Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi
Afyon Kocatepe University – Journal of Science and Engineering
https://dergipark.org.tr/tr/pub/akufemubid

Araştırma Makalesi / Research Article
e-ISSN: 2149-3367
DOI: https://doi.org/10.35414/akufemubid.1595455
AKÜ FEMÜBİD **25** (2025) 055202 (1106-1113)
AKU J. Sci. Eng. **25** (2025) 055202 (1106-1113)

Makine Öğrenimi Modelleri ile Kırılganlık Seviyesi Sınıflandırma ve Veri Tipi Etkisi

Makale Bilgisi / Article Info
Alındı/Received: 03.12.2024
Kabul/Accepted: 19.04.2025
Yayımlandı/Published: 01.10.2025

Frailty Level Classification with Machine Learning Models and Data Type Effect

Mehmet Halit ÖZALP*, Didem ÖLÇER

Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara, Türkiye

© 2025 The Authors | Creative Commons Attribution-Noncommercial 4.0 (CC BY-NC) International License

Öz

Günümüzde kırılganlık, yaşlanan nüfus ile önemli bir sorun haline gelmektedir. Kırılganlık erken teşhis edilmesi ve uygun müdahaleler ile azaltılabilen veya tersine çevrilebilen geriatrik bir sendromdur. Bu nedenle bu çalışmada, KNN (K-Nearest Neighbors), Naïve Bayes, XGBoost (Extreme Gradient Boosting), RF (Random Forests), MLP (Multi-Layer Perceptron), 1D-CNN (1D Convolutional Neural Network) modelleri ile, STAIR testi sırasında toplanan EKG, üç eksenli ivme ölçer ve bu iki sensor verilerinin birleşiminden elde edilen üç ayrı veri seti üzerinde zaman alanı öznitelikleri kullanılarak kırılganlık sınıflandırma problemine çözüm aranmaktadır. Çalışmanın sonucunda RF ve XGBoost modellerinin en başarılı sonuçları verdiği, ayrıca EKG verilerinin kırılganlık tahmininde kullanımının akselerometre verilerinin kullanımından daha yüksek doğrulukta tahminleme yapılabildiği görülmüştür. Ayrıca Kırılganlık tahmininde daha doğru sonuçlar veren verilerin takibi ile birlikte, daha kesin sonuçlarla kişiler, kırılganlık durumlarını izleyebilecek ve sadece gerekli olduğunda bir uzman görüşüne danışacaktır.

Anahtar Kelimeler *Kırılganlık Sınıflandırma; STAIR Testi; İvmeölçer; EKG; Makine Öğrenmesi; Öznitelik Çıkarımı*

Abstract

Today, frailty is becoming an important problem with the ageing population. Frailty is a geriatric syndrome that can be reduced or reversed by early detection and appropriate interventions. Therefore, in this study, KNN (K-Nearest Neighbors), Naïve Bayes, XGBoost (Extreme Gradient Boosting), RF (Random Forests), MLP (Multi-Layer Perceptron), 1D-CNN (1D Convolutional Neural Network) models are used to solve the frailty classification problem by using time domain features on three separate data sets obtained from the combination of ECG, triaxial accelerometer and these two sensor data collected during the STAIR test. As a result of the study, it is seen that RF and XGBoost models give the most successful results, and the use of ECG data in fragility prediction can be predicted with higher accuracy than the use of accelerometer data. In addition, with the follow-up of data that give more accurate results in frailty prediction, people will be able to monitor their frailty status with more precise results and consult an expert opinion only when necessary.

Keywords *Frailty Classification; STAIR Test; Accelerometer; ECG; Machine Learning; Feature Extraction*

1. Giriş

Yaşlanan dünya nüfusu ile birlikte yaşlılarda kırılganlık durumu her geçen gün daha önemli bir sağlık ve araştırma konusu haline gelmektedir. Kırılganlık (frailty) 65 yaş üstü bireylerde genellikle fiziksel, bilişsel ve psikolojik işlevlerde görülen azalmalarla karakterize edilen ve bireyin bağımsız yaşam kapasitesini sınırlayan bir durumdur. Kırılganlık hareket kaybı, sakatlık, hastaneye yatış ve hatta ölüm gibi olumsuz sağlık sorunlarının oluşma riskini arttıran bir sendromdur (Hoogendijk et al. 2019). Neyse ki kırılganlık, egzersiz eğitimi gibi zamanında bir müdahale ile azaltılabilir veya tersine çevrilebilir geriatrik bir sendromdur (Panhwar et al. 2019).

Gelişen teknoloji ile giyilebilir teknolojiler hayatın vazgeçilmez bir parçası olmaya başlamıştır. Akıllı telefon ve akıllı saatler gibi giyilebilir teknolojiler yaygınlaşmış ve bu cihazlar sayesinde günlük fiziksel aktivite (adım sayısı, yürüyüş süresi, koşu süresi gibi) ve kalp atışlarının ölçülerek istatistiksel veri olarak kullanıcılara sunulması sağlık verilerinin izlenmesini gündelik hayatın bir parçası haline getirmiştir. Bu durum kırılganlık probleminin çözümü için giyilebilir teknolojilerin oldukça önemli bir çözüm aracı olduğunu ortaya koymaktadır.

