

Sağlık Alanında Açıklanabilir Yapay Zekâ Kullanımı Üzerine Bir Uygulama

Gülhan TOĞA^{1,a}

¹Erciyes Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, Kayseri, Türkiye

^aORCID: 0000-0001-8835-1769

Makale Bilgileri

Geliş : 30.06.2025

Kabul : 25.09.2025

DOI: 10.21605/cukurovaumfd.1729986

Sorumlu Yazar

Gülhan TOĞA

gpala@erciyes.edu.tr

Anahtar Kelimeler

Açıklanabilir yapay zeka

SHAP

Veri madenciliği

Meme kanseri

Atıf şekli: TOĞA, G., (2025). Sağlık Alanında Açıklanabilir Yapay Zekâ Kullanımı Üzerine Bir Uygulama. Çukurova Üniversitesi, Mühendislik Fakültesi Dergisi, 40(3), 707-723.

ÖZ

Büyük verilerin hayatımıza girmesi ile bu verilerle elde edilebilecek bilgiler de artmıştır. Özellikle derin öğrenmeye yönelik olmak üzere makine öğreniminin yaygın kullanımı sayesinde büyük veri kümelerinden anlamlı çıkarımlar yapabilmek mümkün hale gelmiştir. Ancak son kullanıcılar, oluşturulan modellerin şeffaflık, yorumlanabilirlik ve güvenilirlik gibi bazı eksiklikleri ile karşı karşıya kalmaktadır. Bu eksikliklerin giderilmesi için geliştirilen “Açıklanabilir Yapay Zekâ (AYZ)”, sistematik bir yaklaşım ile oluşturulan modelin ve alınan kararların kullanıcılar tarafından anlaşılmasını ve modellerin güvenilirliğinin nasıl değerlendirilmesi gerektiğini açıklayabilmektedir.

Bu çalışmada sağlık alanında AYZ kullanımı ve faydaları hakkında bilgi verilmektedir. Ayrıca çalışma, sağlık alanında AYZ kullanımına örnek olması açısından, meme kanseri hastalarının çeşitli parametrelere bağlı olarak sağ/ölü durumlarının öngörülmesine olanak sağlayan bir uygulama içermektedir. Öncelikle, makine öğrenmesi algoritmaları kullanılarak sınıflandırma çalışması yapılmıştır. Ardından en başarılı sınıflandırma modelinin sonuçlarının açıklanabilirliği için bir AYZ yöntemi olan SHAP tekniğine başvurulmuştur. Sonuçlar, AYZ kullanımının modelin karar verme sürecinin nedenlerini daha iyi anlamamıza yardımcı olduğunu göstermektedir.

An Application on the Use of Explainable Artificial Intelligence in the Healthcare

Article Info

Received : 30.06.2025

Accepted : 25.09.2025

DOI: 10.21605/cukurovaumfd.1729986

Corresponding Author

Gülhan TOĞA

gpala@erciyes.edu.tr

Keywords

Explainable artificial intelligence

SHAP

Data mining

Breast cancer

How to cite: TOĞA, G., (2025). An Application on the Use of Explainable Artificial Intelligence in the Healthcare. Cukurova University, Journal of the Faculty of Engineering, 40(3), 707-723.

ABSTRACT

With the advent of big data, the potential to derive valuable insights has significantly increased. The widespread use of machine learning, especially deep learning, has made it possible to extract meaningful patterns from large datasets. However, end-users often encounter issues related to the transparency, interpretability, and reliability of these models. To address such concerns, Explainable Artificial Intelligence (XAI) has been developed to systematically clarify how models work and how their decisions can be understood and trusted.

This study explores the use and benefits of XAI in the healthcare domain. As an example, it presents an application that predicts the survival status (alive or deceased) of breast cancer patients based on various clinical parameters. Initially, classification was performed using machine learning algorithms. To interpret the results of the most successful model, the SHAP (SHapley Additive exPlanations) technique, a prominent XAI method, was applied. Findings indicate that XAI enhances understanding of model decisions and supports trustworthy AI applications in healthcare.

1. GİRİŞ

Yapay zeka (YZ) günümüzde birçok endüstride önemli bir rol oynamaktadır. YZ, bilgisayar sistemlerine insan benzeri zekaya sahip olma yeteneği kazandıran bir bilim dalıdır. Bu teknoloji, büyük veri setlerinden anlamlı bilgiler çıkarma, karmaşık problemleri çözme, öngörü yapma ve kararlar alma gibi bir dizi avantaj sağlamaktadır. YZ, birçok farklı alanda çeşitli uygulamalara sahip çok yönlü bir teknolojidir. YZ, finansal hizmetlerde; risk değerlendirmesi, sahtecilik tespiti, portföy yönetimi ve müşteri hizmetleri gibi alanlarda kullanılmaktadır. Perakende ve e-ticarette, büyük veri analizi ve öneri sistemleri gibi YZ teknikleri, müşteri davranışlarını anlama, stok yönetimi, fiyatlandırma ve kişiselleştirilmiş pazarlama gibi alanlarda kullanılmaktadır. YZ Ulaşım ve lojistik alanında; trafik yönetimi, rotalama, teslimat planlaması ve sürücüsüz araç teknolojileri gibi alanlarda önemli bir rol oynamaktadır. Enerji ve çevre sektörlerinde, enerji verimliliği, güneş paneli optimizasyonu, akıllı şebeke yönetimi gibi alanlarda enerji tüketimini optimize etmeye yardımcı olurken, çevre koruma alanında da su kalitesi izleme, hava kirliliği izleme gibi uygulamalarda kullanılmaktadır. Eğitim sektöründe ise YZ, öğrenci performansını izleme, kişiselleştirilmiş öğrenme deneyimleri sunma ve eğitim materyallerini geliştirme gibi uygulama alanlarına sahiptir. YZ, sağlık sektöründe, hastalıkların erken teşhisi, tedavi planlaması, tıbbi görüntü işleme ve ilaç keşfi gibi konularda kullanılabilmektedir [1].

YZ görüldüğü üzere, veri yoğun ve karar verme gerektiren hemen hemen her alanda kullanılabilecek geniş bir potansiyele sahiptir. Ancak, YZ'nin kullanımı bazı dezavantajları da beraberinde getirmektedir. Özellikle, derin öğrenme gibi karmaşık YZ modelleri, genellikle karar alma süreçlerinin nedenlerini açıklamakta zorlanmaktadır ve bu da güvenilirlik, şeffaflık ve yorumlanabilirlik eksikliklerine yol açabilmektedir. [2]. YZ'nin karar alma süreçlerinde nihai karara ulaşmasına olanak sağlayan yapı kara kutu modeller olarak nitelendirildiği için, alınan kararların gerekçelerinin son kullanıcı tarafından anlaşılabilirliği bir sorun olarak karşımıza çıkmakta, bu da güvenilirlik noktasında eksikliklere neden olmaktadır. Bu eksiklikleri gidermek için Açıklanabilir Yapay Zeka geliştirilmiştir [3]. AYZ, yapay zeka modellerinin karar verme süreçlerini şeffaf hale getirerek kullanıcıların modelin işleyişini anlamasına ve alınan kararların güvenilir biçimde değerlendirilmesine olanak sağlamaktadır. AYZ için kullanılan teknikler arasında LIME (Local Interpretable Model-agnostic Explanations), SHAP (SHapley Additive exPlanations) ve ICE (Individual Conditional Expectation) gibi yöntemler bulunmaktadır [4].

Sağlık alanında, YZ teknolojisinin geniş bir kullanım alanı mevcuttur. Hastalara daha iyi bakım sağlamak ve sağlık hizmetlerini optimizasyonu, hastalıkların erken teşhisi, tedavi planlaması, hasta monitörizasyonu, tıbbi görüntüleme analizi ve ilaç keşfi gibi birçok alanda sıklıkla kullanılmaktadır [5]. Ancak, bu alanlarda kullanılan karmaşık YZ modellerinin karar alma süreçlerini açıklamakta zorlanması, sağlık profesyonelleri ve hastalar arasında güven sorunlarına neden olabilir. Bu nedenle AYZ'nin sağlık alanında kullanımı eksikliklerin giderilmesini sağlar. AYZ, karmaşık yapay zeka modellerinin karar alma süreçlerini açıklayarak, sağlık profesyonellerine ve hastalara kararların nedenlerini anlama ve güvenle değerlendirme olanağı sunar. Örneğin, bir hastalığın teşhis edilmesinde kullanılan AYZ modelleri, hangi belirtilerin teşhis için belirleyici olduğunu ve teşhisin nasıl yapıldığını anlatarak, doktorların hastalık durumunu daha iyi anlamalarına yardımcı olabilir [6]. Bu da tedavi planlaması ve hasta bakımında daha iyi kararlar alınmasına olanak tanımaktadır. Sağlık alanında, AYZ'nin kullanımı sadece hastaların daha iyi bakım almasını sağlamakla kalmaz, aynı zamanda sağlık hizmetlerinin etkinliğini artırır, maliyetleri azaltır ve sağlık profesyonellerinin karar alma süreçlerinde daha güvenli hissetmelerini sağlar. Bu nedenle, sağlık alanında AYZ'nin kullanımı giderek daha büyük bir önem kazanmaktadır [7].

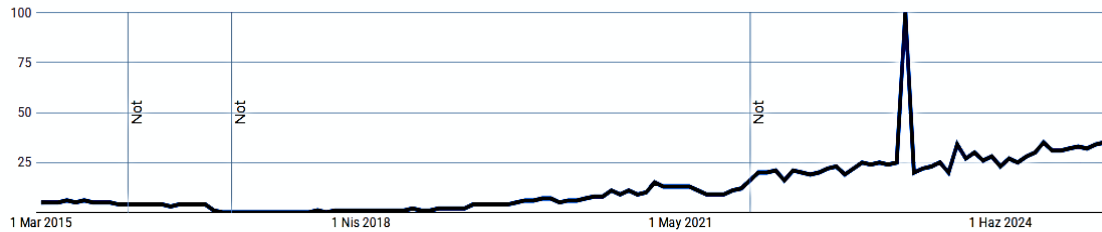
AYZ'nin sağlık alanında kullanımı da hızla yaygınlaşmaktadır. Hastalık teşhisi için literatürde çok farklı örnekleri bulunmaktadır. Maymun Çiçeği deri lezyonlarının sınıflandırılmasında [8], tüberküloz teşhisinde [9], beyin tümörü tespitinde [10], Alzheimer hastalığının sınıflandırılmasında [11], akciğer kanseri tespiti [12] vb. çalışmalarda çoğunlukla da görüntü işleme ile entegre AYZ teknikleri kullanılmaktadır. Ayrıca ameliyat yöntemlerinin değerlendirilmesi aşamasında da literatür hızla genişlemekte ve ameliyat yöntemlerinin hangisinin tercih edileceği noktasında hekimlere açıklanabilir, şeffaf kararlar sunulmaktadır [13]. Sağlık alanında AYZ uygulamaları ile ilgili detaylı literatür taramaları için Yang [14], Tjoa ve Guan [15] ve Sun ve arkadaşları [16]'nın çalışmaları irdelenebilir.

