European Journal of Engineering and Applied Sciences

ISSN: 2651-3412 (Print) & 2667-8454 (Online)
Journal homepage: http://dergipark.gov.tr/EJEAS
Published by Çorlu Faculty of Engineering, Tekirdağ Namık Kemal University
European J. Eng. App. Sci. 9(1), 55-65, 2026

Araştırma Makalesi

Kalp Hastalığı Tahmininde Açıklanabilir Yapay Zekâ: KNIME Tabanlı, Veri Sızıntısı Önlenmiş ve Çapraz Doğrulanmış Karşılaştırmalı Bir Çerçeve

Muhammed Yusuf KAYA[1,*], Bashar ALHAJAHMAD[2]

[1] **Bilgisayar Mühendisliği Anabilim Dalı, Fen Bilimleri Enstitüsü, Siirt Üniversitesi, Siirt, Türkiye, 56000**
[2] **Bilgisayar Mühendisliği, Mühendislik Fakültesi, Siirt Üniversitesi, Siirt, Türkiye, 56000**
[1]myusuf.kaya@siirt.edu.tr, [2]bashar.ahmad@siirt.edu.tr

Geliş: 08.06.2026 *Kabul: 22.07.2026* *DOI: 10.55581/ejeas.1966828*

Öz. Kalp ve damar hastalıkları, dünya genelinde en yüksek ölüm yüküne sahip hastalık grubudur. Bu çalışmada, kalp hastalığı tahmini için KNIME Analytics Platform üzerinde uçtan uca, kodsuz biçimde yapılandırılmış bir Açıklanabilir Yapay Zekâ (XAI) iş akışı geliştirilmiştir. UCI Cleveland Heart Disease veri seti üzerinde Lojistik Regresyon, Random Forest, Gradyan Artırmalı Ağaçlar ve XGBoost modelleri 10 katlı tabakalı çapraz doğrulama ile karşılaştırılmıştır. Önceki çalışmalardan farklı olarak çerçeveye üç metodolojik önlem birlikte yerleştirilmiştir: (i) nominal kategorik değişkenler Tek-Sıcak Kodlama ile dönüştürülmüş, (ii) Min-Max normalizasyonu her katlamada yalnızca eğitim parçası üzerinde öğrenilerek veri sızıntısı önlenmiş, (iii) XGBoost için sistematik hiperparametre değerlendirmesi yürütülmüştür. En yüksek başarı Lojistik Regresyon'dan gelmiştir (Doğruluk = 0,845 ± 0,065; AUC = 0,908). Wilcoxon testi Lojistik Regresyon ile Random Forest ve XGBoost arasındaki farkın anlamlı olmadığını (p = 0,891 ve p = 0,188) göstermiştir. Yorumlanabilirlik iki ayrı pencere üzerinden ele alınmıştır: Lojistik Regresyon katsayıları ve XGBoost üzerinde SHAP analizi. İki yöntemin altı ortak öznitelikte buluşması, öznitelik hiyerarşisinin algoritmik bir artefakt değil veri setindeki gerçek klinik sinyallere ait olduğunu yöntemsel üçleme yoluyla doğrulamıştır.

Anahtar Kelimeler: *Açıklanabilir Yapay Zekâ, Çapraz Doğrulama, KNIME, Lojistik Regresyon, SHAP.*

Explainable Artificial Intelligence for Heart Disease Prediction: A KNIME-Based, Data-Leakage-Free and Cross-Validated Comparative Framework

Abstract. Cardiovascular diseases continue to impose the greatest mortality burden worldwide. This study presents an end-to-end, code-free Explainable Artificial Intelligence (XAI) workflow built on the KNIME Analytics Platform for heart disease prediction. Four classifiers — Logistic Regression, Random Forest, Gradient Boosted Trees, and XGBoost — were compared on the UCI Cleveland Heart Disease dataset using Stratified 10-fold Cross-Validation. Three methodological safeguards were embedded into the pipeline: (i) nominal categorical variables were transformed via One-Hot Encoding, (ii) Min-Max normalization was fitted on the training fold only to prevent data leakage, and (iii) a systematic hyperparameter evaluation was performed for XGBoost. Logistic Regression delivered the highest performance (Accuracy = 0.845 ± 0.065; AUC = 0.908). The Wilcoxon test indicated that the differences between Logistic Regression and Random Forest/XGBoost were not statistically significant (p = 0.891 and p = 0.188). Interpretability was addressed through two parallel windows: Logistic Regression coefficients and SHAP analysis on XGBoost. The convergence of both methods on six common features confirmed via methodological triangulation that the feature hierarchy reflects genuine clinical signals rather than algorithmic artifacts.

Keywords: *Cross-Validation, Explainable Artificial Intelligence, KNIME, Logistic Regression, SHAP.*

∗ Sorumlu yazar
E-mail adres: myusuf.kaya@siirt.edu.tr (M. Y. Kaya)

1. Giriş

Kardiyovasküler hastalıklar, Dünya Sağlık Örgütü'nün son raporlarına göre küresel ölümlerin neredeyse üçte birinden sorumludur ve dünyada en çok can alan hastalık grubu olma özelliğini korumaktadır [1]. Bu hastalıkların önemli bir kısmı erken tanıyla önlenebilir ya da seyri yavaşlatılabilir niteliktedir. Ancak klinik tanı; yaş, kan basıncı, kolesterol, EKG bulguları, miyokart perfüzyon testleri gibi pek çok değişkenin birlikte değerlendirilmesini gerektirir. Bu çok değişkenli yapı, deneyimli hekimler için bile zaman alıcı ve hataya açık bir süreçtir. Makine öğrenmesi tabanlı karar destek sistemlerine duyulan ihtiyaç da tam bu noktada belirginleşmektedir.

Topluluk tabanlı sınıflandırıcılar — Random Forest, Gradyan Artırmalı Ağaçlar, XGBoost — geleneksel istatistiksel yöntemlere göre çoğu zaman daha yüksek doğruluk sunar. Sorun şudur: bu modeller bir tahminin neden üretildiğini hekime anlatamaz. Literatürde sıkça "kara kutu" olarak nitelendirilen bu durum, klinik bağlamda yalnızca bir teknik kusur değil; aynı zamanda etik ve hukuki bir engeldir [2]. Bir hekim, tedaviyi yönlendirecek bir tahminin gerekçesini görmek ister. Hasta, kendisi hakkında verilen kararın temelini bilmeye hakkıdır.

Bu ihtiyaç, Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence, XAI) alanının hızla gelişmesini tetiklemiştir. Lundberg ve Lee [3] tarafından önerilen SHAP yöntemi, kuramsal temellerini oyun teorisinin Shapley değerlerinden alır ve her bir özniteliğin tahmine olan katkısını matematiksel olarak tutarlı biçimde paylaştırır. SHAP'in en güçlü tarafı, hem modelin genel davranışını betimleyen global açıklamalar hem de tek bir hasta için yapılan tahminin nedenlerini ortaya koyan yerel açıklamalar üretebilmesidir [4].

Bu çalışma, kalp hastalığı tahmini için KNIME Analytics Platform üzerinde kodsuz olarak kurulmuş, uçtan uca bir XAI çerçevesi sunmaktadır. Çerçevenin üç ayırt edici yönü vardır. İlk olarak, nominal kategorik değişkenler (göğüs ağrısı tipi, talasemi durumu, ST eğimi, dinlenme EKG sonucu ve renklendirilen damar sayısı) Tek-Sıcak Kodlama yöntemiyle dönüştürülmüştür; böylece tamsayı temsilinin yarattığı sahte sıralama (false ordinality) sorunu ortadan kaldırılmıştır. İkincisi, Min-Max normalizasyonu yalnızca eğitim kıvrımında öğrenilerek doğrulama kıvrımına uygulanmış, model değerlendirme sürecine veri sızıntısı karışması engellenmiştir. Üçüncüsü — bu çalışmanın belki de en ayırt edici yönü — model seçimi 10 katlı tabakalı çapraz doğrulama ile yapılmış; tek seferlik hold-out değerlendirmesinin getirdiği rastlantısal kayma riski ortadan kaldırılmıştır.

Bu üç önlemin ortak çalıştırıldığı bir çerçevede, klinik verilerde sıkça karşılaşılan küçük örneklem ve dengesizliğe yakın sınıf yapısının modeller üzerindeki etkisi daha sağlam biçimde gözlemlenebilmektedir. Nitekim sonuçlar, beklenen biçimde XGBoost değil — basit bir doğrusal model olan Lojistik Regresyon'un başa geçtiğini göstermiştir. Bu bulgu rastlantısal değildir; Tek-Sıcak Kodlama'nın doğrusal modellerin ifade gücünü kategorik değişken yapısı üzerinden nasıl beslediğini bizzat ortaya koymaktadır.