Sağlığın izlenmesi ile ilgili olarak Bozkurt Keser ve Keskin (2023) çalışmalarında, kalp yetmezliği hastalarının tıbbi kayıtlarına dayalı olarak ölüm veya hayatta kalma durumlarının tahmin edilmesini hedeflemektedir. 285 gün süreyle izlenen 299 kalp yetmezliği hastasına ait 13 öznitelikten oluşan veri seti ile çalışılmıştır. Veri seti %75 eğitim ve %25 test olarak ayrılmış RF, Aşırı Gradyan Arttırma (AGA) ve Yapay Sinir Ağları (YSA) modellerine girdi olarak verilmiştir. Sonuç olarak YSA ve RF kalp yetmezliği olmayan hastaları daha iyi tahmin ederken AGA kalp yetmezliğini tahmin etmede daha başarılı görülmüştür. Kırılganlık tahmininde en başarılı model Random Forest (RF) olurken kırılganlık tespitindeki

*Sorumlu Yazar/Corresponding Author: Mehmet Halit ÖZALP e-posta/e-mail: halitt.ozalp@gmail.com

belirleyici değişkenlerin yaş, sandalye kalkma testi, hane halkı zenginliği, denge problemleri ve öz-değerlendirilmiş sağlık durumu olduğu belirlenmiştir.

Kikuchi et. al. (2020), Japonya'da yaşayan yaşlı bireyler arasında fiziksel aktivite yoğunluğu ve belirli süreli hareketsiz zamanın kırılganlık durumu ile olan ilişkisini incelemiştir. 65-84 arasında 511 yaşayan yaşlı bireyin fiziksel aktiviteleri ve hareketsiz zamanları, üç eksenli bir İvmeölçer kullanılarak ölçülmüş ve istatiksel yöntemlerle analiz edilmiştir. Sağlıklı bireylerde fiziksel aktivitenin daha yoğun olduğu kırılgan veya ön kırılgan bireylerde daha az yoğunlukta fiziksel aktivite tespit edilmiştir.

Razjouyan et. al. (2018), toplulukta yaşayan yaşlı ve kırılgan bireylerde kırılganlık durumunu tespit etmek için giyilebilir sensörlerin kullanımını incelemiştir. 60 yaş ve üzeri 153 yaşlı bireye, Fiziksel aktivite ve davranışları izlemek için 48 saat boyunca göğüs bölgesine takılan bir kolye sensör verilmiştir. Sensörden elde edilen verilerin analizi, lojistik regresyon modeli ve ROC eğrisi kullanılarak gerçekleştirilmiştir. Sonuçlar giyilebilir bir sensör kullanarak günlük ev içi izleme yoluyla ön-kırılganlığın tanımlanabileceğini göstermiştir.

Mizuguchi et. al. (2024), yaşlı kalp yetmezliği hastalarında klinik kırılganlık ölçeğini otomatik olarak değerlendirmek için ML ve derin öğrenme tekniklerinin uygulanabilirliğini araştırmışlardır. 75 yaş ve üzeri semptomatik kronik kalp yetmezliği tanısı alan 417 yaşlı hasta değerlendirilmiştir. Yürüyüş analizi, akıllı telefon kamerası kullanılarak derin öğrenme tabanlı bir tahmin kütüphanesi aracılığıyla yapılmıştır. Yürüyüş parametreleri ve 128 anahtar özellik, Light Gradient Boosting Machine modeli kullanılarak tahmin edilmiştir. Bu model ile gerçek ve tahmin edilen klinik kırılganlık ölçeği arasında mükemmel bir uyum göstermiştir.

Szczepanowski et al. (2024), kalp yetmezliği hastalarında ML algoritmalarının ne kadar başarılı olduğunu test etmişlerdir. Çeşitli ML modellerinin denendiği çalışmada AdaBoost modelinin %100'e yakın bir doğruluk oranıyla en iyi performansı gösterdiği belirtilirken kırılganlık skorunun fiziksel, psikolojik ve sosyal bileşenlerinin ise kırılganlığın tahmininde kilit rol oynadığı ifade edilmiştir.

Chen et al. (2024), KOAH hastalarında kırılganlığın etkili bir şekilde tahmin edilebilmesi için bazı risk faktörleri belirlemişler ve bu faktörler üzerine kurulu bir tahminleme modeli geliştirmeye çalışmışlardır. Depresyon, sigara içme durumu, cinsiyet, sosyal aktiviteler, dislipidemi, astım ve yaşanılan yer gibi toplamda 34 değişkenin incelendiği çalışmada Extreme Gradient Boost (XGBoost), RF, Logistic regression, support vector machines (SVM), multilayer perceptron, light gradient-boosting machine (LightGBM) ve K-nearest neighbors (KNN) modelleri test için kullanılmıştır. En yüksek doğruluğa XGBoost modeli, %91,5 doğruluk, %87,3 duyarlılık ve %91,1 özgüllük oranlarıyla en başarılı model olarak çıkarken depresyon, sigara içme, cinsiyet, sosyal aktiviteler, dislipidemi, astım ve yerleşim yeri, kırılganlık gelişiminde en önemli öngörücü faktörler olarak belirlenmiştir.

Erdaş ve Ölçer (2022), kırılganlığın tahmin edilmesi amacıyla, açık kalp cerrahisi sonrası kardiyak rehabilitasyon programına katılan kırılgan hastalardan elde edilen üç eksenli ivmeölçer verilerini çeşitli makine öğrenimi algoritmalarının testinde kullanmıştır. Çalışmalarında Extra Trees algoritması en başarılı ML algoritması olmuştur.

Bu çalışmada literatürden farklı olarak kırılganlık tespiti için kilit sensor verileri incelenecektir. Literatürde kırılganlık tespitinin başarı oranını arttırmak amacıyla bazı kilit değişkenler bulmak için çalışılmıştır. Ancak kırılganlığın sürekli izlenmesi ve bu sayede erken teşhisinin yapılabilmesi için basit sensor verileriyle yüksek doğrulukta sonuç almak kilit bir öneme sahiptir. Bu nedenle üç eksenli ivmeölçer ve elektrokardiyografi (EKG) verileri ile çeşitli algoritmalar test edilecektir.