Mevcut literatürde sağlık alanında AYZ uygulamaları ağırlıklı olarak görüntü işleme temelli tanı sistemlerine odaklanmaktadır. Bununla birlikte, klinik ve demografik parametrelerin açıklanabilir modellerle değerlendirilmesine yönelik çalışmalar sınırlıdır. Bu boşluğu gidermek amacıyla, bu çalışma meme kanseri hastalarının hayatta kalma durumunu makine öğrenmesi algoritmaları ile sınıflandırmakta ve en başarılı modelin sonuçlarını SHAP yöntemi ile yorumlamaktadır. Çalışmanın özgün katkısı, yalnızca doğru tahminler yapmakla kalmayıp, aynı zamanda tahminlerin hangi faktörlerden kaynaklandığını şeffaf biçimde ortaya koymasındır. Böylece sağlık alanında AYZ tabanlı karar destek sistemlerine duyulacak güvenin artırılması hedeflenmektedir.

Bu çalışmada AYZ'nin sağlık alanında ne amaçla ve nasıl kullanılabileceğini açıklamak için detaylı bir örnek verilecektir. Çalışmanın 2. Bölümünde kullanılacak olan yöntemlerden bahsedilecek olup, 3. Bölümde sağlık alanında örnek bir AYZ uygulaması yapılacaktır. 4. Bölümde sonuçlar irdelenecektir. 5. Bölümde gelecek çalışmalardan bahsedilecektir.

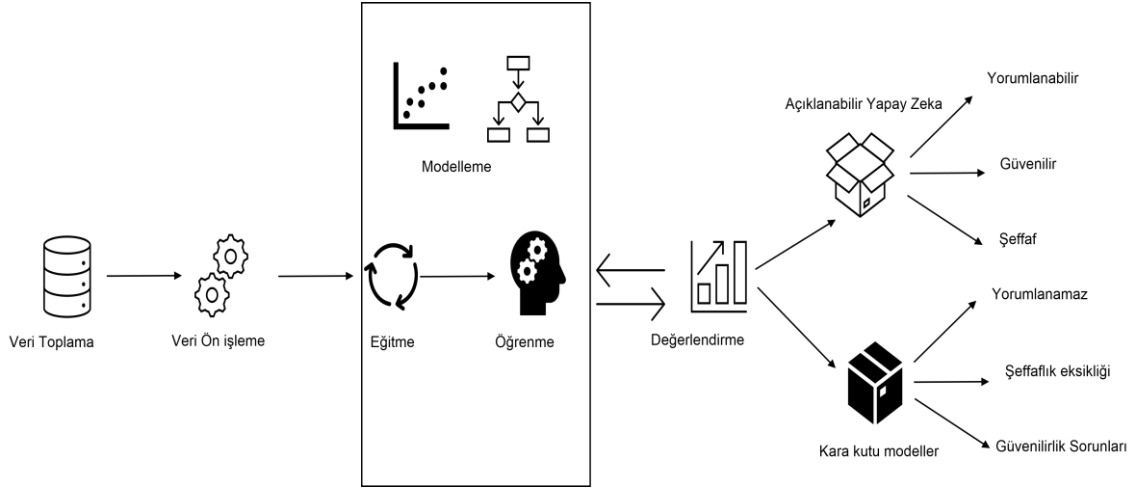
2. YÖNTEM

Şeffaflık, yorumlanabilirlik ve açıklanabilirlik AYZ'nin kara kutu modellere nazaran sağladığı en önemli avantajlardır. Kara kutu modeller, son kullanıcıların modelin davranışını anlamasını zorlaştırmakla beraber, potansiyel önyargıları veya hataları tespit etmelerini veya modeli kararlarından sorumlu tutmalarını engeller. Bu eksikliklerin giderilmesi amacı ile kullanımı gittikçe yaygınlaşan AYZ, akademik çalışmalar ve uygulayıcılar tarafından yaygın şekilde kullanılmaya başlanmıştır. Google Trend verilerinden elde edilen bilgilere göre son 10 yıl için dünya genelinde “Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence)” terimine gösterilen ilgi Şekil 1’de görüldüğü üzere 2018 sonrasında keskin bir şekilde artış göstermiştir. Bu artışın arkasında Avrupa Birliği'nin GDPR (Genel Veri Koruma Tüzüğü) gibi düzenlemelerle otomatik karar alma mekanizmalarında şeffaflığı zorunlu kılması, yapay zekanın sağlık ve finans gibi yüksek riskli alanlarda yaygınlaşmasıyla hesap verilebilirliğe olan ihtiyacın artması ve etkili açıklanabilirlik yöntemlerinin literatüre kazandırılması gibi önemli etkilerin olduğu düşünülmektedir.



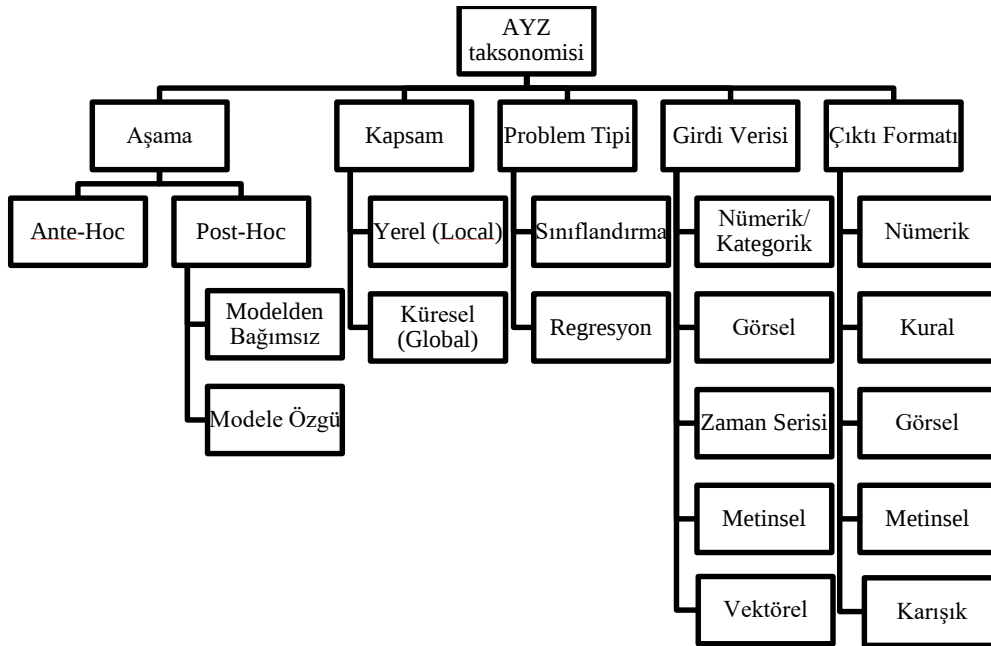
Şekil 1. AYZ Google Trend verileri

AYZ terimi literatürde ilk olarak simülasyon oyunları uygulamalarında yapay zeka tabanlı entitilerin (nesnelerin) davranışlarını açıklamak için Van Lent vd.'nin çalışmasında tanımlanmıştır [17]. Genel olarak AYZ Şekil 2’de gösterilen bir akışa sahiptir. İlk olarak, veri toplama ve ön işleme adımları gerçekleştirilmektedir. Bu aşamada veriler temizlenir, eksik değerler düzenlenir ve modellemeye uygun hale getirilir. İkinci aşamada, eğitim, test ve doğrulama süreçleri ile makine öğrenmesi veya derin öğrenme temelli kara kutu modelleri oluşturulur. Ancak bu modellerin iç işleyişi kullanıcı tarafından doğrudan görülemez veya kolayca yorumlanamaz. Bu modeller yüksek doğruluk oranları sağlayabilse de, karar süreçlerinin neden ve nasıl oluştuğu çoğunlukla anlaşılamamaktadır. Üçüncü aşamada, modelin performansı çeşitli istatistiksel ölçütler (doğruluk, hassasiyet, duyarlılık vb.) ile değerlendirilmektedir. Ancak kara kutu yapının temel sınırlılığı, yalnızca bu performans ölçümleri ile yetinilmesi ve modelin karar mantığının kullanıcıya açıklanamamasıdır. Oluşturulan YZ modelinin başarısının değerlendirilmesinin ardından AYZ için uygun model (LIME, SHAP, ICE vb.) yöntemleri seçilir ve uygulanır [18] [19]. Son aşamada devreye giren AYZ teknikleri, oluşturulan modellerin karar mekanizmasını şeffaflaştırmayı amaçlamaktadır. Bu sayede, kullanıcılar modelin hangi girdilerden nasıl etkilendiğini anlayabilmekte ve sonuçları daha güvenilir biçimde yorumlayabilmektedir [20].



Şekil 2. Kara kutu modeller ve AYZ

Son yıllarda hızla büyüyen bir literatüre sahip olan AYZ'ye ait taksonomi Şekil 3'te verilmiştir [21]. Bu çalışmada modellerin açıklanabilirliği için kullanılan SHAP yöntemi tercih edilmiştir. SHAP yönteminin tercih edilmesinin başlıca nedenlerinden biri sağlık alanındaki modellerde yalnızca lokal değil, küresel açıklamaların da yapılabilmesinin büyük önem taşıyor olmasıdır [22]. SHAP, hem tüm hasta grubu için genel eğilimleri hem de tek bir hasta için bireysel tahminleri yorumlama imkânı sunmaktadır. Ayrıca SHAP, özellikler arasındaki doğrusal olmayan ilişkileri daha iyi yakalayabilmekte [23] ve bu durum karmaşık yapıli klinik verilerin açıklanmasında ek avantaj olarak karşımıza çıkmaktadır. Bu nedenlerle, çalışmada açıklanabilirlik yöntemi olarak SHAP kullanılması uygun bulunmuştur. Kullanılacak olan SHAP AYZ tekniğinin detayları Bölüm 2.1'de verilmiştir.



