random Search:** Rastgele hiperparametre kombinasyonları arasından denemeler yapar ve en iyisini bulur.

**Özellik Çıkarımı:** Özellik çıkarımı, bir veri setindeki ham verilerden modelin öğrenmesini kolaylaştıracak daha anlamlı ve özet özelliklerin seçilmesi veya türetilmesi işlemidir. Modelin performansını arttırmak için kullanılan bu yöntem gereksiz özellikleri azaltarak modelin daha hızlı çalışmasını sağlar.

Bu çalışmada ise, karşılaştırılan makine öğrenmesi modelleri arasında en iyi performansı veren model seçilerek hiperparametre optimizasyonu ve özellik çıkarımı yöntemleri uygulanacaktır.

3.3 **Makine Öğrenmesi Yöntemleri**

Makine öğrenmesi (ML), bilgisayarların veri kullanarak belirli bir görevi yerine getirmeyi öğrenmesini sağlayan bir yapay zeka alt alanıdır. Bu yöntemler, geçmiş verilere dayalı olarak tahminlerde bulunma, sınıflandırma yapma ve örüntüleri belirleme gibi görevlerde kullanılır [11][12].

Bu çalışmada üç farklı makine öğrenimi yöntemi kullanılmıştır. SVM, RF ve ANN. Bu algoritmaların çalışma prensipleri aşağıda açıklanmıştır.

**Destek Vektör Makineleri (SVM)**

**SVM,** sınıflandırma ve regresyon analizinde kullanılan güçlü bir algoritmadır. Sınıflandırma problemlerinde SVM, farklı sınıfları ayıran en iyi hiper düzlemi bulmayı hedefler [13]. Karar sınırını optimize etmek için şu matematiksel formül kullanılır [14]:

* w: Hiper düzlemin ağırlık vektörü,
* b: Bias terimi,
* xi​: Gözlem noktası,
* yi​: Sınıf etiketi.

SVM, doğrusal olmayan durumlarda da çekirdek fonksiyonları (Kernel functions) kullanarak sınıfları ayırabilir. Çekirdek fonksiyonları, veri noktalarını daha yüksek boyutlu bir uzaya dönüştürür.

metin, diyagram, çizgi, yazı tipi içeren bir resim

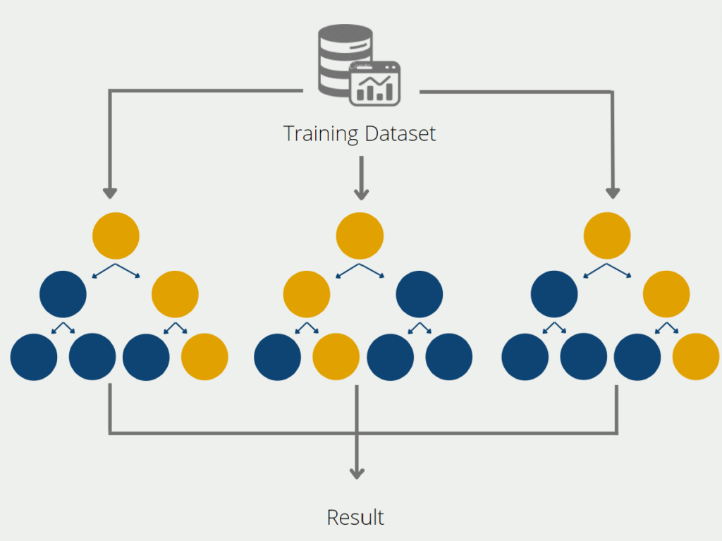
AI-generated content may be incorrect.