2. Materyal ve Metot

2.1 Genel Yapı

STAIR aktivitesi esnasında toplanan üç eksenli ivmeölçer ve EKG verilerinin makine öğrenimi modellerinin besleyerek denetimli bir sınıflandırma problemi olarak ele alınmaktadır. İlgili sensörün çalışma frekansları göz önünde bulundurularak 4 saniye zaman ölçeğinde alt diziler oluşturulmuştur. Ardışık 2 bölüm arasında %50'lik bir zaman örtüşmesine izin verilen bu alt dizilerin içerisinde kırılganlık ile ilgili bilgi barındırdığı varsayımında bulunulmuştur. Sonrasında, ham veri içeren bu dizilerin ivmeölçer verileri için X, Y, Z eksenlerinde ve EKG verisinde ayrı ayrı zaman öznitelikleri çıkarılmıştır.

2.2 Veri Seti

Çalışmada kullanılan veriler, Sokas et al. (2022) tarafından PhysioNet aracılığıyla sağlanan halka açık veri kümesinden elde edilmiştir. Veri seti açık kalp ameliyatı geçirmiş hastalar üzerinden toplanmıştır. Bu veri setinde 80 farklı hastanın her biri toplamda beş farklı fiziksel aktiviteye katılmış ve bu aktiviteler esnasında eş zamanlı olarak EKG ve ivmeölçer verileri toplanmıştır. Bu veriler sensörler arasındaki doğal frekans farkları nedeniyle farklı kayıtlarda tutulmuştur. EKG sensörü 130Hz ile örnekleme

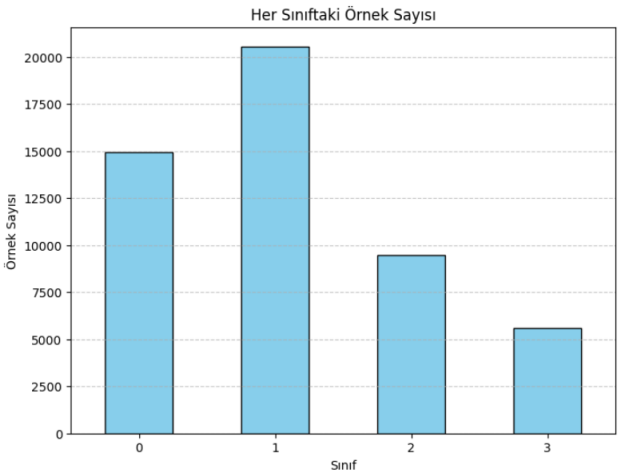
yaparken ivme ölçer için bu frekans 200Hz aralığındadır. 80 hastalık bu veri setinden 1 hasta kaydındaki adreslemelerin hatalı olduğu düşünüldüğünden kayıt veri setinden çıkarılarak 79 hasta kaydı üzerinden çalışmaya devam edilmiştir.

Kırılganlığın tespiti için kullanılan fiziksel aktivite testlerinden biri STAIR testidir. Bu veri setindeki STAIR testinin uygulamasında hastadan yardım almaksızın 12 basamaklı bir merdiveni çıkması istenmektedir. Diğer fiziksel aktiviteler (örneğin yürüme, oturup kalkma gibi) de kaydedilmiş olmasına rağmen, STAIR testinin kırılganlığın tespitinde daha belirleyici olabileceği düşünülmüştür. Bunun nedeni, merdiven çıkma esnasında hastaların daha yüksek efor sarf etmesi ve bu durumun hem ivmeölçer hem de EKG sinyallerinde daha belirgin biyofizyolojik tepkiler oluşturmasıdır. Bu bağlamda, çalışmada kırılganlık seviyelerinin daha iyi ayrıştırılabileceği öngörülerek analizler STAIR testi üzerinden gerçekleştirilmiştir. Edmonton Kırılganlık Ölçeği (EFS), bireylerin kırılganlık düzeylerini değerlendirmek amacıyla kullanılan bir ölçüm aracıdır. EFS, kırılganlıkla ilişkili dokuz alt alanı (örneğin, bilişsel durum ve ilaç kullanımı gibi) inceleyerek genel bir puan oluşturur. Bu ölçek, en fazla 17 puana ulaşabilir ve bu puan en yüksek kırılganlık düzeyini temsil eder (Rolfson at al., 2006). Ölçek kapsamında, alınan puanlara göre kırılganlık dereceleri şu şekilde sınıflandırılır: 0-4 puan kırılganlık olmadığını, 5-6 puan başlangıç seviyesindeki kırılganlığı, 7-8 puan düşük seviyede kırılganlığı, 9-10 puan orta seviyede kırılganlığı ve 11'in üzerindeki puanlar yüksek seviyede kırılganlığı ifade eder (Aygör at al., 2018).

Çalışmada kullanılan veri setinde sınıflandırmalar yapıldığında Şekil 1.'de görülebileceği gibi 4 farklı sınıf mevcuttur. Bunun nedeni kullanılan veri setinde EFS skoru 11 ve üzeri hasta bulunmamasından kaynaklanmaktadır. Çalışmada kullanılan modeller, 14953 örnekten oluşan sınıf 0'a (EFS 0-4 puan), 20555 örnekten oluşan sınıf 1'e (EFS 5-6 puan), 9479 örnekten oluşan sınıf 2'ye (EFS 7-8 puan), 5612 örnekten oluşan sınıf 3'e (EFS 9-10 puan) ait örneklerden oluşan toplamda 50599 örnekten oluşan bu veri seti üzerinde eğitilip test edilmiştir.