Şekil 3. AYZ taksonomisi [21]

2.1. SHAP

Makine öğrenimi modellerinin karar verme süreçlerini şeffaf hale getirmek için geliştirilen SHAP yöntemi, oyun teorisine dayalı olarak her bir özelliğın tahmine katkısı adil biçimde hesaplamayı amaçlar [24]. Bu yaklaşım, Shapley [25] tarafından ortaya konulan Shapley değerlerini temel alır ve her bir özelliğın (özellikle i) model çıktısına yaptığı marjinal katkıyı, tüm olası özellik kombinasyonları üzerinden ortalama olarak tanımlar. SHAP değeriğın matematiksel formülasyonu Eşitlik 1'deki gibidir:

$$\phi_i(f) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (1)$$

Bu denklemde F özellikler kümesini, S ise bu kümenin i 'yi içermeyen bir alt kümesini temsil eder. Ayrıca $\phi_i(f)$, özellik i 'nin katkısı; $f(S)$ ise S altkümesinin modele sağladığı katkıyı ifade etmektedir. Eşitlik, i özelliğinin modele ne kadar katkı sağladığını hesaplamak için, i 'nin dahil edilip edilmemesi durumundaki model çıktısı farklarının ortalamasını alır [26]. Bu yaklaşım, modelin açıklanabilirliğini artırırken karar destek sistemlerinde güvenilirlik açısından önemli avantajlar sunmaktadır.

3. UYGULAMA ÇALIŞMASI

Bu uygulama çalışmasında, meme kanseri hastalarının teşhis aldıktan sonra ölü/ hayatta kalma durumlarının veri madenciliği teknikleri ile tahmini ve ardından modelde ele alınan hasta özelliklerinin model tahmini üzerinden etkisi ve modelin sonuçlarının AYZ teknikleri ile açıklanabilirliği üzerine odaklanılmıştır.

Kullanılan veri seti açık bir kaynaktır ve meme kanseri hasta veri seti, Ulusal Kanser Enstitüsü'nün (National Cancer Institute) nüfus tabanlı kanser istatistikleri sağlayan SEER Programı'ndan elde edilmiştir. Ulusal Kanser Enstitüsü'nün veri tabanında bilimsel çalışmalarda kullanılmak üzere paylaşılmış olan veri seti, 2006-2010 yılları arasında infiltratif duktal ve lobüler karsinom meme kanseri tanısı almış hastaların kayıtlarını içermektedir [27]. Veri seti, meme kanseri teşhisi sonrası 3408 sağ kalım, 616 ölüm olmak üzere 4024 hastaya ait 14 farklı parametrenin bilgilerini içermektedir.

Veri setine ilişkin özellikler aşağıda verilmiştir:

A. Girdi değişkenleri:

1. Demografik Özellikler:

- Yaş, Irk ve Medeni Durum

2. Tümör Özellikleri:

- Tümör Boyutu: Tümörün boyutunu temsil etmektedir.
- T Evresi: Tümörün T evresini (tümör büyüklüğü ve yayılım ölçüsü) temsil etmektedir.
- N Evresi: Tümörün N evresini (lenf düğümü tutulumu ölçüsü) temsil etmektedir.
- 6. Evre: Kanserın genel evresini (T, N ve M evrelerinin kombinasyonu) temsil etmektedir.
- Derece (Grade): Tümör derecesini (kanser hücrelerinin ne kadar anormal görüldüğünü) temsil etmektedir.
- Farklılaşma: Tümör hücrelerinin ne kadar farklılaştığını temsil etmektedir.

3. Hormonal Durum:

- Östrojen Durumu: Tümörün östrojen reseptör durumunu temsil eder, sayısal olarak kodlanmıştır.
- Progesteron Durumu: Tümörün progesteron reseptör durumunu temsil eder, sayısal olarak kodlanmıştır.

4. Bölgesel Lenf Düğümleri:

- Bölgesel Lenf Düğümü İncelendi: İncelenen bölgesel lenf düğümlerinin sayısını temsil etmektedir.
- Bölgesel Lenf Pozitif: Kanser açısından pozitif olan bölgesel lenf düğümlerinin sayısını temsil etmektedir.

5. A Evresi:

- A Evresi: Kanserın A evresini (metastaz ölçüsü) temsil eder, sayısal olarak kodlanmıştır.

B. Çıktı Değişkeni:

Durum: Hastanın hayatta kalma durumunu temsil etmektedir (0: Ölü, 1: Hayatta).

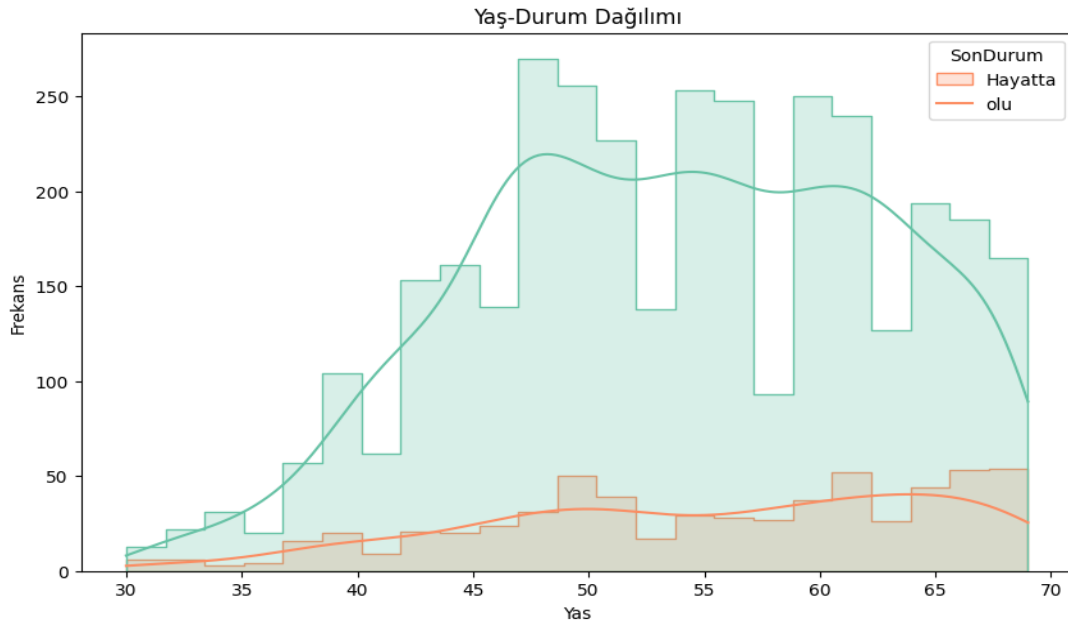
3.1. Veri Ön İşleme

Çalışmanın tüm kodlamaları Python programlama dili aracılığı ile yapılmıştır. Öncelikle veri seti veri ön işleme aşamasına tabii tutulmuştur. Ön işlemede girdi ve çıktı sütunlarında yer alan verilerin yapısı tespit edilmiştir. Verilerin tamsayı, kategorik vs. gibi sınıflandırmaları yapılmıştır. Boş hücre olup olmadığı da eş zamanlı olarak kontrol edilmiştir. Çizelge 1'de verilerin özellikleri görülmektedir.

Çizelge 1. Veri yapıları

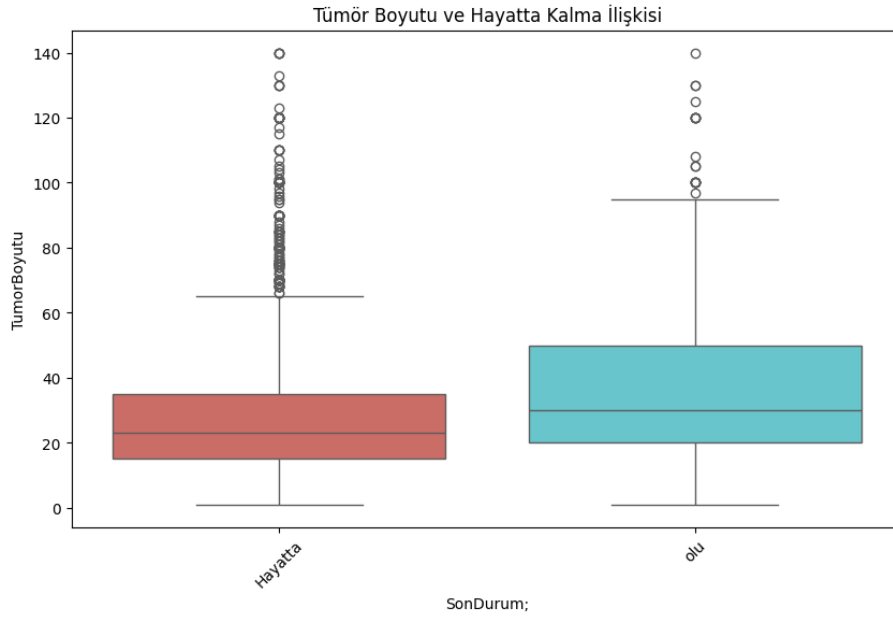
Değişkenler	Dolu hücre	Veri tipi
Yaş	4024	Nümerik
İrk	4024	Kategorik
Medeni durum	4024	Kategorik
T_evresi	4024	Kategorik
N_evresi	4024	Kategorik
6.Evre	4024	Kategorik
Farklılaşma	4024	Kategorik
Derece	4024	Kategorik
A_evre	4024	Kategorik
Tümör boyutu	4024	Nümerik
Östrojen durumu	4024	Kategorik
Progesteron durumu	4024	Kategorik
Bölgesel düğüm incelendi	4024	Nümerik
Bölgesel düğüm pozitif	4024	Nümerik
Son durum	4024	Kategorik

Bazı girdi değişkenlerinin çıktı değişkeni ile ilişkisi görselleştirilmiş olup bunlar Şekil 4 ve Şekil 5'te verilmiştir. Yaş ve meme kanseri sonucu hayatta kalma durumu arasındaki ilişki için Şekil 4'e ve tümör boyutu ve hayatta kalma arasındaki ilişki için ise Şekil 5'e bakılabilir. Şekil 4'te yaşa bağlı ölüm artışı eğilimi olduğu görülmektedir. Ayrıca Şekil 5'te tümör boyutunun artmasının hayatta kalma üzerindeki olumsuz etkisi de gözlemlenebilmektedir.