Şekil 10: SVM formülü ve matrix vektör çarpımı (Albert Cano, 2018)

**Random Forest (RF)**

RF, birden fazla karar ağacından oluşan bir topluluk yöntemidir. Bu algoritma, her bir ağacın bağımsız olarak eğitilmesi ve nihai sınıflandırmanın çoğunluk oyu prensibi ile belirlenmesiyle çalışır. Modelin temel prensipleri şunlardır:

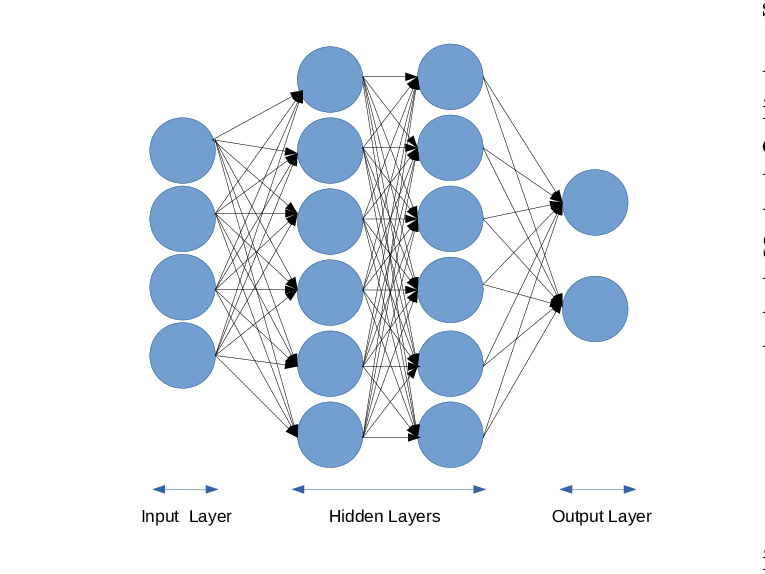
1. Eğitim veri setinden rastgele örnekleme ile alt veri setleri oluşturulur.
2. Her bir alt veri seti ile bir karar ağacı eğitilir.
3. Nihai sınıflandırma, tüm ağaçların tahminlerinin ortalaması veya çoğunluk oyu ile belirlenir.



Şekil 11: RF çalışma mantığı (What is a Random Forest?, 2022 Database Camp)

**Yapay Sinir Ağları (ANN)**

ANN, veriler arasındaki karmaşık ilişkileri öğrenmek için kullanılan çok katmanlı bir yapay zeka modelidir. ANN, tablosal veri üzerinde çalışmak için idealdir ve genellikle sınıflandırma, regresyon ve örüntü tanıma gibi görevlerde kullanılır [11][17].



Şekil 12: ANN katmanları

Bir ANN şu temel bileşenlerden oluşur:

1. **Giriş Katmanı**: ANN’in başlangıç noktasıdır ve verileri ağa ileten katmandır.
2. **Gizli Katmanlar**: Veriler arasındaki ilişkileri öğrenir ve ara işlemlerin gerçekleştirildiği katmandır. Bir veya birden daha fazla katman bulunabilir.
3. **Çıkış Katmanı**: Modelin tahminlerini ve sınıflandırma sonuçlarının alındığı katmandır.

Matematiksel olarak bir ANN’in aktivasyonu şu şekilde ifade edilir:

* z: Aktivasyon çıktısı,
* W: Ağırlık matrisi,
* x: Giriş verisi,
* b: Bias terimi,
* σ: Aktivasyon fonksiyonu

1. **DENEYSEL SONUÇLAR**

Bu çalışmada, rahim ağzı kanseri teşhisinde kullanılan üç farklı makine öğrenmesi yöntemi (SVM,ANN ve RF) değelendirilmiş ve karşılaştırılmıştır. Modellerin performansı doğruluk (Accuracy), F1 skoru ve ROC-AUC gibi metriklerle analiz edilmiştir. Ayrıca, her bir modelin farklı hedef değişkenler üzerindeki etkisi incelenmiş ve görselleştirilmiştir. Deneysel süreç ve sonuçlar aşağıdaki gibi sunulmaktadır:

**4.1 Performans Karşılaştırılması**

Modeller, biyopsi(Biyopsy), kanser teşhisi (DxCancer) ve HPV teşhisi (DxHPV) hedef değişkenlerine göre değerlendirilmiştir. Aşağıda her bir modelin metriklerinin yer aldığı tablo bulunmaktadır:

| **Model** | **Hedef Değişken** | **Doğruluk (Accuracy)** | **F1 Skoru** | **AUC** |
| --- | --- | --- | --- | --- |
| **SVM** | Biopsy | 93.63% | 0.00 | 0.97 |
|  | Dx:Cancer | 97.61% | 0.00 | 0.97 |
|  | Dx:HPV | 98.01% | 0.00 | 0.05 |
| **RF** | Biopsy | 96.02% | 0.64 | 0.98 |
|  | Dx:Cancer | 98.01% | 0.29 | 0.98 |
|  | Dx:HPV | 98.01% | 0.00 | 0.98 |
| **ANN** | Biopsy | 94.42% | 0.50 | 0.90 |
|  | Dx:Cancer | 97.61% | 0.00 | 0.45 |
|  | Dx:HPV | 98.01% | 0.00 | 0.50 |

Tablo 3. Makine öğrenmesi modellerinin hedef değişkenleri üzerinde karşılaştırılması

Tablo 2’ye göre, SVM modeli, genel olarak yüksek doğruluk oranı vermiştir fakat hiçbir F1 skor verememiştir bunun yanı sıra Biopsy ve DxCancer hedef değişkenlerinde iyi sonuçlar verirken Dx:HPV hedef değişkeninde rastgele tahminlerde bulunmuştur ve AUC değeri çok düşük gelmiştir. RF modeli ise, SVM’e benzer şekilde doğruluk oranları vermiştir SVM’den farklı olarak hedef değişkenleri sırasıyla Biopsy, DxCancer ve DxHPV olmak üzere 0.64, 0.29 ve 0 F1 skorları üretmiştir. Ayrıca, 0.98 AUC değeri ile istikrarını korumuştur. Son olarak ANN modeli ise, tıpkı SVM ve RF gibi yüksek doğruluk değerleri elde etmiştir. SVM’e farklı olarak Biopsy hedef değişkeninde F1 skor üretebilmiştir. Ayrıca, istikrarlı olmasa da 0.90, 0.45 ve 0.50 AUC değerleri üretmiştir.

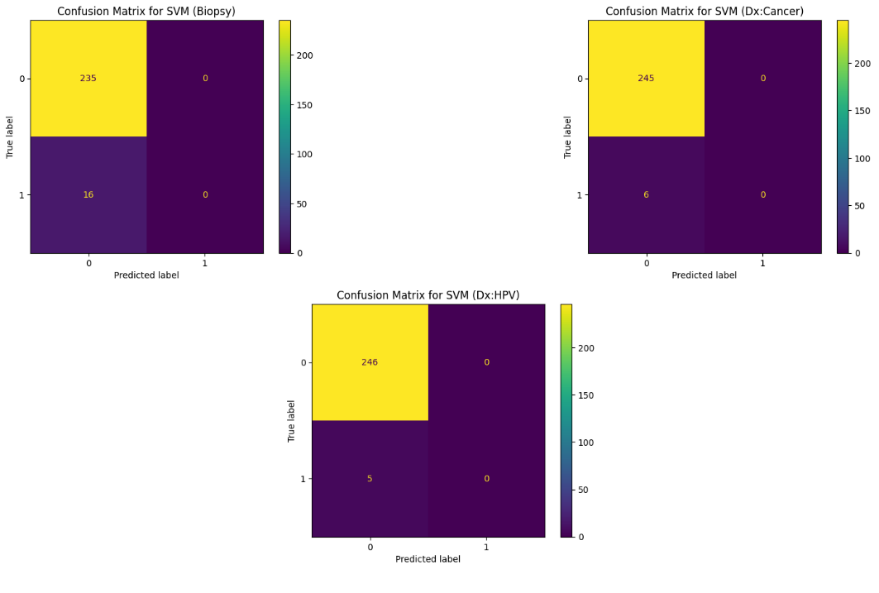
Çalışmanın devamında Tablo 2’deki veriler grafiklerle daha detaylı açıklanmıştır.