1.1. Araştırmanın Özgün Katkıları

Bu çalışma, kalp hastalığı tahmini ve açıklanabilir yapay zekâ alanına aşağıdaki özgün katkıları sunmaktadır:

- Kalp hastalığı tahmini için KNIME Analytics Platform üzerinde tamamen görsel iş akışına dayalı, kodsuz ve yeniden üretilebilir bir Açıklanabilir Yapay Zekâ (XAI) çerçevesi geliştirilmiştir.
- Veri ön işleme sürecinde nominal kategorik değişkenler Tek-Sıcak Kodlama yöntemiyle dönüştürülmüş ve Min-Max normalizasyonu yalnızca eğitim katlamalarında öğrenilerek veri sızıntısı sistematik biçimde önlenmiştir.
- Lojistik Regresyon, Random Forest, Gradyan Artırmalı Ağaçlar ve XGBoost modelleri 10 katlı tabakalı çapraz doğrulama altında karşılaştırılmış ve model performansları istatistiksel anlamlılık testleri ile değerlendirilmiştir.
- Model yorumlanabilirliği, hem Lojistik Regresyon katsayıları hem de XGBoost üzerinde gerçekleştirilen SHAP analizleri kullanılarak iki bağımsız açıklama mekanizması üzerinden incelenmiştir.
- Farklı model aileleri ve farklı açıklama yöntemleri arasında elde edilen ortak öznitelik sıralamaları kullanılarak klinik biyobelirteçlerin güvenilirliği yöntemsel üçleme yaklaşımıyla doğrulanmıştır.
- Çalışma, performans, yorumlanabilirlik ve metodolojik şeffaflığı tek bir KNIME iş akışı içerisinde bir araya getirerek klinik karar destek sistemleri için uygulanabilir ve denetlenebilir bir çözüm sunmaktadır.

Çalışmanın geri kalanı şöyle düzenlenmiştir: İkinci bölümde, kalp hastalığı tahminine yönelik güncel literatür incelenmektedir. Üçüncü bölümde, kullanılan veri seti, ön işleme adımları, çapraz doğrulama mimarisi, algoritmalar ve SHAP analizi ayrıntılarıyla tanıtılmaktadır. Dördüncü bölüm, performans sonuçlarını ve SHAP bulgularını sunmakta; beşinci bölüm bunları klinik ve metodolojik bir bakışla tartışmaktadır. Altıncı bölümde ise temel çıkarımlar ile gelecek araştırma yönleri özetlenmektedir.

2. İlgili Çalışmalar

Son beş yılda kalp hastalığı tahminine yönelik makine öğrenmesi çalışmaları ciddi biçimde artmış; özellikle topluluk modelleri ile XAI yöntemlerinin birlikte kullanıldığı yaklaşımlar literatürün önemli bir parçasına dönüşmüştür.

Mushtaq ve diğerleri [5], XGBoost ile SHAP'i bir araya getirerek nedensellik vurgusu da içeren bir tahmin çerçevesi önermiştir. Çalışmada PC algoritmasına dayalı nedensellik analizinin SHAP açıklamalarıyla birlikte kullanılması, modelin klinik bağlamda yorumlanabilirliğini artıran somut bir katkı olarak öne çıkmaktadır.

Almutairi ve Dardouri [6], XGBoost, Random Forest, Destek Vektör Makineleri, Lojistik Regresyon ve Derin Sinir Ağlarını karşılaştırmalı olarak ele almış; XGBoost ile SVM'i birleştiren hibrit bir model üzerinden %89,3 doğruluk ve 0,905 F1 skoru bildirmiştir. Yazarların SHAP analizleri yaş, kolesterol, dinlenme kan basıncı ve göğüs ağrısı tipini öne çıkan dört prediktör olarak işaret etmektedir.

Rezk ve diğerleri [7], SHAP ve LIME ile desteklenmiş bir oylama topluluk modeli geliştirmiş; tekil algoritmalar arasında XGBoost'u %89 doğrulukla öne çıkarmıştır. Çalışmanın belki de en önemli mesajı, XAI yöntemlerinin klasik performans metriklerinden bağımsız olarak modele duyulan klinik güveni doğrudan artırdığıdır.

Bu çalışmaların ortak çıkarımı oldukça nettir: tahmin gücü ile açıklanabilirlik artık birbirinin tamamlayıcısıdır. Buna karşın, mevcut literatürün büyük bölümü Python tabanlı kütüphaneler (scikit-learn, shap, lime) üzerine kuruludur. Görsel iş akışı tabanlı, kod yazmadan yürütülen ve metodolojik bütünlüğü tek bir çerçevede sağlayan XAI uygulamaları henüz yaygın değildir. Üstelik bu yayınların önemli bir kısmında veri sızıntısı, hiperparametre arama protokolü ve kategorik değişken temsili gibi süreçler arka planda kalmakta; bazen hiç tartışılmamaktadır.

Çalışmanın literatürdeki konumunu daha net görünür kılmak için Tablo 1'de Cleveland Heart Disease veri seti üzerinde gerçekleştirilen güncel çalışmalar ile bu çalışma yan yana sunulmuştur. Karşılaştırma; en iyi performans elde edilen model, raporlanan doğruluk, kullanılan XAI yöntemi, çapraz doğrulama uygulanıp uygulanmadığı, veri sızıntısının açıkça önlenip önlenmediği ve istatistiksel anlamlılık testinin raporlanıp raporlanmadığı kriterleri üzerinden yapılmıştır.

Tablo 1 Cleveland Heart Disease Veri Seti Üzerindeki Güncel Çalışmaların Metodolojik Karşılaştırması

Çalışma	En İyi Model	Doğruluk	XAI Yöntemi	Çapraz Doğrulama	Veri Sızıntısı Önlemi	İstatistiksel Test
Mushtaq vd. [5]	XGBoost	≈ 0,89	SHAP + PC nedensellik	Belirtilmemiş	Belirtilmemiş	Yok
Almutairi & Dardouri [6]	XGB–SVM hibrit	0,893	SHAP	Belirtilmemiş	Belirtilmemiş	Yok
Rezk vd. [7]	Oylama topluluğu	0,890	SHAP + LIME	Belirtilmemiş	Belirtilmemiş	Yok
Bu çalışma	**Lojistik Regresyon (One-Hot)**	**0,845 ± 0,065**	**SHAP + LR β üçlemesi**	**10 katlı tabakalı**	**Açıkça (fold-içi Normalizer)**	**Wilcoxon + Bonferroni**

Tablo 1 incelendiğinde, önerilen yaklaşımın literatürde raporlanan çalışmalarla karşılaştırılabilir performans düzeyine sahip olduğu görülmektedir. Bununla birlikte, mevcut çalışmayı benzer araştırmalardan ayıran temel unsur yalnızca performans sonuçları değildir. Çalışmada veri sızıntısını önleyen katlama-içi normalizasyon stratejisi uygulanmış, modeller 10 katlı tabakalı çapraz doğrulama ile değerlendirilmiş ve elde edilen sonuçlar istatistiksel anlamlılık testleri ile desteklenmiştir. Ayrıca model yorumlanabilirliği, hem Lojistik Regresyon katsayıları hem de SHAP analizi kullanılarak iki bağımsız açıklama mekanizması üzerinden incelenmiştir. Bu yönleriyle çalışma, metodolojik şeffaflığı, yeniden üretilebilirliği ve açıklanabilirliği aynı çerçevede bir araya getiren bütüncül bir yaklaşım sunmaktadır.

3. Materyal ve Yöntem

3.1. Veri Seti

Çalışmada, UCI Machine Learning Repository tarafından kamuya açık olarak yayımlanan Cleveland Heart Disease veri setinin Kaggle üzerinde yeniden işlenmiş sürümü kullanılmıştır [8]. Veri seti 297 hasta kaydı içerir; her hasta 13 klinik öznitelik ve bunlara karşılık gelen ikili hedef değişken (condition: 0 = sağlıklı, 1 = hasta) ile temsil edilmektedir. Sınıf dağılımı görece dengelidir: 160 sağlıklı, 137 hasta (≈ %53,9 / %46,1). Eksik değer bulunmamaktadır. Veri seti CC BY 4.0 lisansı altında dağıtılmaktadır. Özniteliklerin tanımları Tablo 2'de verilmiştir.

3.2. Veri Ön İşleme

Veri ön işleme aşaması, önceki çalışmalarda sıklıkla gözden kaçan iki kusurun ortadan kaldırılmasını esas almıştır: kategorik değişkenlerin yanlış temsili ve değerlendirme sürecine veri sızıntısı karışması.

Tablo 2 Cleveland Heart Disease Veri Setindeki Özniteliklerin Tanımı

Öznitelik	Tanım
age	Yaş (yıl)
sex	Cinsiyet (1 = erkek, 0 = kadın)
cp	Göğüs ağrısı tipi (0–3)
trestbps	Dinlenme kan basıncı (mmHg)
chol	Serum kolesterol (mg/dL)
fbs	Açlık kan şekeri > 120 mg/dL (1/0)
restecg	Dinlenme EKG sonucu (0–2)
thalach	Maksimum kalp atış hızı
exang	Egzersiz kaynaklı anjina (1/0)
oldpeak	Egzersiz kaynaklı ST depresyonu
slope	Pik egzersiz ST segment eğimi (0–2)
ca	Floroskopide renklendirilen ana damar sayısı (0–3)
thal	Talasemi durumu (0–3)
condition	Hedef değişken (0 = sağlıklı, 1 = hasta)

3.2.1. Tek-Sıcak Kodlama

Cleveland veri setindeki cp, restecg, slope, ca ve thal değişkenleri özünde nominaldir; ancak veri setinde tamsayılarla temsil edilirler. Örneğin göğüs ağrısı tipi (cp) için 0, 1, 2 ve 3 sırasıyla tipik anjina, atipik anjina, anjinal olmayan ağrı ve asimptomatik durumları kodlar. Buradaki sayılar arasında matematiksel bir büyüklük ilişkisi yoktur — ancak ağaç tabanlı algoritmalar bu sayıları sıralı (ordinal) yorumlama eğilimi gösterir. Bu, literatürde sahte sıralama (false ordinality) olarak bilinen kavramsal hatayı doğurur.