Şekil 4. Yaş ve son durum değişkenleri arasındaki ilişki grafiği

Bu grafikler ham veriler yardımı ile her girdi değişkeni için çizilebilir. Ancak bu grafiklerde hastanın son durumu ile seçilen girdi değişkeni arasında bir korelasyon görsel olarak algılsa da hastanın sağ kalma/ölme durumu tek bir parametreden kaynaklı olmayıp birçok parametrenin etkileşimi ile ortaya çıkabilecek karmaşık bir olaydır. Farklı parametre etkileşimleri ile YZ yardımı ile herhangi bir vakanın durumunun ne olacağının öngörülmesi ve AYZ yardımı ile kurulan modellerin hangi parametrelerinin ne seviyede olmasının sağ kalma/ölme durumu üzerinde etkili olduğunun irdelenmesi gerekmektedir.



Şekil 5. Tümör boyutu ve son durum değişkenleri arası ilişki grafiği

Veri ön işlemede önemli safhalardan birisi verilerin kodlanmasıdır. Veriler kategorik ya da nümerik gibi farklı yapılarda olabilir. Kategorik değişkenler, birçok makine öğrenimi modelinde olduğu gibi doğrudan analiz için kullanılamadığından, her model için uygun bir formata dönüştürülmelidir. Bir değişkenin bir metin (string) olarak temsil edilmesine ek olarak, eğer değerlerin büyüklüğü veya sıralaması sayısal olarak anlam ifade etmiyorsa, bu değişkenin kategorik bir tür olarak ele alınması gerekir. Kategorik değişkenleri kodlamak için One-Hot Encoding (Tek Sıcaklık Kodlaması), Dummy Encoding (Sahte Değişken Kodlaması), Ordinal Encoding (Sıralı Kodlama), Binary Encoding (İkili Kodlama), Count Encoding (Frekans Kodlama) ve Target Encoding (Hedef Kodlama) gibi çeşitli yöntemler kullanılmaktadır [28]. Verinin yapısına göre farklı kodlama teknikleri seçilebilir. Bu çalışma özelinde değişkenlerin yapısı dikkate alınarak sıralı, ikili ve tek sıcaklık kodlaması yöntemleri uygulanmıştır.

- **Sıralı Kodlama:**

A evresi: Bölgesel, metastaz ve uzak metastaz olmak üzere üç kategoriden oluşmaktadır.

T evresi: T1, T2, T3, T4 olmak üzere dört kategoriden oluşmaktadır.

N evresi: N1, N2, N3 olmak üzere üç kategoriden oluşmaktadır.

6. Evre: IIB, IIIA, IIIB, IIIC olmak üzere dört kategoriden oluşmaktadır.

Farklılaşma: Farklılaşma yok, Çok az farklılaşma, Orta dereceli farklılaşma, Çok fazla farklılaşma olmak üzere dört kategoriden oluşmaktadır.

Kodlama aşamada bu kategoriler sayısal değerler ile değiştirilmiştir. Grup sayısı kadar kod numarası verilmiştir. Örneğin 6. Evre: IIB, IIIA, IIIB, IIIC gruplarından oluşmaktadır. Sırası ile gruplar 1,2,3,4 kod numaraları ile değiştirilmiştir.

- **İkili kodlama ile:**

Östrojen ve Progesteron durumu Pozitif ve Negatif olmak üzere iki kategoriden ve Son durum değişkeni ise Hayatta ve Ölü olmak üzere iki kategorilerinden oluşmaktadır. Bu değişkenler 0-1 şeklinde kodlanmıştır.

- **Tek sıcaklık kodlama ile** Irk ve Medeni durum değişkenleri kodlanmıştır.

Irk ve medeni durum için aşağıda verilen kategoriler yeni birer öznitelik gibi ele alınmış ve sadece hangi satırda bu kategoriler bulunuyorsa o satıra "1" değeri verilmiş, diğerleri "0" şeklinde kodlanmıştır. Irk değişkeni: Irk Beyaz, Irk Diğer, Irk Siyah; Medeni Durum değişkeni: Medeni Durum Ayrı, Medeni Durum Bekar, Medeni Durum Dul, Medeni Durum Evli, Medeni Durum Boşanmış sütunlarına dönüştürülmüştür. Nümerik veriler için de ön işleme gerekmektedir. Amacı, özellikleri benzer bir ölçeğe getirmek olan normalizasyon, sayısal verileri diğer sayısal verilerle aynı ölçeğe dönüştürmek anlamına gelmektedir. Bu, modelin performansını ve eğitim kararlılığını artırmaktadır [29]. Nümerik değişkenleri normalize için L1-

L2 normalizasyonu, bir aralığa ölçeklendirme (Min-Max), kırpma (Clipping), Logaritmik ölçeklendirme (Log scaling) ve Standardizasyon/Z-skoru (Z-score) gibi yöntemler bulunmaktadır. Bu çalışmada sayısal veriler için yaş, tümör boyutu, bölgesel düğüm gibi özellikler min-max normalizasyon yöntemi ile normalize edilmiştir. Kullanılan min-max normalizasyon formülü Eşitlik 2'deki gibidir:

$$norm_x = \frac{x - \min}{\max - \min} \quad (2)$$

Veri ön işleme aşamasının tamamlanmasının ardından tüm veriler kategorik ve normalize edilmiş sürekli değerlere dönüştürülmüştür. Bu aşama sayesinde, daha sağlıklı YZ modellemesi yapılabilecektir. Çizelge 2'de, verilerin bir kısmına ait ön işleme öncesi ve sonrası değerler görülmektedir.

Çizelge 2. Veri ön işleme öncesi (a) ve sonrası (b)

a) Veri ön işleme öncesi										
Yaş	İrk	Medeni durum	T evresi	N evresi	6.Evre	Farklılaşma	Derece	A evre	Östrojen durumu	
68	Beyaz	Evli	T1	N1	IIA	Çok az farklılaşma	3	Bölgesel	Pozitif	
50	Beyaz	Evli	T2	N2	IIIA	Orta dereceli farklılaşma	2	Bölgesel	Pozitif	
58	Beyaz	Boşanmış	T3	N3	IIIC	Orta dereceli farklılaşma	2	Metastaz	Negatif	
58	Siyah	Evli	T1	N1	IIA	Çok az farklılaşma	3	Bölgesel	Pozitif	
47	Beyaz	Boşanmış	T2	N1	IIB	Çok az farklılaşma	3	Bölgesel	Pozitif	

b) Veri ön işleme sonrası										
Yaş	İrk Beyaz	Medeni durum Boşanmış	Medeni durum Evli	T evresi	N evresi	6.Evre	Farklılaşma	Derece	A Evre	Östrojen durumu
0,71	1	0	1	1	1	1	2	3	1	1
0,45	1	0	1	2	2	3	3	2	1	1
0,51	1	1	0	3	3	5	3	2	2	0
0,51	0	0	1	1	1	1	2	3	1	1
0,41	1	1	0	2	1	2	2	3	1	1

3.2. Makine Öğrenmesi

Çeşitli değişkenlere göre çıktı değerine ait bir sınıfın tahmin edildiği modeller sınıflandırma modelleri olarak adlandırılmaktadır. Bu çalışmada makine öğrenmesinde sınıflandırma amacı için geliştirilmiş olan Rassal Orman Sınıflayıcı (ROS), Gradient Boosting Sınıflayıcı (GBS), Lojistik Regresyon (LR), Karar Ağaçları (KA), Destek Vektörü Sınıflayıcı (DVS) ve K-En yakın Komşu (KNN) sınıflayıcı modelleri kullanılmış ve performans yönünden modeller birbiri ile kıyaslanmıştır. Kullanılan yöntemlerin genel çalışma mekanizmaları şu şekilde tanımlanabilir: Lojistik Regresyon, ikili sınıflandırma problemlerinde sık kullanılan, doğrusal ilişkilere dayalı olasılıksal bir modeldir [30]. Karar Ağaçları, veriyi dallara ayırarak sınıflandıran, yorumlanabilirliği yüksek, hiyerarşik yapıda bir yöntemdir [31]. Rassal Orman Sınıflayıcı, birden fazla karar ağacının topluluk (ensemble) yöntemiyle birleştirilmesiyle oluşur; aşırı öğrenmeyi azaltarak daha kararlı sonuçlar üretilmesine olanak tanır [32]. Gradient Boosting Sınıflayıcı, hataları ardışık olarak düzelten zayıf öğrenicilerin (weak learners) birleşiminden oluşan güçlü bir topluluk yöntemidir [33]. Destek Vektörü Sınıflayıcı, sınıflar arasındaki ayrımı en iyi sağlayacak hiper düzlemi bulmaya çalışır; özellikle karmaşık ve yüksek boyutlu verilerde etkili olmaktadır [34]. K-En Yakın Komşu algoritması ise, yeni bir örneği, eğitim verisindeki en yakın komşularının sınıfına göre sınıflandıran, basit ve örnek tabanlı bir algoritmadır [34].

Çalışmada kullanılan algoritmalar farklı yapısal özelliklere sahiptir ve birbirleriyle kıyaslanarak en iyi performans gösteren modelin belirlenmesi amaçlanmıştır. Örneğin, Lojistik Regresyon doğrusal ilişkilerin anlaşılmasına imkân tanırken, Rassal Orman ve Gradient Boosting algoritmaları değişkenler arası karmaşık ilişkileri yakalayabilmektedir. Karar Ağaçları yorumlanabilirlik açısından güçlüdür, KNN ise örnekler arasındaki benzerliklere dayalı tahminler yapabilmektedir. Destek Vektör Makineleri ise sınıflar arasındaki

sınırları optimize etmeye odaklanmaktadır. Bu çeşitlilik sayesinde, farklı bakış açıları sunan modellerin karşılaştırılması ve en uygun modelin seçilmesi hedeflenmiştir.