**4.2 Confusion Matrix (Karışıklık Matrisi)**

Modellerin karışıklık matrisleri, farklı hedef değişkenleri üzerinde değerlendirilmiştir. Karışıklık matrisi, bir sınıflandırma modelinin performansını değerlendirmek için kullanılır. Bu matris, modelin tahmin ettiği sınıflar ile gerçek sınıflar arasındaki ilişkiyi görselleştirir ve aşağıdaki dört değeri içerir:

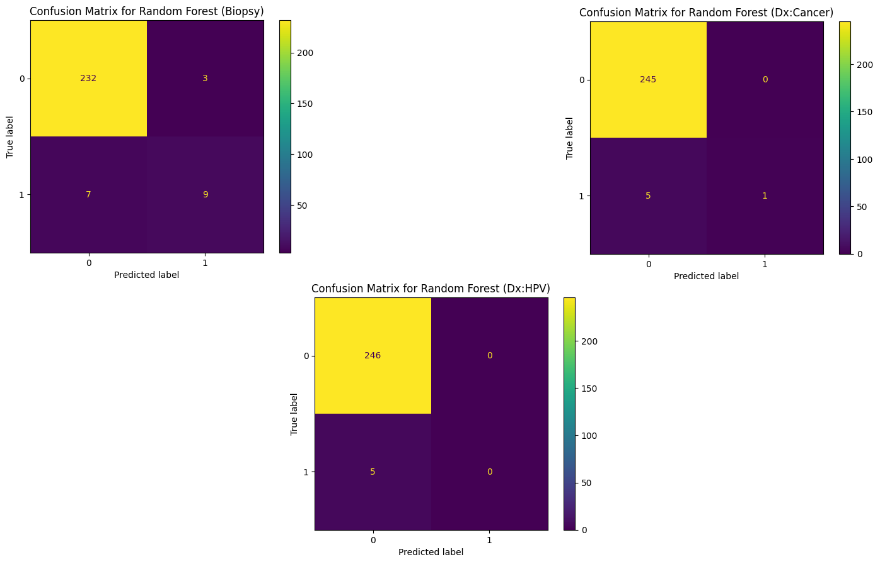
1. **True Positives (TP):** Doğru bir şekilde pozitif olarak tahmin edilen örnekler.
2. **True Negatives (TN):** Doğru bir şekilde negatif olarak tahmin edilen örnekler.
3. **False Positives (FP):** Gerçekte negatif olan ancak pozitif olarak tahmin edilen örnekler (Type I Error).
4. **False Negatives (FN):** Gerçekte pozitif olan ancak negatif olarak tahmin edilen örnekler (Type II Error).

Modellerin hedef değişken üzerindeki karışıklık matrisleri bu başlık altındaki şekillerde açıklamalarıyla birlikte gösterilmektedir.



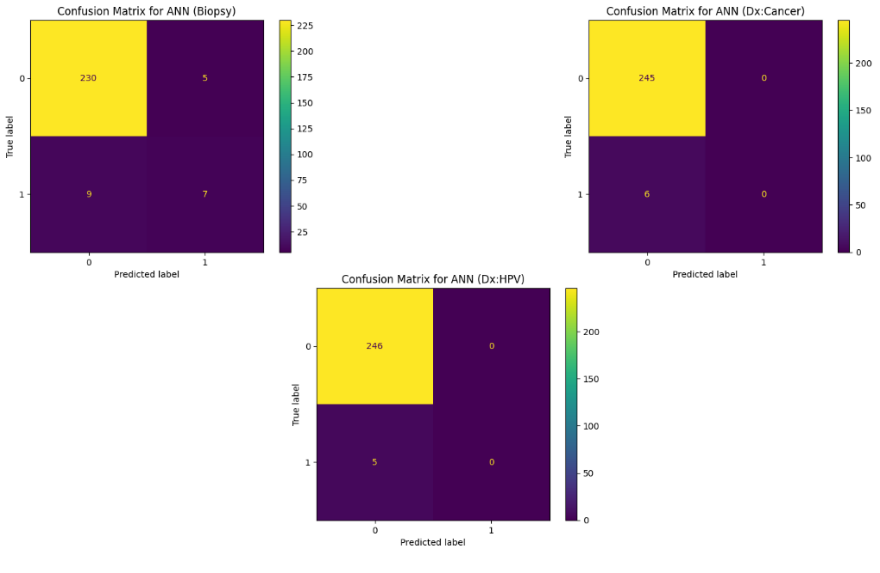
Şekil 13: SVM için karışıklık matris

Şekil 13’de SVM için oluşturulan karışıklık matrisleri, tüm hedef değişkenlerde modelin pozitif sınıfı doğru bir şekilde tahmin edemediğini göstermektedir. **Biopsy**, **Dx:Cancer** ve **Dx:HPV** hedef değişkenlerinde tüm pozitif örnekler yanlış negatif olarak tahmin edilmiş, yalnızca negatif sınıf üzerinde yüksek doğruluk elde edilmiştir. Bu durum, sınıf dengesizliği ve modelin pozitif sınıfı öğrenememe problemine işaret etmektedir.