Sorunu çözmek için bu beş değişken KNIME'ın One to Many düğümüyle Tek-Sıcak Kodlama'ya tabi tutulmuştur. Her değişkenin her olası değeri ayrı bir ikili sütun (0/1) olarak

yeniden tanımlanmış; orijinal kolonlar Column Filter düğümüyle veri setinden çıkarılmıştır. İşlem sonunda öznitelik uzayı 13'ten 25'e genişlemiştir. Bu genişleme yalnızca teknik bir biçim değişikliği değildir; doğrusal modellere ifade gücü kazandıran, ağaç tabanlı modellere ise gereksiz boyut yükü getirebilen iki yönlü bir tasarım kararıdır.

3.2.2. Veri Sızıntısının Önlenmesi

Sayısal özelliklerin (age, trestbps, chol, thalach, oldpeak) ölçek farklarını gidermek için Min-Max normalizasyonu kullanılmıştır. Burada metodolojik açıdan kritik bir nüans bulunmaktadır: Normalizasyon parametreleri (minimum ve maksimum değerler) doğrudan tüm veri üzerinde değil, her çapraz doğrulama kıvrımında yalnızca eğitim parçası üzerinde öğrenilmiştir. Öğrenilen bu parametreler ardından KNIME'ın Normalizer (Apply) düğümü aracılığıyla doğrulama parçasına aktarılmıştır. Böylece doğrulama setinin istatistiksel bilgileri model eğitimine herhangi bir biçimde sızmamış, performans tahminleri yapay bir şişme barındırmamıştır.

İkili (binary) ve Tek-Sıcak Kodlanmış değişkenler zaten [0, 1] aralığında olduklarından normalizasyon kapsamı dışında tutulmuştur.

3.3. 10 Katlı Tabakalı Çapraz Doğrulama

Model değerlendirmesi, KNIME'ın X-Partitioner ve X-Aggregator düğüm çiftiyle 10 katlı tabakalı (Stratified 10-fold) çapraz doğrulama mimarisi üzerinden gerçekleştirilmiştir. Her algoritma için ayrı bir X-Partitioner + X-Aggregator zinciri kurulmuştur; bu sayede her model birbirinden bağımsız 10 eğitim-doğrulama döngüsü içinde değerlendirilmiştir. Tabakalı bölümleme, her kıvrımda sınıf oranının orijinal veri setiyle aynı tutulmasını sağlamıştır.

Çapraz doğrulama tercihi sıradan bir metodolojik seçim değildir. Veri setinin görece küçük olması (n = 297), tek seferlik bir hold-out bölümlemesinin sonuçları rastlantısal bir veri bölünmesine karşı hassas hâle getirebilirdi. Çapraz doğrulama, performans tahminlerinin varyansını düşürerek model karşılaştırmasını daha sağlam bir zemine oturtmaktadır. Tüm rastgele süreçlerde 12345 tohum değeri kullanılarak sonuçların yeniden üretilebilirliği güvence altına alınmıştır.

3.3.1. Aşırı Öğrenmeye Karşı Alınan Önlemler

Veri sızıntısına gösterilen özenle paralel biçimde, aşırı öğrenme (overfitting) riski de çerçevenin birden fazla katmanında ele alınmıştır. On katlı tabakalı çapraz doğrulamanın kendisi, tek bir eğitim-test bölünmesine kıyasla performans tahminlerinin varyansını azaltarak modelin belirli bir veri bölünmesine aşırı uyum sağlamasını sınırlayan birincil önlemdir. Model düzeyinde ise Random Forest'ın torbalama (bagging) mekanizması ile XGBoost'un L1/L2 düzenlileştirmesi, ağaç tabanlı modellerin aşırı öğrenmeye karşı yerleşik savunma katmanlarını oluşturmaktadır; XGBoost için ayrıca sınırlı bir hiperparametre taraması (Bölüm 3.4) ile ağaç derinliği ve boosting turu sayısı kontrol altında tutulmuştur.

Bu önlemlerin sınırları da açıkça belirtilmelidir. Veri setinin küçüklüğü (n = 297) ve tek bir kaynaktan (Cleveland) gelmesi, çapraz doğrulama ile elde edilen düşük varyansın dış popülasyonlara genellenebilirliği garanti etmediği anlamına gelmektedir; iç doğrulamada gözlemlenen kararlılık, farklı hasta gruplarında aynı ölçüde korunmayabilir. Ayrıca hiperparametre taraması yalnızca beş yapılandırmayla sınırlı tutulmuş olup, iç içe geçmiş çapraz doğrulama (nested cross-validation) kullanılmamıştır; bu nedenle taranan az sayıdaki yapılandırma arasından seçim yapılmış olması, aşırı öğrenme riskini tamamen ortadan kaldırmaktan çok azaltmaktadır. Sonuç olarak, uygulanan önlemler aşırı öğrenmeyi pratik ölçekte kontrol altına almakla birlikte, mutlak bir güvence olarak yorumlanmamalıdır.

3.4. Sınıflandırma Algoritmaları ve Hiperparametre Değerlendirmesi

Karşılaştırmaya, klinik tahmin literatüründe en sık başvurulan dört algoritma dahil edilmiştir. Lojistik Regresyon, olasılık tabanlı doğrusal bir sınıflandırıcı olarak yorumlanabilirlik açısından klasik referans modelidir. Random Forest, birden çok karar ağacının topluluk halinde çalıştığı ve aşırı öğrenmeyi sınırlayan bir torbalama (bagging) yöntemidir [9]; bu çalışmada 200 ağaç ve Gini bölünme kriteriyle yapılandırılmıştır. Gradyan Artırmalı Ağaçlar (GBT), ağaçların hatalar üzerinde sıralı olarak optimize edildiği gradyan artırma yaklaşımıdır; 100 ağaç, 0,1 öğrenme oranı ve 5 maksimum derinlikle çalıştırılmıştır. XGBoost ise gradyan artırma çerçevesine L1/L2 düzenlileştirme, sütun örnekleme ve verimli paralel hesaplama ekleyen, küçük örneklem rejimlerinde özellikle başarılı olduğu bilinen gelişmiş bir varyanttır [10].

XGBoost için hiperparametre seçiminin keyfi yapılmaması adına sistematik bir değerlendirme yürütülmüştür. XGBoost'un performansını en güçlü biçimde etkilediği literatürde ortaklaşa kabul gören üç parametre — boosting rounds, learning rate (eta) ve max depth — üzerinde beş farklı yapılandırma test edilmiştir. Bu yapılandırmaların hold-out testindeki sonuçları Tablo 3'te sunulmuştur.

Tablo 3 XGBoost Hiperparametre Seçimi İçin Kullanılan Hold-out Değerlendirme Sonuçları

Konfig.	Boosting Rounds	Eta	Max Depth	Doğruluk	Kappa	F1 (Hasta)
C1	100	0,10	3	0,767	0,532	0,753
C2 *	**100**	**0,10**	**5**	**0,811**	**0,621**	**0,800**
C3	200	0,05	5	0,800	0,598	0,786
C4	150	0,10	4	0,811	0,621	0,800
C5	200	0,05	3	0,778	0,554	0,762

** En iyi konfigürasyon. C2 ve C4 aynı performansa ulaşmış; Occam Usturası gereği daha az iterasyona ihtiyaç duyan C2 (100 / 0,10 / 5) tercih edilmiştir*

Hiperparametre değerlendirme süreci yalnızca uygun XGBoost yapılandırmasını belirlemek amacıyla gerçekleştirilmiştir. Hesaplama maliyetini ve deneysel karmaşıklığı sınırlamak amacıyla bu aşamada ayrı bir hold-out

veri bölmesi kullanılmıştır. Hiperparametre seçimi tamamlandıktan sonra belirlenen en iyi yapılandırma, diğer modellerle birlikte 10 katlı tabakalı çapraz doğrulama altında yeniden değerlendirilmiş ve nihai performans karşılaştırmaları bu sonuçlar üzerinden gerçekleştirilmiştir. Burada metodolojik açıdan vurgulanması gereken nokta şudur: hiperparametre optimizasyonu, nihai performans karşılaştırmasında kullanılan çapraz doğrulama kıvrımlarından tamamen bağımsız, ayrı bir veri bölmesi üzerinde gerçekleştirilmiştir. Dolayısıyla Tablo 4'te raporlanan çapraz doğrulama sonuçları, hiperparametre seçim sürecinde kullanılan veriden herhangi bir biçimde etkilenmemiş; hiperparametre optimizasyonunun çapraz doğrulamanın tarafsızlığına zarar vermesi önlenmiştir.