2.3 Öznitelik Çıkarımı

Üç farklı eksende veri ölçen ivmeölçer ve tek eksende veri ölçen EKG verilerinin her bir ekseni için 14 ayrı zaman alanı özniteliği çıkarılmıştır (Erdaş et al. 2016). Zaman alanı özniteliklerinin kullanılma nedeni, ivmeölçer ve EKG verilerinin zaman içindeki değişikliklerini analiz ederek bu değişimlerden kırılganlığı tespit etmektir. Bu çalışmada frekans alanı öznitelikleri değerlendirilmemiştir, çünkü temel amaç, biyofizyolojik sinyallerin zaman içindeki dinamiklerini ve istatistiksel değişimlerini analiz etmektir. Frekans analizi genellikle periyodik veya ritmik hareketlerin incelenmesi için kullanıldığından, bu çalışmada zaman alanı özniteliklerine odaklanılmıştır.



Şekil 1. Veri Seti Sınıf Dağılımı

Her bir eksen için hesaplanan öznitelikler; ortalama, standart sapma, medyan, çeyrekler arası fark, en düşük ve en yüksek değerler, bu değerler arasındaki fark, mutlak sapmaların ortalaması ve medyanı, ortalamanın üzerindeki ve altındaki değerlerin sayıları, lokal maksimumların (tepe noktalarının) sayısı, veri setinin çarpıklığı ve basıklığı şeklinde sıralanmıştır. Gerçekleştirilen bu öznitelik çıkarımları sonucunda ivme ölçer verileri üç eksen için ayrı ayrı yapıldığından 42 adet öznitelik ile, EKG verileri tek eksene sahip olduğundan 14 adet öznitelik ile, EKG ve ivme ölçer verilerinin birleşimi ise 56 adet öznitelik ile temsil edilmektedir.

2.4 Makine Öğrenmesi Modelleri

KNN (K-Nearest Neighbors) algoritması ML metotları içerisinde özellikle sınıflandırma problemleri içinde oldukça popüler ve tercih edilen bir algoritmadır. KNN her bir veri noktasını, belirli bir mesafe metriğine göre "k" adet en yakın komşusuyla ilişkilendirir. Bu mesafe Euclidean, Chebyshev, Manhattan gibi mesafeler ile ölçülür. Algoritmanın temel parametresi "k" parametresidir ve bu, komşu olarak seçilecek veri noktalarının sayısını ifade eder (Barrashet al. 2019).

Naïve Bayes algoritması, olasılık temelli ve özellikler arasında bağımsızlık varsayımına dayanan basit bir ML algoritmasıdır. Bayes teoremini kullanarak çalışır ve genellikle sınıflandırma çalışmalarında kullanılır. Naive Bayes sınıflandırıcısının temel avantajı, özellikle büyük özellik alanlarında etkili olması ve hesaplama açısından verimli olmasıdır (Sano et al. 2023).

XGBoost (Extreme Gradient Boosting), denetimli öğrenme (supervised learning) problemleri için geliştirilmiş, güçlü

ve esnek bir ML algoritmasıdır. XGBoost, temel olarak, gradyan artırma (gradient boosting) yönteminin optimize edilmiş ve genişletilmiş bir versiyonudur. Gradyan artırma, karar ağaçları gibi bir dizi modelin, hataları azaltarak birbirini takip edecek şekilde eğitildiği bir topluluk öğrenme tekniğidir. Her bir yeni modelin, mevcut modelin tahmin hatalarını öğrenmeye çalıştığı ve her eklenen karar ağacıyla birlikte daha doğru sonuç veren bir algoritmadır. Aynı zamanda her yeni ağaç eklenirken modelin toplam hatasını azaltmak için her bir ağaçta mevcut olan ağırlıkları güncellemektedir. Aşırı öğrenmenin önüne geçmek için karmaşıklığı kontrol altında tutar. Paralel veri işlemeleri ve eksik verileri otomatik olarak işleme kabiliyeti ile oldukça avantajlı bir algoritmadır (Chen et al 2016).

RF (Random Forests) algoritması, Leo Breiman tarafından 2001 yılında önerilen ve makine öğreniminde kullanılan güçlü bir sınıflandırma ve regresyon yöntemidir. RF, birden fazla karar ağacının birleştirilmesiyle oluşturulan bir modeldir. Her bir ağaç, eğitim verilerinin rastgele bir alt kümesi üzerinde eğitilir ve her düğümde, rastgele seçilmiş bir özellik alt kümesi kullanılarak bölünme işlemi gerçekleştirilir (Breiman 2001).

MLP (Multi-Layer Perceptron), yapay sinir ağlarının temel bir türüdür ve çoğunlukla sınıflandırma ve regresyon gibi problemler için kullanılır. Bu model, giriş katmanı, bir veya daha fazla gizli katman ve çıkış katmanı olmak üzere birkaç katmandan oluşur ve her katmandaki nöronlar, bir sonraki katmandaki tüm nöronlarla bağlantılıdır. Giriş katmanındaki nöronlar, modelin işlediği veri özelliklerini temsil ederken, gizli katmanlar veriyi işleyerek öğrenme sürecine katkıda bulunur. Çıkış katmanı, modelin elde ettiği tahmin sonuçlarını oluşturur. MLP'de nöronlar, doğrusal olmayan aktivasyon fonksiyonları kullanarak veriyi işler (örneğin, ReLU veya sigmoid gibi). Modelin eğitim süreci, geri yayılım algoritmasıyla gerçekleştirilir; bu algoritma, modelin tahmin hatalarını minimize etmek için ağırlıkları güncellemektedir (Wang et al. 2024).