Şekil 14: RF için karışıklık matris

Şekil 14’de RF, Biopsy hedef değişkeninde genel olarak iyi bir performans göstermiştir. Ancak Dx:Cancer ve Dx:HPV hedef değişkenlerinde pozitif sınıfta tahminler sınırlı kalmıştır. Bu sonuçlar, modelin pozitif sınıfı tanımada zorlandığını ve veri setindeki dengesizliğin pozitif tahmin performansını etkileyebileceğini göstermektedir. Özellikle Dx:HPV için, modelin pozitif sınıfta hiç doğru tahmin yapamaması önemli bir eksikliktir.



Şekil 15: ANN için karışıklık matrisi

Şekil 15’de ANN, Biopsy hedef değişkeninde pozitif ve negatif sınıflar arasında kısmen dengeli bir performans göstermiştir, ancak Dx:Cancer ve Dx:HPV hedef değişkenlerinde pozitif sınıfta tamamen başarısız olmuştur. Bu durum, ANN’in dengesiz veri setinde pozitif sınıfı öğrenme zorluğu yaşadığını göstermektedir.

Bu sonuçlar, sınıf dengesizliği gibi problemler karşısında RF’in daha dayanıklı olduğunu ve rahim ağzı kanseri teşhisinde en uygun model olarak tercih edilmesi gerektiğini göstermektedir.

**4.3 Precision-Recall Eğrileri**

Modellerin**,** hassasiyet (Precision) ve duyarlılık (Recall) arasındaki ilişkiler Şekil 16,17 ve 18’de gösterilmiştir. Tablo 2.’de sunulan F1 Skoru hassasiyet ve recall’un dengeli bir ölçütüdür. Hassasiyet, duyarlılık ve F1 skor metriklerinin formülleri aşağıdaki gibi hesaplanmaktadır:

metin, diyagram, çizgi, plan içeren bir resim

AI-generated content may be incorrect.

Şekil 16: SVM için Precision-Recall Eğrisi

Şekil 16’de sunulan **SVM için Precision-Recall eğrisi**, **Biopsy** hedef değişkeninde Precision ve Recall arasında orta seviyede bir ilişki sergilemiştir. Eğri, belirli noktalarda Precision’ın düştüğünü göstermesine rağmen genel olarak dengeli bir performans sunmaktadır. **Dx:Cancer** hedef değişkeninde ise Precision ve Recall değerleri oldukça düşük seviyelerde kalmış, modelin bu değişken için sınırlı bir ayrıştırma yeteneğine sahip olduğunu göstermiştir. **Dx:HPV** hedef değişkeninde ise Precision ve Recall değerleri neredeyse sıfır seviyesindedir. Bu durum, SVM’in HPV teşhisinde başarısız olduğunu ve tahminlerinin rastgele olduğunu ortaya koymaktadır. Genel olarak, SVM, **Biopsy** ve **Dx:Cancer** değişkenlerinde kısmen etkili bir performans sergilerken, **Dx:HPV** için tatmin edici bir sonuç üretememiştir.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 17: RF için Precision-Recall Eğrisi

Şekil 17’de sunulan RF için Precision-Recall eğrisi, tüm hedef değişkenlerde güçlü bir ilişki göstermiştir. **Biopsy** hedef değişkeninde Precision ve Recall yüksek seviyelerde dengeli bir şekilde ilerlemiş, modelin güçlü bir sınıflandırma performansı sunduğunu göstermiştir. **Dx:Cancer** hedef değişkeninde de benzer bir başarı gözlemlenmiştir. **Dx:HPV** hedef değişkeni için eğri, modelin bu hedefte de oldukça yüksek bir performansa sahip olduğunu ortaya koymuştur.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 18: ANN için Precision-Recall Eğrisi