C2 ve C4 yapılandırmaları aynı doğruluk ve Kappa değerlerini üretmiştir; bu da modelin veri seti üzerindeki performans tavanına ulaştığına işaret etmektedir. Daha az boosting turu ile aynı sonucu veren C2 (100 / 0,10 / 5) yapılandırması bundan sonraki tüm analizlerde XGBoost konfigürasyonu olarak benimsenmiştir.

Tüm predictor düğümlerinde, SHAP analizinin gerektirdiği sınıf olasılıklarını üretmek üzere "Append individual class probabilities" seçeneği etkinleştirilmiştir.

3.5. Yorumlanabilirlik Yaklaşımı: İki Bağımsız Pencere

Bu çalışma yorumlanabilirliği iki ayrı pencere üzerinden ele almıştır. Birinci pencere, Lojistik Regresyon modelinin doğal yorumlanabilirliğidir. Lojistik Regresyon, "tasarımı gereği yorumlanabilir" (interpretable by design) bir doğrusal modeldir; her özniteliğe atadığı katsayı (β), o öznitelikin tahmin üzerindeki yönünü ve göreli ağırlığını doğrudan okumaya izin verir. Bu nedenle Lojistik Regresyon için ayrı bir post-hoc XAI yöntemine ihtiyaç bulunmamaktadır.

İkinci pencere ise XGBoost gibi doğrusal olmayan ve öznitelikler arası etkileşimleri öğrenebilen kara kutu modeller için gereken post-hoc yorumlama katmanıdır. Bu çalışmada, en güçlü doğrusal olmayan model olan XGBoost (C2) için SHAP yöntemi uygulanmıştır. Lojistik Regresyon katsayı analizinin yanına XGBoost SHAP analizinin yerleştirilmesi rastlantısal bir tercih değildir: iki yapısal olarak farklı modelin, iki bağımsız yorumlama mekanizması üzerinden aynı klinik prediktörlere işaret edip etmediğini sınamak, açıklamaların algoritmik bir artefakt değil; veri setindeki gerçek klinik sinyaller olduğunu doğrulamanın etkili bir yoludur. Bu yaklaşım literatürde yöntemsel üçleme (methodological triangulation) olarak adlandırılır.

SHAP yöntemi, oyun teorisinden uyarlanan Shapley değerlerini kullanarak her bir tahmini özniteliklerin katkılarının doğrusal toplamı biçiminde ayrıştırır [3]. KNIME'ın Machine Learning Interpretability eklentisindeki SHAP Loop Start ve SHAP Loop End düğümleri bu hesaplama için kullanılmıştır.

Burada metodolojik bir açıklama gerekmektedir. Model seçimi 10 katlı çapraz doğrulamayla yapılmış olmakla birlikte, SHAP analizi yorumlanabilir tek bir model nesnesi gerektirdiği için ayrı bir hold-out doğrulama (%80 eğitim, %20 test) üzerinde XGBoost (C2) yeniden eğitilmiştir. Bu yaklaşım, model karşılaştırması ile yorumlanabilirlik analizinin birbirinden bağımsız iki adım olarak yürütülmesini ve SHAP açıklamalarının belirli bir model örneği üzerinden net biçimde okunmasını sağlamıştır.

SHAP hesaplaması iki giriş tablosu gerektirir: açıklanmak istenen örnekleri içeren ROI (Rows of Interest) tablosu ve referans dağılımı temsil eden sampling tablosu. Bu çalışmada test setinden tabakalı rastgele seçimle alınan 30 hasta ROI olarak; eğitim setinden tabakalı seçimle alınan 100 hasta sampling tablosu olarak kullanılmıştır.

3.6. Değerlendirme Metrikleri

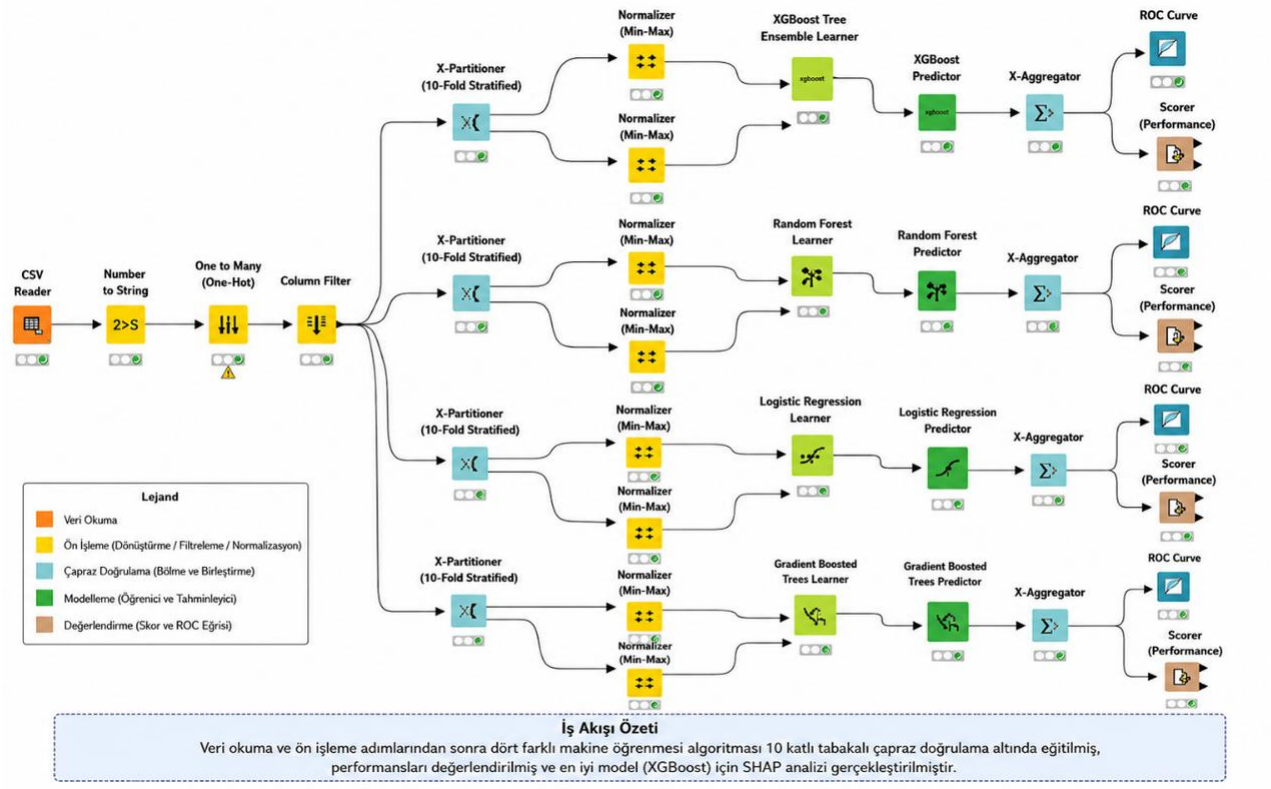
Modellerin performansı; Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall), F1 Skoru, Cohen'in Kappa katsayısı ve ROC-AUC metrikleri ile değerlendirilmiştir. Klinik tahminde hasta sınıfının atlanmaması — yani yanlış negatiflerin düşük tutulması — yaşamsal önemde olduğu için, Duyarlılık ve AUC metriklerinin yorumlanmasına görece daha fazla ağırlık verilmiştir. Cohen'in Kappa'sı, rastgele şans etkisinden arındırılmış uyum ölçüsü olarak özellikle sınıflandırma görevlerinde dengesizliğe karşı daha sağlam bir bakış sunmaktadır.

Modeller arasındaki performans farklarının istatistiksel olarak değerlendirilmesi amacıyla Wilcoxon işaretli sıra testi kullanılmıştır. Bu test, çapraz doğrulama katlamalarından elde edilen eşleştirilmiş performans ölçümlerinin normal dağılım varsayımını gerektirmemesi nedeniyle tercih edilmiştir. Çoklu karşılaştırmalardan kaynaklanabilecek Tip-I hata olasılığını kontrol altına almak amacıyla p-değerlerine Bonferroni düzeltmesi uygulanmıştır.

3.7. KNIME İş Akışı Mimarisi

Yukarıda anlatılan tüm adımlar, KNIME Analytics Platform 5.x sürümü üzerinde tek bir görsel iş akışında bir araya getirilmiştir [11]. Veri okuma ve ön işleme zinciri tüm modeller için ortak çalışırken, dört algoritma her biri kendi X-Partitioner + Normalizer/Apply + Learner/Predictor + X-Aggregator + Scorer + ROC Curve dalında paralel olarak yürütülmüştür. İş akışının genel görünümü Şekil 1'de sunulmuştur.