1D-CNN (1D Convolutional Neural Network), tek boyutlu veriler üzerinde çalışan bir tür yapay sinir ağıdır. Genellikle zaman serisi verileri, ses verileri, sinyaller veya metinler gibi tek boyutlu veriler üzerinde kullanılmaktadır. Bu tür ağlar, verilerdeki özellikleri ve desenleri çıkarmak için bir dizi filtre veya çekirdek kullanarak evrişim (convolution) işlemleri gerçekleştirir. Adaptif öğrenme, hız, verimlilik, öznitelik çıkarımı yapabilme gibi hususlardaki avantajları sayesinde oldukça tercih edilen bir algoritmadır (Kiranyazet al. 2015).

3. Bulgular

KNN, Naïve Bayes, XGBoost, RF, MLP ve 1D-CNN modelleri ile Kırılganlığın tahmini ve Kırılganlık tahmini için hangi sensör verisinin daha iyi sonuç verdiğinin tespiti amaçlanmaktadır.

Bu çalışmada model performansını değerlendirmek için 10 katlı çapraz doğrulama (k-fold cross-validation) yöntemi kullanılmıştır. Çapraz doğrulama sürecinde, veri seti her iterasyonda farklı eğitim ve test bölümlerine ayrılarak model eğitilmiş ve test edilmiştir. Genelleştirilmiş bir sonuç elde etmek amacıyla, her iterasyonda hesaplanan performans metriklerinin ortalaması alınmıştır. Modelin farklı veri kümelerinde gösterdiği değişkenliği ölçmek için standart sapma değerleri hesaplanmış ve sonuçların tutarlı olduğu görülmüştür.

KNN algoritmasında en iyi 'k' değeri ve uzaklık metriğinin bulunması için Grid Search kullanılmıştır. Kullanılan Grid Search modelinde potansiyel K kümesi 1,3,5,7,9,11,13,15 ve potansiyel uzaklık metrikleri Öklid, Manhattan, Chebyshev şeklindedir. Bu fonksiyon verilen potansiyel k kümesindeki ve uzaklık metrikleri kümesindeki tüm elemanları dener ve verilen çapraz doğrulama metriğine göre doğrulayarak en iyi sonucu veren k ve uzaklık metriğini geri döndürür.

XGBoost ve RF için karar ağacı sayısı 100 olarak seçilmiştir. MLP için tek bir gizli katman (hidden layer) içermekte ve bu katmanda 100 nöron bulunmakta ve maksimum iterasyon sayısı olarak 1000 ve random_state 101 olarak belirlenmiştir. Son olarak 1D-CNN için 512 filtre içeren tek bir convolutional layer ardından bir flatten katmanı, 1024 ve 512 nörondan oluşan 2 dense katmanı ve bir çıkış katmanından oluşmuş ve relu ile softmax aktivasyon fonksiyonları kullanılmıştır.

Algoritmaların başarılarını değerlendirmek amacıyla Doğruluk (accuracy) (Aydemir 2022), kesinlik (precision) (Erdaş et al. 2021), duyarlılık (Recall) (Erdaş et al. 2021), F1 skoru (F1 Score) (Erdaş et al. 2021) ve Matthews Korelasyon Katsayısı (MCC/Matthews Correlation Coefficient) (Alan ve Karabatak 2020) metrikleri kullanılmıştır.

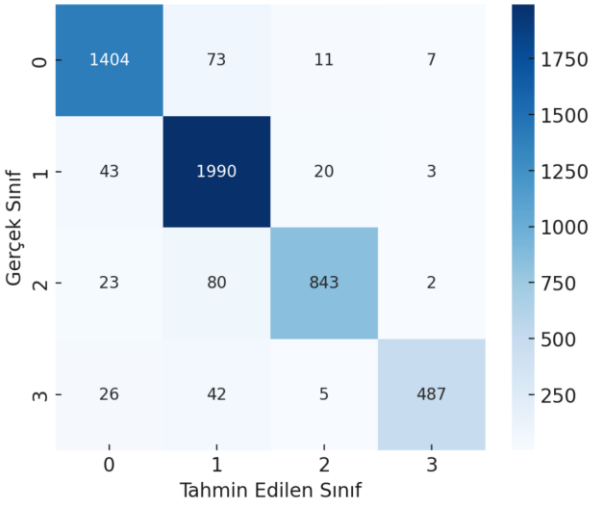
Çizelge 1'de STAIR testi esnasında toplanan ivmeölçer verisi ile 4 saniyelik zaman aralığında oluşturulmuş veri seti için sonuçlar görülmektedir. Grid Search ile yapılan aramada KNN için "k" değeri 7, mesafe ölçütü ise Manhattan olarak bulunmuştur. En iyi sonucu 0,905 MCC oranı ile RF modeli verirken en kötü sonuç ise Naive Bayes modelinde olduğu görülmektedir.