Modeller için eğitim ve test verisi bölümlendirmesi, %80 eğitime ve %20 test olacak şekilde yapılmıştır. Makine öğrenmesi aşamasında kullanılan sınıflandırma modellerine ait parametreler Python scikit-learn kütüphanesinde tanımlanan varsayılan değerler olarak kullanılmıştır. Veri seti, sınıf sayılarının 3408 sağ vaka ve 616 ölü vaka olması nedeni ile dengesiz bir yapıdadır. Dengesiz veri setlerinde model başarısı düşebilmekte, makine öğrenmesi algoritmalarında performansı artırmak için SMOTE (Synthetic Minority Over-sampling Technique) [35], Sınıf Ağırlığı (SA)[36] gibi farklı tekniklerin algoritmalara entegre edilmesi gerekmez. Bu çalışmada, dengesiz veri setinin tahmin performansına olan etkisini azaltmak için Sınıf Ağırlığı ve SMOTE yöntemlerine yer verilmiştir. Sınıf Ağırlığı yöntemi, modelin eğitim sırasında farklı sınıfların örneklerine farklı ağırlıklar vermesini sağlar. Azınlık sınıfının örneklerine daha yüksek ağırlık verilirken, çoğunluk sınıfının örneklerine daha düşük ağırlık verilir. Modelin kayıp fonksiyonu (loss function) hesaplanırken, her örneğin katkısı o örneğin ait olduğu sınıfın ağırlığı ile çarpılır. Daha yüksek ağırlık verilen azınlık sınıfı örneklerindeki yanlış tahminler, modelin toplam kaybına daha fazla katkıda bulunur. Bu da modelin eğitim sırasında azınlık sınıfı örneklerini doğru tahmin etmeye daha fazla odaklanmasını sağlar. SMOTE yöntemi ise dengesiz veri kümelerinde azınlık sınıfının örneklerini sentetik olarak artırmak için kullanılan popüler bir üst örnekleme (oversampling) tekniğidir [35]. Gradient Boosting ve KNN algoritmaları yapıları gereği sınıf ağırlığı modellerinin kullanımına uygun değildir. İki yaklaşım dikkate alınarak makine öğrenmesi modelleri uygulanmış, elde edilen sonuçlar Çizelge 3'te verilmiştir.

Çizelge 3. Makine öğrenmesi modelleri sonuç değerleri

Makine öğrenmesi tekniği	Yaklaşım	Doğruluk	Kesinlik (Ölü)	Duyarlılık (Ölü)	F1 skoru (Ölü)	Kesinlik (Hayatta)	Duyarlılık (Hayatta)	F1 skoru (Hayatta)
ROS	SA	0,85	0,54	0,13	0,21	0,86	0,98	0,92
LR	SA	0,70	0,27	0,62	0,38	0,91	0,72	0,80
DVS	SA	0,71	0,27	0,61	0,37	0,91	0,72	0,81
KA	SA	0,78	0,28	0,30	0,29	0,88	0,87	0,87
GBS	SMOTE	0,79	0,33	0,39	0,35	0,89	0,86	0,87
ROS	SMOTE	0,79	0,30	0,32	0,31	0,88	0,87	0,87
KNN	SMOTE	0,69	0,23	0,48	0,31	0,89	0,72	0,80
DVS	SMOTE	0,73	0,27	0,50	0,35	0,90	0,77	0,83
KA	SMOTE	0,75	0,25	0,34	0,28	0,88	0,82	0,85

Modellerin performanslarını değerlendirmek için doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru değerleri hesaplanmıştır. Bu metriklerin hesaplanmasında kullanılan temel kavramlar: Gerçekte pozitif olan (örneğin, “ölü” sınıfı) ve doğru tahmin edilen örnekleri ifade eden “TP (True Positive)”, Gerçekte negatif olan (örneğin, “hayatta” sınıfı) ve doğru tahmin edilen örnekleri ifade eden TN (True Negative), Gerçekte negatif olduğu halde pozitif tahmin edilen örnekleri ifade eden FP (False Positive) ve Gerçekte pozitif olduğu halde negatif tahmin edilen örnekleri ifade eden FN (False Negative) parametreleridir. Ardından bu parametrelere bağlı olarak modelleri birbiri ile kıyaslayabilmek adına aşağıdaki formüller yardımı ile doğruluk, kesinlik, duyarlılık ve F1 skoru değerleri hesaplanmıştır.

Tüm doğru sınıflandırmaların toplam örnek sayısına oranını ifade eden Doğruluk değeri Eşitlik 3'teki gibi hesaplanmaktadır.

$$\text{Doğruluk} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3)$$

Pozitif olarak tahmin edilenlerin içinde gerçekten pozitif olanların oranını ifade eden kesinlik değeri Eşitlik 4'teki gibi hesaplanmaktadır:

$$\text{Kesinlik} = \frac{(TP)}{(TP+FP)} \quad (4)$$

Gerçekte pozitif olanların içinde doğru tahmin edilenlerin oranını ifade eden Duyarlılık değeri Eşitlik 5 kullanılarak hesaplanmaktadır:

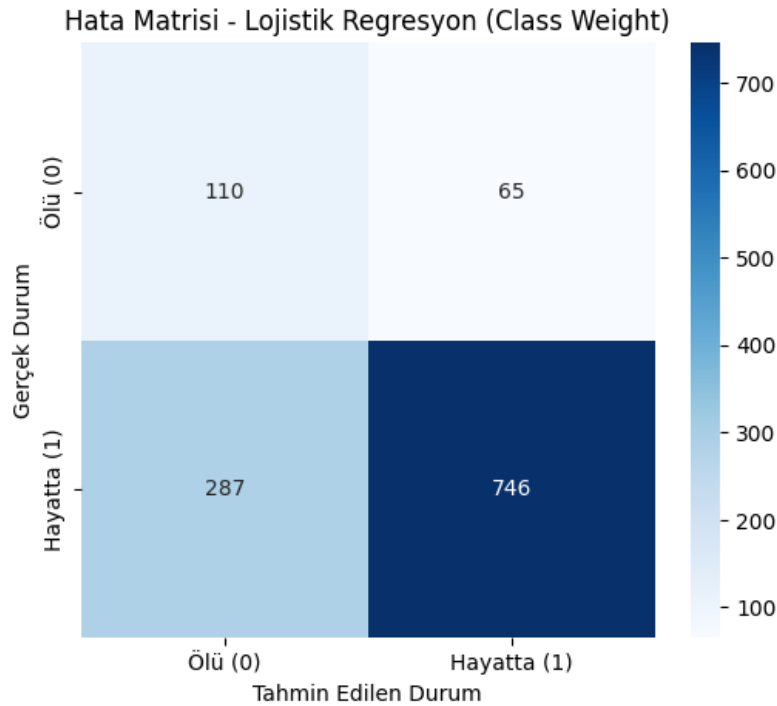
$$\text{Duyarlılık} = \frac{(TP)}{(TP+FN)} \quad (5)$$

Kesinlik ve duyarlılığın harmonik ortalamasını ifade eden F1 skoru değeri ise Eşitlik 6 ile hesaplanmaktadır:

$$F1 = 2X \frac{\text{Kesinlik} \times \text{Duyarlılık}}{(\text{Kesinlik} + \text{Duyarlılık})} \quad (6)$$

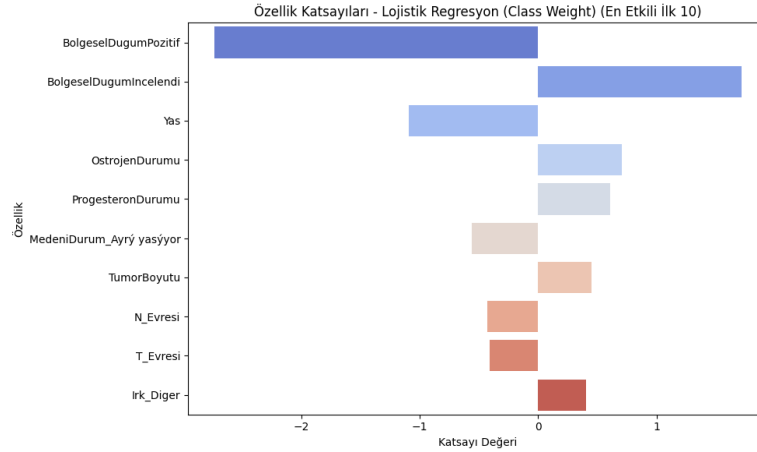
Bu ölçütler sayesinde modeller yalnızca doğruluk açısından değil, aynı zamanda azınlık sınıfı olan “ölü” grubunu doğru tahmin edebilme başarısı açısından da daha kapsamlı bir şekilde değerlendirilmiştir. Ayrıca özellikle dengesiz sınıf dağılımına sahip veri kümelerinde Doğruluk değerine göre model performansına karar vermek yanıltıcı olabilmektedir. Yüksek doğruluk, modelin sadece çoğunluk sınıfını doğru tahmin etmesinden kaynaklanabilir. F1 Skoru ise hem Duyarlılık hem de Kesinliği dikkate aldığı için dengesiz veri kümelerinde daha bilgilendiricidir. Yüksek bir F1 Skoru, modelin hem yanlış pozitifleri hem de yanlış negatifleri düşük tutarak iyi bir denge kurduğunu göstermektedir.

Çizelge 3’te görüldüğü üzere genel doğru tahmin açısından ve hayatta kalanların doğru sınıflandırılması açısından Rassal Orman Sınıflayıcı model daha başarılıdır. Dengesiz veri seti üzerinde performans artırıcı tüm çalışmalara rağmen sayıca az olan sınıfın tahminleme başarısı % 62’yi geçememektedir. Ancak bu değer SMOTE ve Sınıf Ağırlığı yöntemleri entegre edilmeden önce % 15 civarında olduğu tespit edilmiştir. Özellikle ele alınan veri setinde “Ölü” sınıf sayısının azlığı dikkate alındığında ve ayrıca ölü bir hastayı yanlış sınıfa atamanın maliyeti daha yüksek olacağı için ölü sınıfına ait duyarlılık değeri baz alınarak bir tercih yapılmalıdır. Şekil 6’da F1 skoru açısından en başarılı olan Lojistik Regresyon sınıflayıcı modeline ait hata matrisleri verilmiştir. LR hata matrisinde de görüldüğü üzere 746 sağ hasta sağ olarak tahmin edilirken, 110 hasta ölü hasta ölü sınıfında tahmin edilmiştir.



Şekil 6. Lojistik Regresyona ait hata matrisi

Makine öğrenmesi aşamasında LR modeli için özellik önemine ait bilgiler elde edebiliriz. Şekil 7’de ve Çizelge 4’te bu iki yönetime ait özellik önem grafiği ve değerleri görülmektedir.