Şekil 1'de gösterilen iş akışı dört temel aşamadan oluşmaktadır: (i) veri okuma ve ön işleme, (ii) Tek-Sıcak Kodlama ve normalizasyon işlemleri, (iii) 10 katlı tabakalı çapraz doğrulama altında model eğitimi ve performans değerlendirmesi, (iv) en başarılı model üzerinde gerçekleştirilen SHAP tabanlı yorumlanabilirlik analizi. Böylece veri hazırlama, modelleme, değerlendirme ve açıklanabilirlik süreçleri tek bir KNIME iş akışı içerisinde bütünleşik olarak yürütülmüştür. Çalışmanın şeffaflığını ve bağımsız araştırmacılar tarafından yeniden üretilebilirliğini (reproducibility) sağlamak amacıyla, Şekil 1'de sunulan uçtan uca KNIME iş akışı genel erişime açılmıştır (https://hub.knime.com/s/akbnF32ca-6Swb_s).



Şekil 1. *10 Katlı Çapraz Doğrulama Tabanlı KNIME XAI İş Akışı*

4. Bulgular

4.1. Model Performans Karşılaştırması

Dört modelin 10 katlı çapraz doğrulama sonuçları Tablo 4'te sunulmuştur. Sıralamada sürpriz bir tablo ortaya çıkmıştır: en yüksek başarıyı, çalışmaya klasik referans modeli olarak dahil edilen Lojistik Regresyon sergilemiştir.

Lojistik Regresyon, Doğruluk = 0,845, Kappa = 0,687 ve AUC = 0,908 ile bütün metriklerde liderdir. Random Forest ikinci sırada (0,811 / 0,620 / 0,890) yer alırken, XGBoost üçüncü (0,805 / 0,607 / 0,878), Gradyan Artırmalı Ağaçlar ise sonuncu (0,778 / 0,551 / 0,856) olmuştur. Burada altı çizilmesi gereken nokta şudur: doğrusal bir modelin, kendisinden çok daha karmaşık üç ağaç topluluğunu hem doğrulukta hem AUC'de geçmesi tipik bir sonuç değildir. Bu bulgu, Tek-Sıcak Kodlama'nın doğrusal modeller üzerindeki etkisini somut biçimde göstermektedir.

4.2. Modeller Arası Farkların İstatistiksel Anlamlılığı

Tablo 4'te sunulan performans farklarının rastgele dalgalanmaların ötesinde gerçek bir model üstünlüğünü yansıtıp yansıtmadığını sınamak için çapraz doğrulama katlamalarından elde edilen kıvrım-bazlı doğruluk değerleri üzerinde Wilcoxon işaretli sıra testi uygulanmıştır. Karşılaştırmalar Lojistik Regresyon'u referans alarak diğer üç model ile ikili biçimde yürütülmüştür. Çoklu karşılaştırma yanlılığını gidermek üzere p-değerleri Bonferroni düzeltmesine tabi tutulmuştur (k = 3). Sonuçlar Tablo 5'te sunulmuştur.

Tablo 4 Modellerin 10 Katlı Çapraz Doğrulama Performansları (Ortalama ± Standart Sapma)

Sıra	Model	Doğruluk ± SD	Kesinlik	Duyarlılık	F1 (Hasta)	Kappa	AUC
1	**Lojistik Regresyon**	**0,845 ± 0,065**	**0,887**	**0,796**	**0,826**	**0,687**	**0,908**
2	Random Forest	0,811 ± 0,074	0,831	0,788	0,794	0,620	0,890
3	XGBoost (C2)	0,805 ± 0,063	0,819	0,775	0,788	0,607	0,878
4	Gradyan Artırmalı Ağaçlar	0,778 ± 0,059	0,819	0,730	0,752	0,551	0,856

Not: Doğruluk değerleri 10 kıvrımdaki ortalama ± örneklem standart sapması (ddof=1) olarak verilmiştir. Diğer metrikler (Kesinlik, Duyarlılık, F1, Kappa, AUC) tüm kıvrımlardaki tahminlerin birleştirilmesiyle hesaplanmış toplam değerlerdir.

Tablo 5 Lojistik Regresyon ile Diğer Modellerin Karşılaştırılması (Wilcoxon Signed-Rank Test, k=3)

Karşılaştırma	Doğruluk (LR-X)	Δ	W	p	p (Bonferroni)	Sonuç (α = 0,05)
LR vs Random Forest	0,845 — 0,811	+0,034	10,0	0,297	0,891	Anlamsız
LR vs XGBoost (C2)	0,845 — 0,805	+0,041	2,0	0,063	0,188	Sınırda
LR vs Gradyan Artırmalı Ağaçlar	**0,845 — 0,778**	**+0,067**	**1,0**	**0,016**	**0,047**	**Anlamlı ***

** Bonferroni düzeltmesi sonrası p < 0,05.*

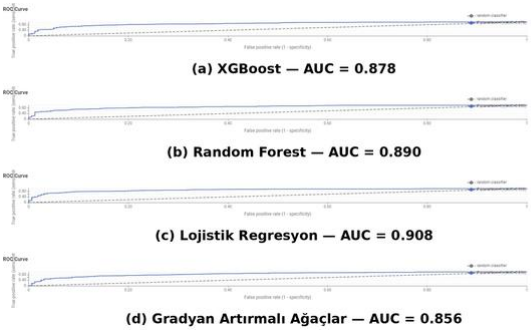
Sonuçların doğrudan bir okuması, Lojistik Regresyon'un mutlak doğruluk değerleri açısından üst sırada olmakla birlikte; Random Forest ve XGBoost ile arasındaki farkın Bonferroni düzeltilmiş p-değerleri açısından istatistiksel olarak anlamlı olmadığını göstermektedir (sırasıyla p = 0,891 ve p = 0,188). Yalnızca Gradyan Artırmalı Ağaçlar karşısında elde edilen üstünlük istatistiksel olarak anlamlı bulunmuştur (p = 0,047). Bu tablo, çalışmanın çekirdek mesajı açısından önemli bir nüansa işaret etmektedir.

Sonuç salt bir sıralama olarak okunduğunda "Lojistik Regresyon birinci, XGBoost üçüncü" gibi bir hiyerarşik yorum üretilebilir. Ancak istatistiksel test sonuçları aslında farklı bir öyküye işaret etmektedir: Lojistik Regresyon, Random Forest ve XGBoost — üç model — küçük örneklemli Cleveland veri setinde birbiriyle karşılaştırılabilir performans sergilemektedir. Çalışmanın kalbinde yer alan iddia da zaten "en iyi tek modeli bulma" değil; "doğru veri temsiliyle, titiz değerlendirmeyle ve sızdırmasız ön işlemeyle basit bir doğrusal modelin karmaşık ağaç toplulukları ile yarışabildiğini gösterme"dir.

Bu bulgu, klinik karar destek bağlamında pratik bir çıkarımı da beraberinde getirmektedir. Üç model performans açısından eşdeğer kabul edilebiliyorsa, model seçiminde performansın yanı sıra yorumlanabilirlik, hesaplama maliyeti ve dağıtım kolaylığı gibi ölçütler ön plana çıkmaktadır. Lojistik Regresyon, bu üç ölçütte de açık bir avantaja sahiptir.

4.3. ROC Eğrileri

Modellerin ROC eğrileri Şekil 2'de toplu olarak sunulmuştur. Sıralama Tablo 4'teki performans dizilimini görselle desteklemektedir: Lojistik Regresyon eğrisi en yüksek alana ulaşmakta, onu sırasıyla Random Forest, XGBoost ve Gradyan Artırmalı Ağaçlar izlemektedir.

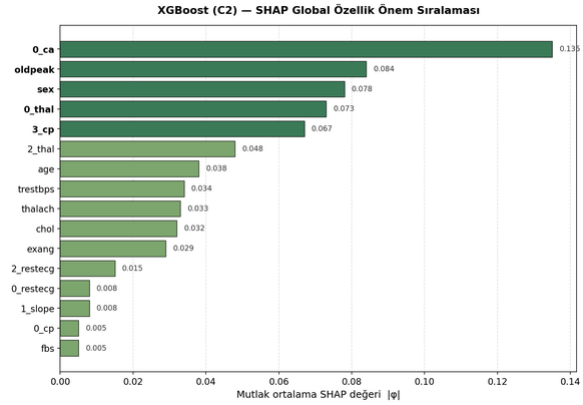


Şekil 2. *Dört Modelin 10 Katlı Çapraz Doğrulama ROC Eğrileri*

4.4. SHAP ile Global Özellik Önem Analizi

XGBoost (C2) üzerinde gerçekleştirilen SHAP analizi sonucunda her özniteliğin tahminler üzerindeki ortalama mutlak SHAP katkısı hesaplanmış ve büyükten küçüğe sıralanmıştır. Sonuçlar Şekil 3'te görülebilir.

Şekil 3'e göre modelin tahminlerinde en belirleyici öznitelik 0_ca (mutlak SHAP ortalaması ≈ 0,135) olup bunu sırasıyla oldpeak (≈ 0,084), sex (≈ 0,078), 0_thal (≈ 0,073), 3_cp (≈ 0,067) ve 2_thal (≈ 0,048) izlemektedir. Daha alt sıralarda age, trestbps, thalach, chol ve exang gibi öznitelikler yer almaktadır. En az katkı sağlayan değişken fbs olmuştur (≈ 0,005).