Şekil 18’de sunulan **ANN için Precision-Recall eğrisi**, **Biopsy** hedef değişkeninde orta seviyede bir performans göstermiştir. Precision ve Recall değerleri bazı aralıklarda istikrarsızlık göstermiştir, ancak genel olarak kabul edilebilir bir sonuç sunmuştur. **Dx:Cancer** hedef değişkeninde ise Precision ve Recall oldukça düşük seviyelerde kalmış ve modelin sınırlı bir başarısı olduğunu göstermiştir. **Dx:HPV** hedef değişkeninde Precision ve Recall değerleri sıfır seviyesine yakın olup, ANN'in bu hedef değişkeni üzerinde başarısız olduğunu göstermiştir. ANN, yalnızca Biopsy hedef değişkeninde makul bir performans sunarken diğer değişkenlerde yetersiz kalmıştır.

* 1. **ROC Eğrileri**

Modellerin ROC eğrileri, farklı hedef değişkenler için aşağıdaki Şekil 19, 20 ve 21’de sunulmaktadır. Bu eğriler, modelin pozitif sınıfı doğru bir şekilde ayırma performansını görselleştirmektedir.

**metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.**

Şekil 19 SVM için ROC eğrisi

**Şekil 19’de** sunulan SVM için ROC eğrisi, **Biopsy** hedef değişkeninde AUC (Alan Altında Kalan Alan) değeri 0.97 ile oldukça yüksektir. BU SVM’in biyopsi sınıflandırmasında güçlü bir ayrıştırma yeteneğine sahip olduğunu göstermektedir. **DxCancer** hedef değişkeninde yine 0.97’dir. Model, kanser teşhisinde biyopsiye benzer bir doğruluk göstermiştir. **DxHPV** hedef değişkeninde ise AUC değeri 0.05 gibi çok düşük bir değerde kalmıştır. Bu SVM’in HPV teşhisinde oldukça düşük bir performans gösterdiğini ve rastgele bir tahminde bulunduğunu göstermiştir. SVM, biyopsi ve kanser teşhisinde etkili ancak HPV teşhisinde başarısız performans sergilemiştir.

metin, çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 20: RF için ROC eğrisi

**Şekil 20’de** sunulan RF için ROC eğrisi, **Biopsy** hedef değişkeninde 0.98 ile çok yüksektir. Bu RF’in biopsy sınıflandırmasında oldukça başarılı olduğunu göstermektedir. **DxCancer**  hedef değişkeninde ise AUC değeri yine 0.98’dir. Model, kanser teşhisinde oldukça başarılıdır. **DxHPV**  hedef değişkeninde ise AUC değeri 0.98 olarak görülmektedir ve bu modelin, HPV teşhisinde de tutarlı ve yüksek bir performans gösterdiğini göstermekedir. RF, tüm hedef değişkenlerde oldukça yüksek bir doğrulukla çalışmıştır ve en dengeli performansı sergilemiştir.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

AI-generated content may be incorrect.

Şekil 21 ANN için ROC eğrisi

**Şekil 21’de** sunulan ANN için ROC eğrisi, **Biopsy** hedef değişkeninde AUC değeri 0.90 ile kabul edilebilir bir düzeydedir. Bu ANN’in, biyopsi sınıflandırmasında başarılı olduğunu göstermektedir. **DxCancer** hedef değişkeninde AUC değeri 0.45 gibi düşük bir seviyededir. Bu ANN’in kanser teşhisinde çok sınırlı bir performansa sahip olduğunu işaret etmektedir. **DxHPV**  hedef değişkenin de ise AUC değeri 0.50’dir. Bu değer, modelin rastgele bir tahmin seviyesinde olduğunu göstermektedir. ANN, biyopsi için yeterince iyi bir performans sergilemiş olsa da kanser ve HPV teşhisinde düşük bir başarı göstermiştir.