Şekil 7. Özellik önem sıralaması grafiği

Şekil 7’de ve Çizelge 4’te LR sınıflandırma modelinin tahmininde en etkili parametrenin Pozitif Bölgesel Düğüm olduğu görülmektedir. Ardından sırası ile incelenen bölgesel düğüm sayısı, yaş, östrojen durumu, progesteron durumu gibi parametreler gelmektedir. Şekil 7’de LR modelinde model performansı üzerinde en etkili parametrelerin önem sırası ve ne derece önemli olduğu çubuk grafiklerle görselleştirilmiştir. Katsayı değeri negatif bölgede olan parametrelerin sağ kalım üzerinde olumsuz yönlü etkisi olduğu anlaşılmaktadır.

Çizelge 4. Özellik önem dereceleri

Lojistik regresyon		
	Özellik	Önem
1	Bölgesel düğüm pozitif	2,73
2	Bölgesel düğüm incelendi	1,71
3	Yaş	1,08
4	Östrojen durumu	0,70
5	Progesteron durumu	0,61
6	Medeni durum_ayrı	0,55
7	Tümör boyutu	0,45
8	N_evresi	0,43
9	T_evresi	0,40
10	Irk_diger	0,40

Makine öğrenmesi modellerinin performans ölçümleri, modelin ne kadar iyi tahmin ettiği ve nerede zorlandığı hakkında bilgi verirken; bu sonuçları modelin neden belirli bir tahmin yaptığını veya hangi özelliklerin belirli bir sonuç üzerinde en önemli etkiye sahip olduğunu açıklayamamaktadır. Burada devreye AYZ teknikleri girmektedir. SHAP gibi yöntemler, bireysel tahminler için "yerel" açıklamalar sunarak, hangi özelliklerin (örneğin yaş, tümör boyutu, evre vb.) belirli bir hastayı "Ölü" veya "Hayatta" olarak tahmin etmede en etkili olduğunu ve bu etkinin yönünü anlamamızı sağlamaktadır. Bu durum, yalnızca model performansını değerlendirmek için değil, aynı zamanda modele güven oluşturmak, çıktıları doğrulamak, olası önyargıları engellemek ve hatta alan uzmanlarının (doktorlar gibi) modelin tahminlerini yorumlamalarına ve kullanmalarına yardımcı olmak için de önemlidir.

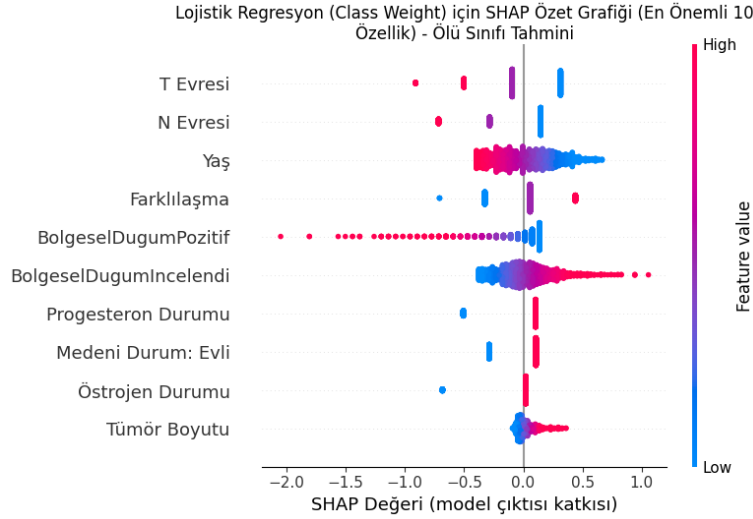
3.3. SHAP Tekniği ile Açıklanabilir Yapay Zeka Analizi

Bu aşamada, sadece “Ölü” sınıfını tahmin etmede başarı gösteren LR sonuçları AYZ tekniklerinden olan SHAP tekniği yardımı ile çıkarımlarda bulunulmuştur. LR katsayıları genel eğilimleri gösterirken, ROS gibi daha karmaşık modeller birer kara kutu gibi davranabilmektedir. Ancak SHAP, her bir tekil gözlem için, her bir özelliğin modelin nihai tahminine ne kadar ve hangi yönde katkıda bulunduğunu matematiksel olarak sağlam bir temelde açıklayarak yerel yorumlanabilirlik sağlamak ve model performansının ötesinde, güven oluşturmaya ve alınan kararları anlamlandırmaya yardımcı olmaktadır.

Bu bağlamda, hangi özelliklerin çıktığı ne yönde etkilediğini anlamak için SHAP özet tablosundan faydalanmak mümkündür. Şekil 8’de “Ölü” sınıfının tahmininde kullanılan LR modeli çıktıları üzerinden

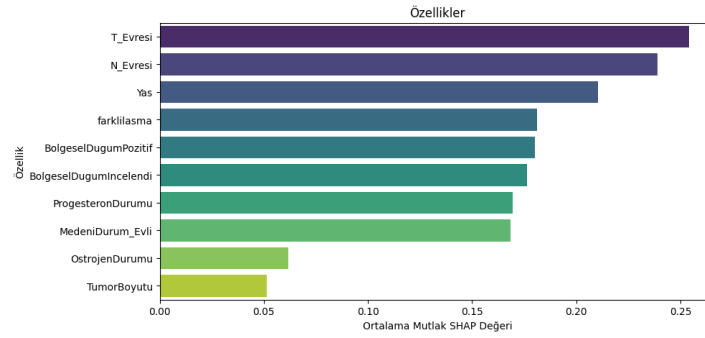
en önemli 10 özelliğe ait SHAP özet tablosu verilmektedir. Şekil 8’de eğer özellik nümerik bir girdi ise; kırmızı veriler ilgili özelliğin yüksek sayısal değerlerini ifade ederken, mavi veriler düşük sayısal değerleri ifade etmektedir. SHAP değerinde 0 değerinin sağ tarafı “Hayatta” sınıfının tahminini 1 olarak bulmamıza neden olan özellikleri ifade ederken sol tarafı “Hayatta” sınıfının değerini 0 olarak tahmin etmemize neden olan değerleri göstermektedir (“Ölü” olarak tahmin edilen sonuçlar). Ayrıca kategorik sınıflandırma yapılan özellikler için mavi renk, kod numarası düşük olan grubu ifade ederken, kırmızı renkli veriler kod değeri yüksek olan verileri ifade etmektedir. Bu sonuçlara göre Ölü sınıfı tahminlemede en başarılı olan LR modeline ait Şekil 8’de verilen SHAP özet tablosu yorumlandığında yapılan çıkarımlar aşağıdadır:

- T Evresi için T4, T3 ve T2 sırası ile (tümörün meme duvarına veya cilde yayılımının artmasını ifade eder) “Ölü” sınıfı tahminini artırmaktadır.
- N Evresi için N3, N2 sırası ile (kanserin daha uzak veya daha fazla sayıda lenf düğümüne yayıldığı durumu ifade eder) “Ölü” sınıfı tahminini artırmaktadır.
- Yaş değişkeni için ileri yaşlarda olmak “Ölü” sınıfı tahminini artırmaktadır.
- Farklılaşma değişkeni için “Farklılaşma yok” ve “Çok az farklılaşma” olması (kansere hücrelerinin normal meme hücrelerine çok az benzediği anlamına gelir ki bu agresif bir kanser türüne işaret eder [37]) “Ölü” sınıfı tahminini artırmaktadır.



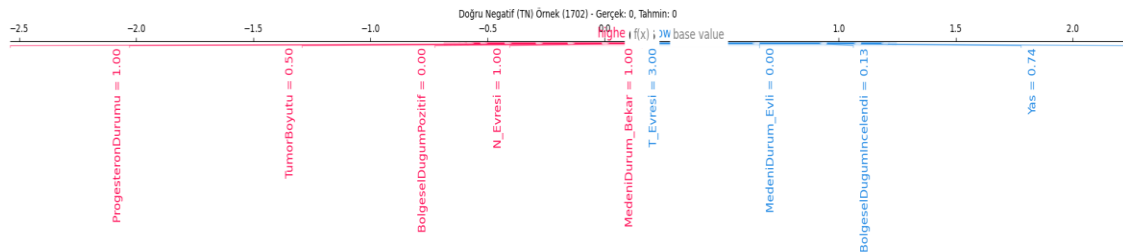
Şekil 8. Ölü sınıfının tahmini için kullanılan LR modeline ait en önemli 10 özellik için SHAP özet tablosu

- Bölgesel düğüm pozitif değişkeni için kanser hücrelerinin lenfatik sisteme yüksek yayılımını ifade eden değerler “Ölü” sınıfı tahminini artırmaktadır.
- Bölgesel düğüm incelendi değişkeni için daha az sayıda lenf düğümü incelemesi “Ölü” sınıfı tahminini artırmaktadır.
- Negatif progesteron ve östrojen değeri meme kanserinin hormonlara duyarsız olduğu ve bu nedenle hormonal tedavinin işe yaramadığı grup olarak “Ölü” sınıfı tahminini artırmaktadır.
- Tümör boyutu küçük değerler için “Ölü” grubuna ait tahmin olasılığı artıyor görülmektedir. Ancak bu durum tek başına değerlendirme için yeterli gözükmemektedir. Negatif progesteron ve östrojen değerli küçük bir tümör, daha büyük boyutlu ancak pozitif progesteron ve östrojen değerli bir tümörden daha agresif olabilmektedir.
- Evli olmayan grup için “Ölü” sınıfı tahmini artırmaktadır. Literatürde birçok çalışmada evli olmanın meme kanseri dahil olmak üzere birçok kanserde bağımsız bir sağkalım artırıcı faktör olduğu gösterilmiştir [38]. Ayrıca Şekil 9’da LR modeli için “Ölü” sınıfını tahmin etmede en etkili 10 özellik ve etki oranlarına ilişkin grafik görülmektedir. Buradan modelin “Ölü” değeri tahminlemede en çok etkilendiği özellikler önem sırasına göre görülmektedir. Elde edilen bulguların bilimsel gerçeklerle de uyumlu olduğu görülmektedir.