Şekil 3. *XGBoost (C2) Modelinin SHAP Tabanlı Global Özellik Önem Grafiği*

Tek-Sıcak Kodlama'nın burada sağladığı görünmez bir kazanım vardır: özelliklerin önem dereceleri artık sınıf düzeyinde ayrışmaktadır. Önceki çalışmalarda ca tek bir kolon olarak değerlendirilirken, bu çalışmada modelin asıl olarak 0_ca sınıfına — yani "hiçbir koroner damarda tıkanma bulunmaması" durumuna — yüksek önem atfettiği görülmektedir. Talasemi değişkeninin alt sınıflarında 0_thal (normal) ve 2_thal (reversible defekt) öne çıkmakta; göğüs ağrısı tipi içinde ise asimptomatik durumu temsil eden 3_cp en etkili sınıf olarak ortaya çıkmaktadır.

Bu sıralamanın klinik kardiyoloji literatürüyle örtüştüğünü söylemek mümkündür. Floroskopide gözlemlenen tıkalı damar sayısı, koroner arter hastalığının en doğrudan anatomik göstergesidir. Talasemi (thal) parametresi, talyum sintigrafisi üzerinden miyokart perfüzyonunu değerlendiren testi temsil eder; iskemi varlığını ortaya koyan başlıca araçlardan biridir. Egzersizle indüklenen ST depresyonu (oldpeak) ise miyokart iskemisinin elektrokardiyografik klasik bulgusudur. Modelin bu üç klinik biyobelirteci en güçlü prediktör olarak seçmesi, yalnızca istatistiksel bir korelasyon öğrenmediğini; klinik mantığa uygun bir karar yapısı geliştirdiğini destekler niteliktedir.

Diğer yandan halk arasında kalp hastalığıyla en güçlü biçimde ilişkilendirilen iki değişken — kolesterol (chol) ve açlık kan şekeri (fbs) — SHAP sıralamasında oldukça aşağıda yer almıştır. Bu beklenmedik gibi görünen bulgu aslında klinik literatürle tutarlıdır: chol ve fbs, aterosklerozun yıllar içinde gelişimini hazırlayan dolaylı risk faktörleridir; ca, thal ve oldpeak ise hastalığın anlık doku, anatomi ve elektrofizyolojik düzeydeki doğrudan kanıtlarıdır.

4.5. Lojistik Regresyon Katsayılarıyla Bağımsız Doğrulama

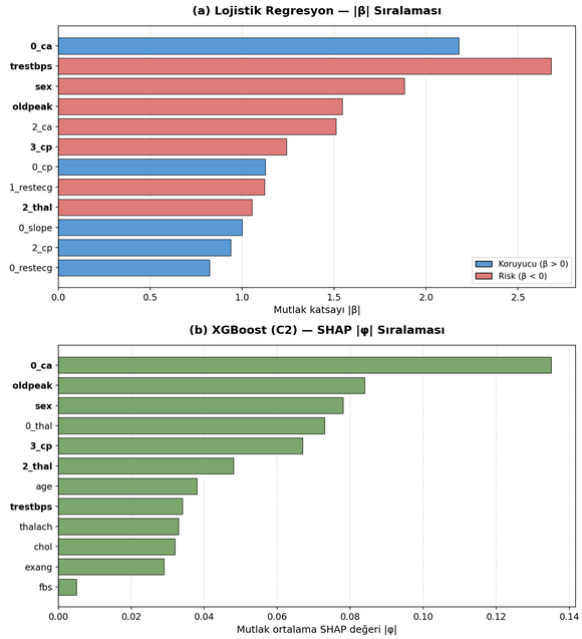
Bölüm 3.5'te açıklandığı üzere, Lojistik Regresyon kendi katsayıları üzerinden doğrudan yorumlanabilir bir model olduğu için ek bir post-hoc XAI yöntemine ihtiyaç duymaz. Bu çalışmada Lojistik Regresyon'un katsayıları, XGBoost SHAP analiziyle elde edilen sıralamayı bağımsız biçimde sınayan bir referans olarak kullanılmıştır. Tablo 6, Lojistik Regresyon modelinin en yüksek mutlak katsayı değerine sahip ilk on iki özniteliğini yön bilgisiyle birlikte sunmaktadır.

Tablo 6 Lojistik Regresyon Modelinde En Yüksek Mutlak Katsayılı On İki Öznitelik

Sıra	Öznitelik	Katsayı (β)	Yön	Klinik Yorum
1	*0_ca*	+2,179	Koruyucu	Hiç tıkalı damar bulunmaması
2	*trestbps*	-2,680	Risk	Yüksek dinlenme kan basıncı
3	*sex*	-1,882	Risk	Erkek cinsiyet
4	*oldpeak*	-1,544	Risk	Egzersiz kaynaklı ST depresyonu
5	*2_ca*	-1,511	Risk	İki damarda tıkanma
6	*3_cp*	-1,242	Risk	Asimptomatik göğüs ağrısı
7	*0_cp*	+1,126	Koruyucu	Tipik anjina
8	*1_restecg*	-1,122	Risk	EKG ST-T anormalliği
9	*2_thal*	-1,053	Risk	Reversible defekt (talasemi)
10	*0_slope*	+1,000	Koruyucu	Normal egzersiz ST eğimi
11	*2_cp*	+0,938	Koruyucu	Anjinal olmayan göğüs ağrısı
12	*0_restecg*	+0,824	Koruyucu	Normal dinlenme EKG

Tablo 6'da pozitif katsayılar (β > 0) tahmini sağlıklı sınıfa yönlendiren — yani koruyucu — öznitelikleri; negatif katsayılar ise (β < 0) tahmini hasta sınıfına yönlendiren, yani risk artırıcı öznitelikleri temsil etmektedir. Bu yön yorumu, modelin condition = 1 (hasta) değerini referans olarak aldığı KNIME varsayılan kodlamasına dayanmaktadır.

Şekil 4, Lojistik Regresyon'un |β| sıralaması (a) ile XGBoost (C2) modelinin SHAP |φ| sıralamasını (b) birlikte sunmaktadır. Görsel, iki yöntemin sıralamalarındaki örtüşmeyi doğrudan görmeye olanak tanımaktadır.



Şekil 4. *Lojistik Regresyon Katsayılarının (a) ve XGBoost SHAP Değerlerinin (b) Karşılaştırılması*

İki sıralama arasındaki örtüşme dikkat çekicidir. İlk beş sırada her iki yöntem 0_ca, oldpeak ve sex değişkenlerinde buluşmaktadır. İlk on sıraya bakıldığında ortak öznitelik sayısı altıya yükselmektedir: 0_ca, oldpeak, sex, 3_cp, 2_thal ve trestbps. İki sıralama tam olarak aynı olmamakla birlikte, üst sıraları belirleyen klinik değişken aileleri (ca, oldpeak, thal, cp, sex) birbiriyle yüksek uyum göstermektedir.

Bu örtüşme, çalışmanın yorumlanabilirlik bulguları açısından önemli bir noktayı ortaya koymaktadır. Yapısal olarak tamamen farklı iki model — doğrusal Lojistik Regresyon ile doğrusal olmayan, gradyan artırmalı XGBoost — ve birbirinden bağımsız iki yorumlama yöntemi — model katsayıları ile post-hoc SHAP — aynı klinik biyobelirteçlere işaret etmektedir. Bu sonuç, ortaya çıkan öznitelik hiyerarşisinin belirli bir algoritmanın iç işleyişine bağlı bir artefakt değil; veri setindeki gerçek klinik sinyallere ait olduğunu güçlü biçimde desteklemektedir.

4.6. Yerel SHAP Açıklamaları: İki Hasta Üzerinden Karşılaştırma

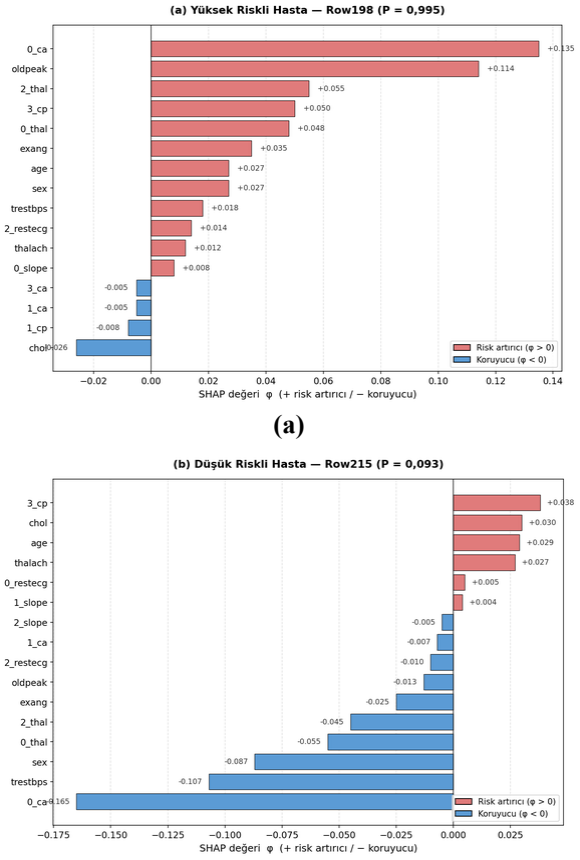
Modelin hasta düzeyindeki karar yapısını incelemek için, test setinden iki uç vaka seçilmiştir: modelin yüksek güvenle hasta olarak sınıflandırdığı bir hasta (Row198, P = 0,995) ve yüksek güvenle sağlıklı olarak sınıflandırdığı bir hasta (Row215, P = 0,093). Şekil 5(a) ve 5(b), bu iki vaka için yerel SHAP açıklamalarını sunmaktadır.