**4.5 Random Forest Hiperparametre Optimizasyonu ve Özellik Önem Sıralaması Sonuçları**

Bu başlıkta, üst başlıklarda karşılaştırılan modeller arasında en iyi performansı veren RF modeli kullanılarak hiperparametre optimizasyonu ve özellik çıkarımı yöntemleri uygulanacaktır. RF modeli için hedef değişken **Biopsy** olarak belirlenmiştir.

**Hiperparametre Optimizasyonu**

Hiperparametre optimizasyonunun iki yaygın çeşidi vardır. Bunlar Grid Search ve Random Search’dur. Hiperparametre Optimizasyonu ve Özellik Çıkarımı başlığı altında bu iki çeşidin kullanım amacı açıklanmıştır. Çalışmanın devamında, Grid Search optimizasyon yöntemi kullanılacaktır.

Grid Search yöntemi sonucunda, RF modeli için en iyi hiperparametre kombinasyonu şu şekilde belirlenmiştir:

|  |  |
| --- | --- |
| **Parametre** | **Değer** |
| max\_depth | None |
| min\_samples\_leaf | 1 |
| min\_samples\_split | 10 |
| n\_estimators | 50 |

Tablo 4. Seçilen hiperparametreler ve değerleri

Tablo 3’de verilen parametreler, modelin sınıflandırma performansını optimize etmek için seçilmiş ve bu parametrelerle yeniden eğitilen model, sınıflandırma görevinde en yüksek performansı göstermiştir.

**Random Forest Performans Metrikleri**

Çalışmanın sonuç kısımlarında belirleyici olan RF modelinin hiperparametre optimizasyonunda hedef değişken olan Biopsy değişkeni üzerindeki performans metrikleri Tablo 5’de sunulmaktadır. Tablo 5 incelendiğinde, RF modeli Biopsy değişkeni üzerinde 0.96 doğruluk, 0.69 F1 skor ve 0.98 AUC değerlerini elde etmiştir. Elde edilen bu sonuçlar neticesinde özellik çıkarımı ve özellik önem sıralaması belirlenirken ciddi katkıda bulunacaktır.

|  |  |
| --- | --- |
| **Performans Metrikleri** | **Değer** |
| Doğruluk | None |
| F1 Skor | 1 |
| AUC | 10 |

Tablo 5. Seçilen hiperparametreler ve değerleri

**Özellik Çıkarımı ve Özellik Önem Sıralaması**

RF modeli, hedef değişken olan **Biopsy** için tahmin yaparken kullandığı özelliklerin önem derecelerini belirlemiştir. Tablo 5’de en önemli 10 özellik ve önem skorları sunulmuştur:

| **Özellik** | **Önem Skoru** |
| --- | --- |
| Schiller Testi | 0.385655 |
| Hinselmann Testi | 0.161460 |
| Sitoloji (Cytology) | 0.075954 |
| Yaş (Age) | 0.060403 |
| İlk Cinsel Deneyim (First sexual intercourse) | 0.053016 |
| Hormonal Kontraseptif Kullanımı (Hormonal Contraceptives - years) | 0.047056 |
| Cinsel Partner Sayısı (Number of sexual partners) | 0.022869 |
| Hamilelik Sayısı (Num of pregnancies) | 0.021784 |
| Sigara Kullanımı (Smokes - packs/year) | 0.021683 |
| Rahim İçi Araç Kullanım Süresi (IUD - years) | 0.021577 |

Tablo 6. Biopsy hedef değişkenine göre özellik önem sıralaması.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

AI-generated content may be incorrect.

Şekil 22: Random Forest Özellik Önem Sıralaması

RF modeline göre, Tablo 5’e ve Şekil 22’ye bakıldığında sınıflandırmada en etkili özellik Schiller Testi olmuştur. Bu özellik, rahim ağzı kanseri teşhisinde en güçlü belirleyici faktör olarak öne çıkmıştır. Hinselmann Testi ve Sitoloji de model için önemli özellikler olarak belirlenmiştir. Yaş, ilk cinsel deneyim yaşı ve hormonal kontraseptif kullanımı gibi özellikler de modelde belirli bir etkiye sahiptir. Ancak, klinik testlerinden sonra ikinci derecede önemlidir.