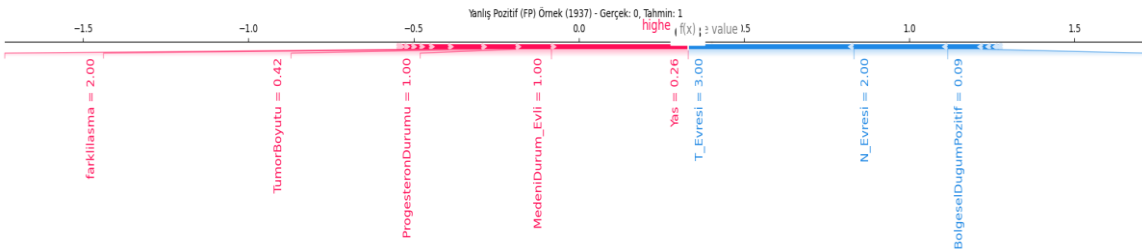


Şekil 9. SHAP analizine göre LR modelinin “Ölü” sınıfına ait tahmin yapmasında etkili olan 10 girdi ve ortalama mutlak SHAP değerleri

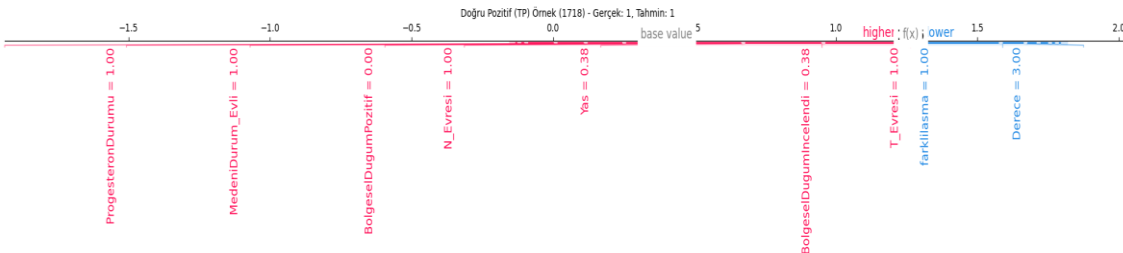
AYZ ayrıca, tekil bir örnek için modelin neden belirli bir tahmin ürettiğini derinlemesine anlamak için çok kullanışlıdır. Bu durumda karar grafiğinden faydalanabilir. Şekil 10 (a, b, c ve d)'da belirli hastaların çıktısının tahmin edilmesine etki eden parametreler ve ne yönde etki ettiği görülmektedir. Karar grafiği üzerinde tahmin hattının sağa doğru ilerlemesine neden olan özellik değerlerinin, bu özelliği “Ölü” sınıfı olarak tahmin etme yönünde katkı sağladığını göstermektedir.



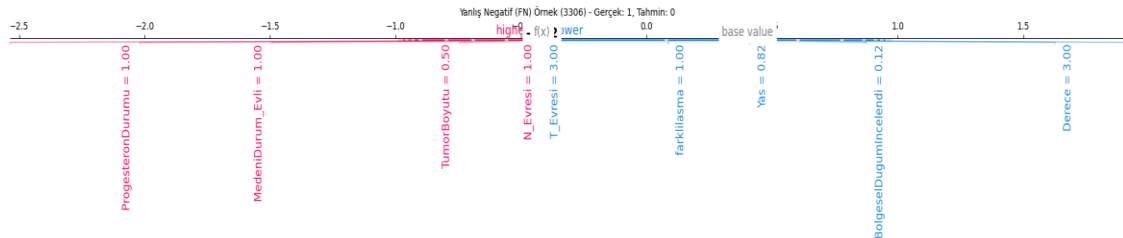
(a). Doğru negatif tahmin; örnek no: 1702



(b). Yanlış pozitif tahmin; örnek no: 1937



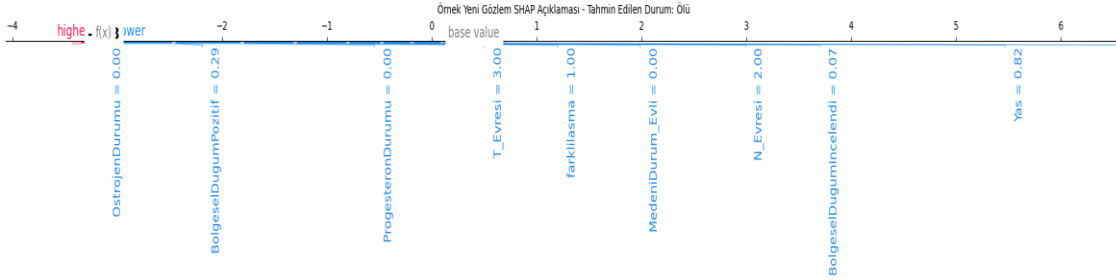
(c). Doğru pozitif tahmin; örnek no: 1718



(d). Yanlış negatif tahmin; örnek no: 3306

Şekil 10. Tahmin sonuçları üzerine girdi değişkenlerinin etkisini gösteren SHAP kuvvet grafiği

Ayrıca yeni gelen bir vakanın belirlenen en iyi makine öğrenmesi algoritmasına göre sonucunun ne olacağı ve ardından bu sonucun tahmin edilmesinde etkili olan girdi değişkenlerinin neler olduğu da SHAP yardımı ile öngörülebilmektedir. Bu bağlamda belirli girdi değerlerine sahip yeni bir hasta için modelin tahmini ve bunu etkileyen girdi değerleri incelenebilmektedir. Örnek hastaya ait özellikler şu şekildedir: Yaş: 62, Irk: Beyaz, Medeni Durum: Dul, Tümör Boyutu: 38, Derece: 3, T Evresi: T3, N Evresi: N2, 6.Evre: IIIC, A Evre: Bölgesel, Östrojen Durumu: Negatif, Progesteron Durumu: Negatif, Farklılaşma: Çok az farklılaşma, Bölgesel Düğüm İncelendi:5, Bölgesel Düğüm Pozitif:14. Belirtilen yeni özelliklere sahip vaka için LR-SA yöntemi sonucu “Ölü” sınıfı olarak tahmin edilmiştir. Şekil 11’de SHAP kuvvet grafiği analizi ile sonuca etki eden parametreler irdelenmiştir. Vakanın “Ölü” sınıfında tahmin edilmesinin en önemli etkenleri olarak Bölgesel Düğüm Pozitif değişkeninin değeri, Progesteron ve Östrojen varlığının negatif olması tespit edilmiştir.



Şekil 11. Belirli özelliklere sahip bir vakanın tahminine etki eden değişkenlere ait SHAP kuvvet grafiği

4. TARTIŞMA VE SONUÇ

Bu çalışmada elde edilen bulgular, literatürdeki benzer araştırmalarla karşılaştırıldığında genel olarak literatüre uyumluluk göstermektedir. Öncelikle, literatürde belirtildiği gibi [39] dengesiz veri setleri ile yapılan çalışmada SMOTE tekniğinin kullanımının, meme kanseri hastalarının sağkalım tahmininde kullanılan makine öğrenmesi algoritmalarının başarısını artırdığı yönünde bulgular elde edilmiştir. Ayrıca SHAP'ın meme kanseri hastalarında sağkalım tespitinde daha önceki çalışmalarda da belirtilen etkin kullanımını [40] sayesinde makine öğrenmesi modellerinden klinik olarak anlamlı sonuçlar çıkarılmıştır. SHAP yöntemi ile elde edilen çıkarımlar, Şekil 8 ve 9'da da verildiği üzere literatürde meme kanseri sağkalımına etki eden risk faktörleriyle benzerlik göstermektedir. Yaş, T evresi, N evresi gibi değişkenlerin model çıktıları üzerinde belirleyici etkilerinin bulunduğu bulgusu, mevcut literatürde rapor edilen sonuçlarla örtüşmektedir [41]. İlave olarak medeni durumun da bulguları destekler şekilde sağkalım üzerinde etkisi olduğu tespit edilmiştir [38]. Bu durum, AYZ'nin meme kanseri sağkalım tahminlemede ve nedenlerini açıklamada başarılı sonuçlar verdiğinin göstergesidir. Ayrıca genelleme yapılacak olursa, sağlık alanında AYZ kullanımı sayesinde hastalar, hastalıklar ve tedaviler ile ilgili özel ve genel çıkarımlar yapılabileceği anlaşılmaktadır.

AYZ, makine öğrenmesi ile oluşturulan modellerin çalışma mekanizmasını anlamamız, böylece model tahminlerinin neden oluştuğunu anlamlandırmamız açısından oldukça önemlidir. Bu modeller sayesinde daha şeffaf ve anlamlandırılabilir çıkarımlar yapmak mümkündür. Bu durum özellikle sağlık gibi yanlış karar vermenin maliyetinin yüksek olduğu alanlarda karar vericilere önemli bir katkı sağlamaktadır. Literatür örneklerinden de anlaşıldığı üzere çok farklı alanlarda kara kutu modelleri açıklama ve anlamlandırma yeteneği olan AYZ, sağlık alanında tanı, teşhis, tedavi, girişimsel ve girişimsel olmayan tedavilerin sonuçlarının tahmini için yaygın olarak kullanılmaya başlanmıştır.

Bu çalışmada öncelikle sağlık alanında AYZ'nin hangi amaçlarla kullanıldığına ilişkin literatür sunulmuş, ardından dengesiz sınıf kümelerinden oluşan meme kanseri verileri üzerinde çeşitli makine öğrenmesi algoritmaları uygulanmıştır. Veri setinin dengesiz sınıf dağılımının negatif etkilerini gidermek için SMOTE ve sınıf ağırlığı yöntemlerinden faydalanılmış bu yöntemler sayesinde model performansları önemli ölçüde iyileştirilmiştir. Modellerin başarısı özellikle F1 skoru baz alınarak değerlendirilmiş, genel doğruluk açısından en yüksek performans %85 ile Rassal Orman modelinde, ölü sınıfına ait en yüksek duyarlılık ise %62 ile Lojistik Regresyon modelinde elde edilmiştir. F1 skoru dikkate alındığında en başarılı modelin Lojistik Regresyon olduğu görülmüştür. Ardından bu modelin test sonuçları dikkate alınarak AYZ tekniklerinden SHAP ile sınıflandırma sonuçlarının nedenleri araştırılmıştır. Elde edilen çıkarımlar, meme

kanseri hastalarının hangi özellikler dolayısıyla riskli sınıfta yer alabileceği hakkında bilgi vermiş ve sonuçların literatürle uyumlu olduğu gözlemlenmiştir.