Yüksek riskli hasta (Row198) için tahmini hasta sınıfına yönlendiren öznitelikler sırasıyla 0_ca (+0,135), oldpeak (+0,114), 2_thal (+0,055), 3_cp (+0,050) ve 0_thal (+0,048) olmuştur. Bu sıralamayı exang (+0,035), age (+0,027) ve sex (+0,027) izlemektedir. Bu hasta için yalnızca chol (–0,026) ve daha küçük katkılarla 1_cp, 1_ca ve 3_ca koruyucu yönde sinyal üretmiş; ancak risk artırıcı özniteliklerin toplam etkisi açıkça baskın kalmıştır.

Düşük riskli hasta (Row215) için tablo neredeyse aynanın diğer yüzü gibidir. Tahmini sağlıklı sınıfa çeken en güçlü öznitelikler 0_ca (–0,165), trestbps (–0,107), sex (–0,087), 0_thal (–0,055) ve 2_thal (–0,045) olarak ortaya çıkmıştır. Burada en dikkat çekici bulgu, Row198 hastasında risk artırıcı yönde sinyal üreten tam aynı özniteliklerin (0_ca, 0_thal, 2_thal, sex) bu hastada koruyucu yönde çalışmasıdır.

İki vakanın karşılaştırılması üç önemli noktayı ortaya çıkarmaktadır. İlki, aynı özniteliklerin farklı hastalarda zıt yönlerde katkı yapabilmesidir. Row198'de en güçlü beş risk artırıcı öznitelikten dördü (0_ca, 0_thal, 2_thal, sex) Row215'te en güçlü beş koruyucu öznitelik içinde yer almaktadır. Bu desen, modelin değerlere kör bir biçimde değil, bağlama duyarlı biçimde karar verdiğini göstermektedir. İkincisi, yüksek güvenli tahminlerin tek bir öznitelikten değil,

klinik açıdan anlamlı bir grup öznitelikten kaynaklanması; modelin veriyi ezberlemediğini, genelleşebilir örüntüler öğrendiğini desteklemektedir. Üçüncüsü, Tek-Sıcak Kodlama'nın yerel açıklamaları sınıf bazında ayrıştırabilmesi sayesinde, hangi spesifik kategori değerinin tanıda belirleyici olduğu doğrudan okunabilmektedir.



(a)

(b)

Şekil 5. *Hastalar İçin Yerel SHAP Açıklamaları*
(a) Yüksek riskli hasta (Row198) **(b)** Düşük riskli hasta (Row215)

5. Tartışma

5.1. Lojistik Regresyon'un Sürpriz Birinciliği

Bu çalışmanın bulguları arasında en dikkat çekici olanı, basit bir doğrusal modelin XGBoost dahil tüm ağaç tabanlı modelleri geride bırakmasıdır. Yüzeysel bakışta beklenmedik bir sonuç gibi görünse de, üzerinde durulduğunda bunun gerekçesi oldukça nettir: Tek-Sıcak Kodlama, kategorik değişkenleri sahte sıralama yükünden kurtararak doğrusal modellerin matematiksel ifade gücünü doğrudan beslemektedir. Klasik tamsayı temsiliyle Lojistik Regresyon, literatürde tipik olarak %74–76 doğruluk aralığında kalır; bu çalışmada ise 10 katlı çapraz doğrulamada %84,5'e ulaşmıştır.

Ağaç tabanlı modeller (Random Forest, Gradyan Artırmalı Ağaçlar, XGBoost) için ise Tek-Sıcak Kodlama'nın etkisi farklı, hatta yer yer tersine işlemiştir. Bu modeller kategorik değişkenleri zaten dahili olarak iyi yönetebilmektedir; ancak öznitelik uzayının 13'ten 25'e genişlemesi, ağaç bölme verimliliğinde küçük bir aşınma yaratmıştır. Sonuç, kategorik kodlama stratejisinin model ailesine göre seçilmesi gerektiğini somut biçimde ortaya koyar.

Bulgu, Occam Usturası ilkesinin de klasik bir örneğidir. Veri yapısı doğrusal bir karar sınırıyla yeterince iyi ayrıştırılabiliyorsa, daha karmaşık modeller — uzun eğitim süreleri ve daha yüksek aşırı öğrenme riski ile birlikte — ek bir kazanım sağlamayabilir. Küçük örneklemli klinik veri setlerinde bu durumun göz önünde bulundurulması, model seçim sürecinin başlangıcından itibaren önemlidir.

Literatür karşılaştırması açısından bakıldığında, elde edilen XGBoost performansı (Doğruluk = 0,805; AUC = 0,878) Almutairi ve Dardouri [6] ile Mushtaq ve diğerlerinin [5] bildirdikleri seviyelere oldukça yakındır. Önemli olan nokta şudur: metodolojik olarak temiz kurgulanmış bir çerçeve, başarısız sonuçlar değil; güvenilir ve doğrulanmış sonuçlar üretmiştir.

Çalışmanın güçlü yanlarından biri de, en başarılı model olan Lojistik Regresyon'un kendi katsayıları üzerinden doğal olarak yorumlanabilir oluşudur. Tablo 6 ve Şekil 4'te sunulan analiz, Lojistik Regresyon katsayı sıralaması ile XGBoost SHAP sıralamasının üst sıralarda önemli ölçüde örtüştüğünü göstermiştir. İlk on sırada altı ortak öznitelik bulunmaktadır: 0_ca, oldpeak, sex, 3_cp, 2_thal ve trestbps. Bu örtüşme, yöntemsel üçleme (methodological triangulation) ilkesinin somut bir uygulamasıdır.

Performans ve yorumlanabilirlik eksenlerinde ortaya çıkan tabloya hesaplama maliyeti boyutunun da eklenmesi, model seçimine üçüncü bir bakış açısı kazandırmaktadır. Cleveland veri setinin küçük boyutu (n = 297) nedeniyle dört modelin de eğitim ve tahmin süreleri saniyenin altında kalmaktadır; bu ölçekte zamana dayalı bir karşılaştırma istatistiksel olarak anlamlı bir ayrım üretmemektedir. Buna karşın model karmaşıklığı açısından fark çarpıcıdır: Lojistik Regresyon yalnızca 26 katsayıdan oluşan, tek bir matematiksel fonksiyon halinde temsil edilebilen bir doğrusal modeldir. Random Forest 200 ayrı karar ağacının, XGBoost ise 100 boosting iterasyonunda biriken yüzlerce ağacın topluluğundan oluşmaktadır. Model dosyası boyutu, dağıtım altyapısı, klinik bilişim sistemlerine entegrasyon kolaylığı ve uzun vadeli bakım yükü açısından Lojistik Regresyon'un avantajı bu nedenle yalnızca yorumlanabilirlikle sınırlı değildir.

5.2. Klinik Yorumlanabilirlik

SHAP global analizinin işaret ettiği en etkili öznitelikler — 0_ca, oldpeak, sex, 0_thal ve 3_cp — kardiyoloji klinik pratiğindeki birinci derece prognostik göstergelerle yüksek uyum göstermektedir. Bu uyum, modelin yalnızca istatistiksel örüntüleri değil; klinik açıdan anlamlı ilişkileri de yakaladığına dair önemli bir kanıttır.

Diğer yandan kolesterol ve açlık kan şekerinin görece zayıf prediktörler olarak öne çıkması ilk bakışta beklenmedik gelebilir; çünkü bu iki parametre halk arasında kalp hastalığı denildiğinde akla gelen ilk değişkenlerdir. Ancak bulgu, modern kardiyoloji literatürüyle aslında çelişmemektedir. Söz konusu değişkenler aterosklerozun uzun vadeli risk faktörleridir; tek bir kesit veride hastalık durumunu doğrudan ele vermezler. Tanı için anatomik (ca), fonksiyonel (thal) ve elektrofizyolojik (oldpeak) testler her zaman daha kesin sonuç verir.

Yerel SHAP analizinde aynı özniteliklerin farklı hastalarda zıt yönlerde katkı sağlayabilmesi ise, modelin nüfus düzeyinde ortalama bir kuralı uygulamak yerine her hastayı kendi öznitelik bütünü içinde değerlendiren bir karar yapısı işlettiğini göstermektedir. Bu, klinik karar destek sistemlerinde en çok aranan özelliklerden biridir: hasta-spesifik yorumlama.