Çizelge 1. STAIR testi İvmeölçer verileri 4 saniye

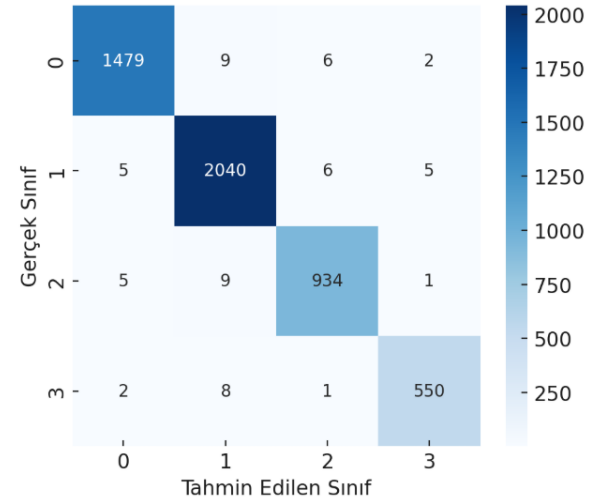
Modeller	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	MCC
KNN	0,825 ±0,004	0,826 ±0,005	0,825 ±0,005	0,823 ±0,005	0,747 ±0,006
NAİVE BAYES	0,225 ±0,009	0,364 ±0,016	0,225 ±0,009	0,192 ±0,011	0,079 ±0,013
XGBoost	0,926 ±0,004	0,926 ±0,004	0,926 ±0,006	0,926 ±0,005	0,894 ±0,006
RF	0,934 ±0,004	0,935 ±0,004	0,934 ±0,006	0,933 ±0,005	0,905 ±0,007
MLP	0,728 ±0,016	0,739 ±0,015	0,728 ±0,022	0,724 ±0,016	0,608 ±0,022
1D CNN	0,801 ±0,007	0,804 ±0,013	0,772 ±0,013	0,785 ±0,006	0,713 ±0,008

Çizelge 2. STAIR testi EKG verileri 4 saniye

Modeller	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	MCC
KNN	0,883±0,003	0,883 ±0,003	0,883±0,005	0,882±0,004	0,832±0,005
NAİVE BAYES	0,241±0,001	0,600 ±0,058	0,241±0,006	0,161±0,022	0,159±0,011
XGBoost	0,989±0,002	0,989 ±0,002	0,989±0,003	0,989±0,002	0,984±0,003
RF	0,989±0,002	0,989 ±0,001	0,989±0,003	0,989±0,002	0,984±0,002
MLP	0,851±0,012	0,852 ±0,012	0,851±0,015	0,850±0,013	0,787±0,018
1D CNN	0,878±0,014	0,875 ±0,018	0,868±0,020	0,870±0,016	0,826 ±0,021



Şekil 2. İvmeölçer- RF Karmaşıklık Matrisi



Şekil 3. EKG- RF Karmaşıklık Matrisi

Şekil 2'de verilen RF modeline ait karmaşıklık matrisi, modelin sınıfları yüksek doğrulukla ayırt edebildiğini göstermektedir. Tüm sınıflar için doğru tahmin sayıları oldukça yüksektir ve yanlış sınıflandırmalar diğer modellere kıyasla daha düşüktür. Sınıf 1, 1990 doğru tahminle en iyi performans gösteren sınıftır, ancak 43 örnek yanlışlıkla Sınıf 0 ve 20 örnek Sınıf 2 olarak tahmin edilmiştir. Sınıf 0, 1404 doğru tahminle başarılı bir şekilde sınıflandırılmıştır, ancak 73 örnek Sınıf 1 olarak yanlış tahmin edilmiştir. Sınıf 2, 843 doğru tahminle güçlü bir başarı sergilese de 80 örneğin Sınıf 1 olarak yanlış sınıflandırılması dikkat çekmektedir. Sınıf 3, 487 doğru tahminle en düşük hata oranına sahip sınıftır, ancak 42 örneğin Sınıf 1 olarak tahmin edilmesi modelin bu sınıflar arasında küçük bir hata yaptığını göstermektedir. Genel olarak, RF modeli yüksek doğruluk oranı ve düşük hata oranları ile en iyi performans gösteren modellerden biridir, ancak Sınıf 1 ve Sınıf 2 arasındaki yanlış tahminlemeler modelin daha da iyileştirilebileceğini göstermektedir.

Çizelge 2'de STAIR testi esnasında toplanan EKG verisi ile 4 saniyelik zaman aralığında oluşturulmuş veri seti için sonuçlar görülmektedir. Grid Search ile yapılan aramada KNN için "k" değeri 5 olarak belirlenirken, mesafe ölçütü olarak ise Manhattan bulunmuştur. En iyi sonucu 0,984 MCC oranı ile RF ve XGBoost modelleri verirken en kötü sonuç ise Naive Bayes modelinde olduğu görülmektedir.

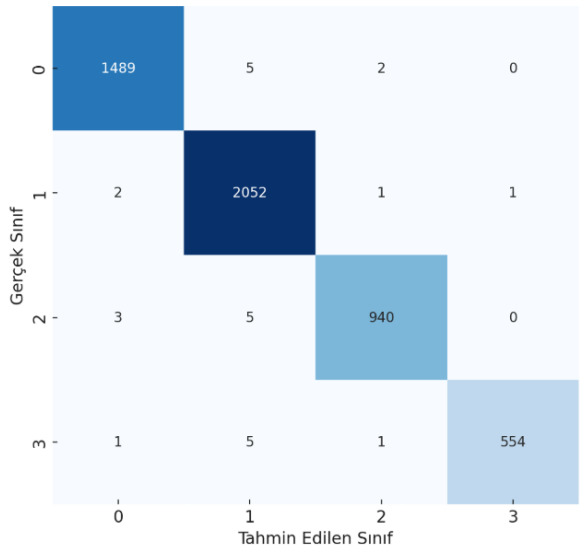
Şekil 3'te verilen RF modeline ait karmaşıklık matrisi, modelin son derece yüksek doğruluk oranına sahip olduğunu ve sınıfları başarılı bir şekilde ayırt edebildiğini göstermektedir. Tüm sınıflar için yanlış sınıflandırma oranları oldukça düşük seviyededir. Sınıf 0, 1479 doğru