1. **SONUÇLAR**

Bu çalışmada rahim ağzı kanseri teşhisinde SVM, RF ve ANN modelleri kullanılarak performans karşılaştırılmaları gerçekleştirilmiştir. Hedef değişken olarak Biopsy, DxCancer ve DxHPV belirlenmiş belirlenmiş, modellerin sınıflandırma performansları çeşitli metrikler ve görsellerle değerlendirilmiştir. Son olarak, hiperparametre optimizasyonu sonrası RF modeli seçilmiş ve özellik önem sıralaması analiz edilmiştir. Bu analiz, klinik testlerin (Schiller, Hinselmann ve Sitoloji) rahim ağzı kanseri teşhisinde belirleyici faktörler olduğunu ortaya koymaktadır. Diğer yandan, demografik faktörler ve hormonel etkiler de belirli ölçüde katkı sağlamaktadır. RF modeline göre elde edilen sonuçlar, modelin yüksek doğruluk ve ayrıştırma performansı sergilediğini göstermektedir. Özellikle AUC değerinin 0.98 olması, modelin pozitif ve negatif sınıfları başarıyla ayırdığını kanıtlamaktadır. Ancak F1 skorunun 0.69 seviyesinde kalması, modelin küçük veri setindeki sınıf dengesizliğinden etkilendiğini göstermektedir.

KAYNAKLAR

[1] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, “Classification of Cervical Cancer Using Artificial Neural Networks,” Procedia Computer Science, vol. 89, pp. 465–472, 2016, doi: 10.1016/J.PROCS.2016.06.105.

[2] F. X. Bosch and S. de Sanjosé, “The epidemiology of human papillomavirus infection and cervical cancer,” Disease Markers, vol. 23, no. 4, pp. 213–227, 2007, doi: 10.1155/2007/914823.

[3] D. Saslow, D. Solomon, H. W. Lawson, M. Killackey, et al., “American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer,” CA: A Cancer Journal for Clinicians, vol. 62, no. 3, pp. 147–172, 2012, doi: 10.3322/CAAC.21139.

[4] T. Wu, E. Lucas, F. Zhao, P. Basu, and Y. Qiao, “Cancer Biology & Medicine,” Cancer Biology & Medicine, Sep. 2024, doi: 10.20892/j.issn.2095-3941.2024.0198.

[5] M. Glučina, A. Lorencin, N. Anđelić, and I. Lorencin, “Cervical Cancer Diagnostics Using Machine Learning Algorithms and Class Balancing Techniques,” Applied Sciences, vol. 13, no. 2, p. 1061, 2023, doi: 10.3390/APP13021061.

[6] H. Li et al., “Large-scale cervical precancerous screening via AI-assisted cytology whole slide image analysis,” arXiv preprint arXiv:2407.19512v1, 2024.

[7] M. M. Rahaman et al., “DeepCervix: A Deep Learning-based Framework for the Classification of Cervical Cells Using Hybrid Deep Feature Fusion Techniques,” Computers in Biology and Medicine, vol. 136, 2021, doi: 10.1016/j.compbiomed.2021.104649.

[8] H. A. Sarhangi, D. Beigifard, E. Farmani, and H. Bolhasani, “Deep Learning Techniques for Cervical Cancer Diagnosis based on Pathology and Colposcopy Images,” arXiv preprint arXiv:2310.16662v1, 2023.

[9] K. Poudel et al., “AI-Assisted Cervical Cancer Screening,” arXiv preprint arXiv:2403.11936v2, 2024.

[10] M. Rahimi, A. Akbari, F. Asadi, and H. Emami, “Cervical cancer survival prediction by machine learning algorithms: a systematic review,” BMC Cancer, vol. 23, no. 1, pp. 1–10, 2023, doi: 10.1186/S12885-023-10808-3.

[11] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[12] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, 2009.

[13] C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[14] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Mining and Knowledge Discovery, vol. 2, no. 2, pp. 121–167, 1998.

[15] L. Breiman, “Random Forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[16] D. R. Cutler et al., “Random Forests for Classification in Ecology,” Ecology, vol. 88, no. 11, pp. 2783–2792, 2007.

[17] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[18] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” Journal of Machine Learning Research, vol. 13, pp. 281–305, 2012.