Bu çalışma, yalnızca doğru tahminler yapabilen bir model geliştirmekle sınırlı kalmayıp, aynı zamanda model kararlarının nedenlerini şeffaf bir biçimde açıklaması ve klinik karar süreçlerine açıklanabilirlik getirmesi açısından literatüre özgün bir katkı sunmaktadır. Literatürdeki birçok çalışmanın görüntü tabanlı uygulamalara odaklandığı dikkate alındığında, bu yaklaşımın sağlık verilerinin farklı boyutlarını değerlendirmede özgün bir katkı sunduğu düşünülmektedir.

5. GELECEK ÇALIŞMALAR

Bu çalışma, meme kanseri sağkalım tahmininde AYZ kullanımının önemini ortaya koysa da, bulguları daha da güçlendirmek ve eksikliklerini gidermek amacıyla çeşitli yönlerde gelecek araştırmalara açıktır. İlk olarak, makine öğrenmesi aşamasında parametre optimizasyonu ile performans artırımı sağlanabilir. Çalışmada uygulanan SHAP yönteminin yanı sıra, diğer AYZ tekniklerinin karşılaştırmalı bir analizi yapılabilir. Bu sayede, farklı açıklanabilirlik yaklaşımlarının uzmanlar tarafından anlaşılabilirlik, güvenilirlik ve kullanılabilirlik açısından nasıl algılandığı değerlendirilebilir. Veri setinin niteliği ve kapsamı üzerine de önemli iyileştirmeler yapılabilir. Mevcut çalışmadaki dengesiz veri yapısının model performansı üzerindeki sınırlayıcı etkisi göz önüne alındığında, daha büyük ve dengeli veri setleriyle çalışılması modelin genelleme yeteneğini artıracaktır. Ayrıca, genetik faktörler, tedavi yöntemleri veya yaşam tarzı faktörlerin de modele dahil edilmesi, tahminlerin klinik geçerliliğine önemli ölçüde katkı sağlayabilir. Bir diğer önemli çalışma alanı ise, bu modellerin gerçek dünyada klinik ortamlara entegrasyonudur. Geliştirilen tahminleme modellerinin ve AYZ yönteminin entegre bir klinik karar destek sistemi haline getirilmesi ve hekimlerle yapılacak kullanılabilirlik çalışmalarından geçirilmesi bu gibi çalışmaların değerini artıracaktır. Son olarak, bu çalışmada sunulan metodolojik yaklaşımın akciğer, prostat veya kolon kanseri gibi diğer kanser türlerine uyarlanarak test edilmesi, yaklaşımın genel geçerliliğini gösterecektir.

6. KAYNAKLAR

1. Kaplan, M., Çakar, F. & Bingöl, H. (2024). Sağlık alanında yapay zekanın kullanımı: Derleme. *Muş Alparslan Üniversitesi Sağlık Bilimleri Dergisi*, 4(3), 75-85.
2. Adadi, A. & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.
3. Terzi, R. (2021). Sağlık sektöründe açıklanabilir yapay zekâ. *Yapay zekâ ve büyük veri çalışmaları, Siber Güvenlik ve Mahremiyet*, 157-175
4. Kalasampath, K., Spoorthi, K.N., Sajeev, S., Kuppa, S.S., Ajay, K. & Maruthamuthu, A. (2025). A literature review on applications of explainable artificial intelligence (XAI). *IEEE Access*, 13, 41111-41140.
5. Yildirim, H. (2024). Hastalıkların teşhis ve tedavisi için tıbbi görüntülemelerden yapay zeka kullanımı ile daha hassas ve daha hızlı sonuç elde etme konusundaki gelişmeler: FDA (Amerikan Gıda ve İlaç Dairesi) tarafından onaylanmış teknolojilerle geliştirilebilecek yeni uygulamalar. *Academic Social Resources Journal*, 6(24), 558-568.
6. Alkhanbouli, R., Abdulla Almadhaani, H.M., Alhosani, F. & Simsekler, M.C.E. (2025). The role of explainable artificial intelligence in disease prediction: A systematic literature review and future research directions. *BMC Medical Informatics and Decision Making*, 25(1), 110.
7. Albahri, A.S., Duhaim, A.M., Fadhel, M.A., Alnoor, A., Bager, N.S., Alzubaidi, L., Albahri, O.S., Alamoodi, A.H., Bai, J., Salhi, A., Santamaria, J., Ouyang, C., Gupta, A., Gu, Y. & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, 96, 156-191.
8. Akin, K.D., Gurkan, C., Budak, A. & Karataş, H. (2022). Açıklanabilir yapay zeka destekli evrimsel sinir ağları kullanılarak maymun çiçeği deri lezyonunun sınıflandırılması. *European Journal of Science and Technology*, 40, 106-110.
9. Nafisah, S.I. & Muhammad, G. (2024). Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence. *Neural Computing and Applications*, 36(1), 111-131.
10. Orman, A., Köse, U. & Yiğit, T. (2021). Brain tumor detection via explainable convolutional neural networks: Açıklanabilir evrimsel sinir ağları ile beyin tümörü tespiti. *El-Cezeri Journal of Science and Engineering*, 8(3), 1323-1337.

11. Viswan, V., Shaffi, N., Mahmud, M., Subramanian, K. & Hajamohideen, F. (2024). Explainable artificial intelligence in Alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1), 1-44.
12. Wani, N.A., Kumar, R. & Bedi, J. (2024). DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. *Computer Methods and Programs in Biomedicine*, 243, 107879.
13. Zhang, Y., Weng, Y. & Lund, J. (2022). Applications of explainable artificial intelligence in diagnosis and surgery. *Diagnostics*, 12(2), 237.
14. Yang, C.C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. *Journal of Healthcare Informatics Research*, 6(2), 228-239.
15. Tjoa, E. & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), 4793-4813.
16. Sun, Q., Akman, A. & Schuller, B.W. (2025). Explainable artificial intelligence for medical applications: A review. *ACM Transactions on Computing for Healthcare*, 6(2), 1-31.
17. van Lent, M., Fisher, W. & Mancuso, M. (2004). An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the 16th Conference on Innovative Applications of Artificial Intelligence (IAAI'04)* (900-907). AAAI Press.
18. Barredo Arrieta, A., Diaz-Rodriguez, N., Del Ser, J., Benetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115.
19. Belle, V. & Papantonis, I. (2021). Principles and practice of explainable machine learning. *Frontiers in Big Data*, 4, 688969.
20. Kalasampath, K., Spoorthi, K.N., Sajeev, S., Kuppa, S.S., Ajay, K. & Maruthamuthu, A. (2025). A literature review on applications of explainable artificial intelligence (XAI). *IEEE Access*, 13, 41111-41140.
21. Rawal, A., McCoy, J., Rawat, D.B., Sadler, B.M. & Amant, R.S. (2022). Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence*, 3(6), 852-866.
22. Nguyen, H.T.T. & Cao, H.Q. (n.d.). Evaluation of explainable artificial intelligence: SHAP, LIME, and CAM. [Manuscript in preparation].
23. Holzinger, A., Saranti, A., Molnar, C., Biecek, P. & Samek, W. (2022). Explainable AI methods-A brief overview. In A. Holzinger et al. (Eds.), *xxAI - Beyond Explainable AI* (13-38). Springer.
24. Salih, A.M., Raisi, Z., Galazzo, I.B., Radeva, P., Petersen, S.E., Lekadir, K. & Menegaz, G. (2025). A perspective on explainable artificial intelligence methods: SHAP and LIME. *Advanced Intelligent Systems*, 7(1), 2400304.
25. Shapley, L. (2020). A value for n-person games. In H. W. Kuhn (Ed.), *Classics in game theory* (69-79). Princeton University Press.
26. Meng, Y., Yang, N., Qian, Z. & Zhang, G. (2021). What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(3), 3.
27. Teng, J. (2019). SEER breast cancer data. *IEEE DataPort*.
28. Potdar, K., Pardawala, T. & Pai, C. (2017). A comparative study of categorical variable encoding techniques for neural network classifiers. *International Journal of Computer Applications*, 175(4), 7-9.
29. Ahsan, M.M., Mahmud, M.A.P., Saha, P.K., Gupta, K.D. & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3), 3.
30. Hua, Y., Stead, T.S., George, A. & Ganti, L. (2025). Clinical risk prediction with logistic regression: Best practices, validation techniques, and applications in medical research. *Academic Medicine & Surgery*, 1(1), 1-13.
31. Patel, H.H. & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, 6(10), 74-78.
32. Masetic, Z. & Subasi, A. (2016). Congestive heart failure detection using random forest classifier. *Computer Methods and Programs in Biomedicine*, 130, 54-64.
33. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937-1967.
34. Wasule, V. & Sonar, P. (2017). Classification of brain MRI using SVM and KNN classifier. In *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)* (218-223). IEEE.

35. Blagus, R. & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, 14(1), 106.
36. Anand, A., Pugalenth, G., Fogel, G.B. & Suganthan, P.N. (2010). An approach for classification of highly imbalanced data using weighting and undersampling. *Amino Acids*, 39(5), 1385-1391.
37. Rakha, E., Toss, M. & Quinn, C. (2022). Specific cell differentiation in breast cancer: A basis for histological classification. *Journal of Clinical Pathology*, 75(2), 76-84.
38. Aizer, A.A., Chen, M.-H., McCarthy, E., Mendu, M.L., Koo, S., Wilhite, T.J., Graham, P.L., Choueiri, T.K., Hoffman, K.E., Martin, N.E., Hu, J.C. & Nguyen, P.L. (2013). Marital status and survival in patients with cancer. *Journal of Clinical Oncology*, 31(31), 3869-3876.
39. Ghavidel, A. & Pazos, P. (2025). Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: A systematic review. *Journal of Cancer Survivorship*, 19(1), 270-294.
40. Moncada-Torres, A., van Maaren, M.C., Hendriks, M.P., Siesling, S. & Geleijnse, G. (2021). Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. *Scientific Reports*, 11(1), 1-13.
41. Park, S.W., Park, Y.L., Lee, E.G., Chae, H., Park, P., Choi, D.W., Choi, Y.H., Hwang, J., Ahn, S., Kim, K., Kim, W.J., Kong, S.-Y., Jung, S.-Y. & Kim, H.J. (2024). Mortality prediction modeling for patients with breast cancer based on explainable machine learning. *Cancers*, 16(22), 3799.