tahmin ile başarılı bir şekilde sınıflandırılmış olup yalnızca 9 örnek Sınıf 1, 6 örnek Sınıf 2 ve 2 örnek Sınıf 3 olarak yanlış tahmin edilmiştir. Sınıf 1 için 2040 doğru tahmin yapılmış ve sadece 5'er örnek Sınıf 0 ve Sınıf 3, 6 örnek ise Sınıf 2 olarak yanlış tahmin edilmiştir. Sınıf 2, 934 doğru tahmine sahip olup yalnızca 9 örnek Sınıf 1, 5 örnek Sınıf 0 ve 1 örnek Sınıf 3 olarak yanlış tahmin edilmiştir. Sınıf 3 ise 550 doğru tahmin ile en düşük hata oranına sahip sınıf olup, yalnızca 8 örnek Sınıf 1, 2 örnek Sınıf 0 ve 1 örnek Sınıf 2 olarak yanlış tahmin edilmiştir. Genel olarak, RF modeli yüksek doğruluk oranı ve düşük hata oranları ile en başarılı modellerden biri olarak öne çıkmaktadır. Yanlış tahminlerin büyük kısmı komşu sınıflar arasında gerçekleşmiştir, ancak bu oran oldukça düşüktür. XGBoost modeli ile benzer bir doğruluk göstermektedir ve her iki model de veri seti için oldukça başarılı sınıflandırma performansı sergilemektedir.

Çizelge 3. STAIR testi İvmeölçer ve EKG verileri 4 saniye

Modeller	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	MCC
KNN	0,932 ±0,003	0,933± 0,004	0,932 ±0,004	0,932 ±0,004	0,817 ±0,005
NAİVE BAYES	0,269 ±0,005	0,415 ±0,018	0,269 ±0,003	0,213 ±0,011	0,163 ±0,001
XGBoost	0,995 ±0,001	0,995 ±0,001	0,995 ±0,001	0,995 ±0,001	0,993 ±0,002
RF	0,995 ±0,001	0,995±0,001	0,995 ±0,002	0,995 ±0,001	0,992 ±0,002
MLP	0,913 ±0,010	0,914 ±0,012	0,913 ±0,012	0,913 ±0,010	0,774 ±0,014
1D CNN	0,929±0,006	0,930 ±0,007	0,923 ±0,007	0,926 ±0,006	0,813±0,008

Çizelge 3'de STAIR testi esnasında toplanan ivmeölçer ve EKG verileri ile 4 saniyelik zaman aralığında oluşturulmuş veri seti için sonuçlar görülmektedir. Grid Search ile yapılan aramada KNN için "k" değeri 5 olarak belirlenirken, mesafe ölçütü olarak ise Manhattan bulunmuştur. En iyi sonucu 0,993 MCC oranı ile RF ve XGBoost modelleri verirken en kötü sonuç ise Naive Bayes modelinde olduğu görülmektedir.



Şekil 4. İvmeölçer ve EKG- RF Karmaşıklık Matrisi

Şekil 4'te verilen RF modeline ait karmaşıklık matrisi, modelin sınıfları oldukça yüksek doğrulukla ayırt edebildiğini göstermektedir. Model, yanlış tahminleri minimum seviyeye indirerek tüm sınıflar için başarılı bir sınıflandırma yapmıştır. En yüksek doğruluğa sahip sınıf, 2052 doğru tahmin ile Sınıf 1 olurken, diğer sınıflar da benzer şekilde yüksek doğruluk oranları sergilemiştir. Yanlış tahminlerin büyük çoğunluğu komşu sınıflar arasında gerçekleşmiş olup, bu oran oldukça düşüktür. Özellikle Sınıf 3, yalnızca 7 yanlış tahmin ile en iyi ayrılan sınıflardan biri olmuştur. Genel olarak, RF modeli, yüksek doğruluk oranı ve düşük hata oranları ile en başarılı modellerden biri olarak öne çıkmaktadır ve sınıflar arasında güçlü bir ayrım yapabilmektedir.

4. Sonuçlar ve Tartışma

STAIR testi esnasında kırılganlığın değerlendirilmesi konusunda en başarılı modeller hiç şüphesiz ki XGBoost ve RF modelleri olmuştur. Naïve Bayes modeli ise hiçbir veri setinde başarılı olamamıştır. MLP modelinde veri seti değişikliğinin başarıya etkisi akselerometre veri seti ile diğer veri setleri karşılaştırıldığında oldukça net bir biçimde ortaya çıkmaktadır.

KNN ve 1D-CNN modelleri için ise akselerometre veri setinde 0,70 seviyelerinde olan MCC değeri EKG ve tümleşik veri setlerinde 0,80 bandının üstüne çıktığı ve dolayısıyla daha güvenilir tahminleme yapabildiği görülmüştür.

Bununla birlikte 4 saniyelik akselerometre ve EKG verilerinin birlikte kullanılması ile oluşturulan veri seti ile kırılganlığın tespitinde en başarılı sonuçlar alınmıştır. Bu sonuç göstermektedir ki merdiven inip çıkarken ivme ölçer ve kalp ritminin birlikte değerlendirilmesi kırılganlığın doğru tahminlenme oranını kayda değer oranda arttırmaktadır.

Szczepanowski et al. (2024), bir anket ile elde ettiği fiziksel ve psikolojik değişkenleri kullandığı çalışmasında %100'lük Doğru Negatif tahminleri ve %93'lük Doğru Pozitif tahminlerinin ortalaması alındığında %96,5'lik değer ile bu çalışmadaki ivme ölçer verileri kıyaslanabilir.

Bu kıyaslama neticesinde psikolojik değişkenlerin kırılganlığın tahmininde etkisi olduğunu görülebilmektedir.