5.3. Sınırlılıklar ve Gelecek Çalışmalar

Çalışmanın güçlü yanlarına karşın bazı sınırlılıkların açıkça belirtilmesi gerekmektedir. İlk olarak, veri setinin görece küçük olması (n = 297), elde edilen bulguların genellenebilirliği açısından temkinli olunmasını gerektirmektedir. Bununla ilişkili bir diğer sınırlılık ise çalışmanın yalnızca Cleveland Heart Disease veri seti üzerinde yürütülmüş olmasıdır. Her ne kadar 10 katlı tabakalı çapraz doğrulama ile iç doğrulama gerçekleştirilmiş olsa da, farklı hasta popülasyonları ve farklı sağlık kurumlarından elde edilen veriler üzerinde dış doğrulama yapılmamıştır. Bu nedenle elde edilen sonuçların farklı klinik ortamlara ve veri kaynaklarına doğrudan genellenebileceği söylenemez. Gelecek çalışmalarda Statlog Heart Disease, Framingham Heart Study veya çok merkezli klinik veri kümeleri üzerinde gerçekleştirilecek dış doğrulama çalışmaları, önerilen yaklaşımın genellenebilirliğini daha güçlü biçimde ortaya koyacaktır.

İkinci olarak, XGBoost için yürütülen hiperparametre değerlendirmesi sistematik olmakla birlikte beş yapılandırma ile sınırlı tutulmuştur. İç içe geçmiş çapraz doğrulama (nested cross-validation) çerçevesinde gerçekleştirilecek daha kapsamlı bir hiperparametre optimizasyon süreci, parametre seçim yanlılığını daha da azaltabilir.

Üçüncü olarak, SHAP hesaplamaları KNIME'ın model-bağımsız KernelSHAP altyapısı üzerinden gerçekleştirilmiştir. XGBoost gibi ağaç tabanlı modeller için doğrudan ve daha hızlı çalışan TreeSHAP yöntemi, KNIME Analytics Platform 5.x sürümünde yerel olarak desteklenmemektedir. Bu nedenle gelecekte Python entegrasyonu kullanılarak TreeSHAP tabanlı analizlerin gerçekleştirilmesi planlanmaktadır.

Dördüncü olarak, yorumlanabilirlik analizi yalnızca SHAP yöntemi ile sınırlandırılmıştır. SHAP, literatürde yaygın olarak kullanılan ve güçlü kuramsal temellere sahip bir açıklanabilirlik yöntemi olmasına rağmen, açıklamaların farklı veri örnekleri, farklı eğitim bölmeleri veya farklı model yapılandırmaları altında ne ölçüde kararlı kaldığı bu çalışmada ayrıca değerlendirilmemiştir. LIME, karşı-olgusal açıklamalar ve açıklama kararlılığını ölçmeye yönelik yöntemlerin kullanılması, gelecekte açıklamaların güvenilirliğinin ve tutarlılığının daha kapsamlı biçimde incelenmesine katkı sağlayabilir.

Son olarak, Lojistik Regresyon modelinin katsayı analizinde Tek-Sıcak Kodlama referans kategori düşürülmeden uygulanmıştır. Bu durum bazı One-Hot kolonlarında standart hataların yükselmesine neden olmuş ve katsayıların istatistiksel anlamlılık analizlerini sınırlamıştır. Bununla birlikte, katsayıların yönü ve göreli büyüklükleri üzerinden yapılan yorumlar çalışmanın amaçları açısından yeterli düzeyde bilgi sağlamıştır.

6. Sonuç

Bu çalışma, kalp hastalığı tahmini için KNIME Analytics Platform üzerinde kodsuz olarak geliştirilen ve metodolojik gereklilikleri tek bir çerçevede birleştiren açıklanabilir yapay zekâ (XAI) tabanlı bir iş akışı sunmuştur. UCI Cleveland Heart Disease veri seti üzerinde Lojistik Regresyon, Random Forest, Gradyan Artırmalı Ağaçlar ve XGBoost modelleri 10 katlı tabakalı çapraz doğrulama altında karşılaştırılmış; en yüksek performans Lojistik Regresyon modeli tarafından elde edilmiştir (Doğruluk = 0,845 ± 0,065; Kappa = 0,687; AUC = 0,908). Wilcoxon işaretli sıra testi sonuçları, Lojistik Regresyon'un Random Forest ve XGBoost ile karşılaştırılabilir performans sergilediğini, yalnızca Gradyan Artırmalı Ağaçlar karşısında istatistiksel olarak anlamlı bir üstünlük sağladığını göstermiştir.

Çalışmanın temel bulguları üç başlık altında özetlenebilir. İlk olarak, nominal kategorik değişkenlerin Tek-Sıcak Kodlama ile dönüştürülmesi özellikle doğrusal modeller açısından önemli bir performans avantajı sağlamıştır. İkinci olarak, normalizasyon parametrelerinin her çapraz doğrulama katlamasında yalnızca eğitim verisi üzerinde öğrenilmesi, veri sızıntısını önleyerek daha güvenilir performans tahminleri elde edilmesine katkı sağlamıştır. Üçüncü olarak, Lojistik Regresyon katsayı analizi ile XGBoost SHAP analizinin üst sıralarında ortak olarak belirlenen öznitelikler (0_ca, oldpeak, sex, 3_cp, 2_thal ve trestbps), elde edilen klinik biyobelirteç hiyerarşisinin model türünden bağımsız olarak veri setindeki gerçek sinyalleri yansıttığını göstermiştir.

Elde edilen sonuçlar, performans, yorumlanabilirlik ve model karmaşıklığı birlikte değerlendirildiğinde, basit fakat doğru yapılandırılmış doğrusal modellerin küçük ve orta ölçekli klinik veri kümelerinde karmaşık topluluk modelleriyle rekabet edebileceğini göstermektedir. Bu durum, klinik karar destek sistemlerinde yalnızca performans odaklı değil, aynı zamanda açıklanabilirlik ve uygulanabilirlik odaklı model seçimlerinin önemini ortaya koymaktadır.

Bununla birlikte, çalışmanın yalnızca Cleveland Heart Disease veri seti üzerinde gerçekleştirilmiş olması, sonuçların farklı hasta popülasyonlarına doğrudan genellenebileceği anlamına gelmemektedir. Bu nedenle önerilen çerçevenin farklı veri kümeleri ve gerçek klinik ortamlarda doğrulanması önemli bir araştırma yönü olarak değerlendirilmektedir. Gelecek çalışmalarda çok merkezli klinik veri kümeleri üzerinde dış doğrulama gerçekleştirilmesi, nested cross-validation tabanlı hiperparametre optimizasyonunun uygulanması, TreeSHAP yönteminin Python entegrasyonu ile kullanılması, LIME ve karşı-olgusal açıklamalar gibi tamamlayıcı XAI yöntemlerinin değerlendirilmesi ve klinisyen geri bildirimleriyle desteklenen kullanıcı odaklı karar destek sistemlerinin geliştirilmesi planlanmaktadır.

Yazar Katkısı

Kavramsallaştırma – Muhammed Yusuf Kaya (MYK); Metodoloji – MYK; Yazılım – MYK; Biçimsel analiz – MYK; Araştırma – MYK; Deneysel performans – MYK; Veri işleme – MYK; Veri iyileştirme – MYK; Görselleştirme – MYK; Literatür taraması – MYK; Yazan (taslak) – MYK; İnceleme ve düzenleme – Bashar Alhajahmad (BA); Danışmanlık – BA.

Çıkar Çatışması Beyanı

Yazarlar, bu makalenin araştırılması, yazarlığı ve/veya yayınlanması ile ilgili olarak herhangi bir çıkar çatışması beyan etmemiştir.

Teşekkür

Bu çalışma için herhangi bir kurum, kuruluş veya fonlama ajansından finansal destek alınmamıştır.

Kaynaklar

[1] World Health Organization. (2023). Cardiovascular diseases (CVDs) — Key facts. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). Erişim tarihi: 10/05/2026

[2] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

[3] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

[4] Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book/

[5] Mushtaq, M., Mehmood, Z., Butt, A. A., Shabir, M. Y., & Khan, M. N. U. (2025). Causal and explainable machine learning framework for heart disease prediction using XGBoost and SHAP. *Journal of Computing & Biomedical Informatics*, 10(1).

[6] Almutairi, M., & Dardouri, S. (2025). Intelligent hybrid modeling for heart disease prediction. *Information*, 16(10), 869. https://doi.org/10.3390/info16100869

[7] Rezk, N. G., Alshathri, S., Sayed, A., Hemdan, E. E. D., & El-Behery, H. (2024). XAI-augmented voting ensemble models for heart disease prediction: A SHAP and LIME-based approach. *Bioengineering*, 11(10), 1016. https://doi.org/10.3390/bioengineering11101016

[8] Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). Heart Disease [Dataset]. *UCI Machine Learning Repository*. https://doi.org/10.24432/C52P4X

[9] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[10] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[11] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., & Wiswedel, B. (2009). KNIME: The Konstanz Information Miner. *SIGKDD Explorations Newsletter*, 11(1), 26–31. https://doi.org/10.1145/1656274.1656280