Erdaş ve Ölçer (2022), yaptıkları çalışmada kırılganlığın tahminlemesini bir regresyon problemi olarak ele almış ve kalkıp yürüyüp sonrasında geri dönülmekten oluşan TUG testi esnasında toplanmış veriler kullanılmıştır. Bu çalışmada kırılganlığın tahmininde Rastgele Ağaçları ve Ekstra Ağaçlar gibi ağaç temelli modeller en yüksek başarı oranlarına sahip olmuşlardır. Bu çalışmalar ile ivme ölçer verilerinin kullanılarak elde edilen sonuçların tutarlı olduğu görülmüştür.

İlginç olarak sadece EKG verilerinin kullanılarak kırılganlık tahminlemesi yapılan sonuçlar akselerometre verilerinin kullanılarak yapılan tahminlerden daha isabetli olduğu görülmüştür.

Bu çalışmanın sonucunda fiziksel aktivite esnasında Kırılganlığın tahmin edilebilmesi için tek başına akselerometre verilerinin yeterli olmadığı, EKG verilerinin akselerometre verilerine nazaran Makine Öğrenmesi modellerinin kırılganlığı tahmin etmesinde daha yararlı olduğu sonucuna ulaşılmıştır.

Akselerometre ve EKG verilerinin birlikte kullanımında ise RF ve XGBoost modellerinin MCC değerlerinin arttığı ancak diğer modellerin başarılarının sadece EKG verilerinin kullanılmasından daha başarısız sonuç verdiği gözlenmiştir. Bunun nedeninin veri setinin boyutunun artması ile modelin karmaşıklaşması ve bu nedenle başarının düşmesi olduğunu düşünüyoruz. Model karmaşıklığının azaltılması ile bu konu tekrar test edilmesi gerekmektedir. Aynı zamanda zaman ölçeğindeki değişikliklerin başarıya olan etkisi de ayrı bir araştırma konusu olarak göze çarpmaktadır. Gelecek çalışmalar, bu iki veri tipinin entegrasyonu üzerine yoğunlaşarak, kırılganlık tespitinde diğer aktivitelerinde eklenmesi ile daha kapsamlı ve etkili modellerin geliştirilmesine katkıda bulunabilir. Ayrıca, farklı açık kaynak veri setleri üzerinde yapılacak karşılaştırmalı analizler, bu çalışmanın bulgularının doğrulanmasına ve modelin genelleme yeteneğinin değerlendirilmesine yardımcı olacaktır.

Etik Standartlar Bildirgesi

Bu çalışma Dr. Öğretim Üyesi Didem Ölçer danışmanlığında Mehmet Halit Özalp tarafından 22.11.2024 tarihinde tamamlanan "Makine öğrenmesi ile kırılganlık tahmini ve zaman ölçeği etkisi" başlıklı yüksek lisans (Tez no: 920935) tezinden türetilmiştir.

Yazarlık Katkı Beyanı

Yazar 1: Araştırma, Deneyleme, Doğrulama, Yazma-orijinal taslak, Görselleştirme,

Yazar 2: Kaynaklar, Araştırma, Deneyleme, Doğrulama, Metodoloji, Yazma – orijinal taslak,

Çıkar Çatışması Beyanı

Yazarların bu makalenin içeriğiyle ilgili olarak beyan edecekleri hiçbir çıkar çatışması yoktur.

Verilerin Kullanılabilirliği

Bu çalışma sırasında oluşturulan veya analiz edilen tüm veriler, yayınlanan bu makaleye dahil edilmiştir.

5. Kaynaklar

Alan, A., ve Karabatak, M., 2020. Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 32(2), 531-540. https://doi.org/10.35234/fumbd.738007

Aydemir, Ö., 2022. Dengesiz veri kümelerinin sınıflandırılmasında poligon alan metriğinin sınıflandırıcı performans değerlendirilmesi için kullanılması. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 27(2), 194–205. https://doi.org/10.53433/yyufbed.1066340

Aygör, H., Fadıloğlu, Ç., Şahin, S., Aykar, F., and Akçiçek, F., 2018. Validation of Edmonton Frail Scale into elderly Turkish population. Archives of Gerontology and Geriatrics, 76, 133-137. https://doi.org/10.1016/j.archger.2018.02.003

Barrash, S., Shen, Y., and Giannakis, G.B., 2019. Scalable and Adaptive kNN for Regression Over Graphs. IEEE CAMSAP 2019, 241-245. https://doi.org/10.1109/CAMSAP45676.2019.9022509

Bozkurt Keser, S., ve Keskin, K., 2023. Kalp Yetmezliği Hastalarının Sağ Kalım Tahmini: Sınıflandırmaya Dayalı Makine Öğrenmesi Algoritmalarının Bir Uygulaması. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 23, 362-369. https://doi.org/10.35414/akufemubid.1033377

Breiman, L., 2001. Random Forests. Machine Learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., and Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-784. https://doi.org/10.1145/2939672.2939785

Chen, Y., Yu, Y., Yang, D., Zhang, W., Kouritas, V., and Chen, X., 2024. Developing and validating machine learning-based prediction models for frailty occurrence in those with chronic obstructive pulmonary disease. Journal of Thoracic Disease, 16(4), 2482-2498. https://doi.org/10.21037/jtd-24-416

Erdaş, Ç.B., Atasoy, I., Açıcı, K., and Oğul, H., 2016. Integrating Features for Accelerometer-based Activity Recognition. Procedia Computer Science, 98, 522-527. https://doi.org/10.1016/j.procs.2016.09.070

Erdaş, Ç.B., and Ölçer, D., 2022. Prediction of Frailty Grade Using Machine Learning Models. In 2022 Medical Technologies Congress (TIPTEKNO), Antalya, Türkiye, 1-4. https://doi.org/10.1109/TIPTEKNO56568.2022.9960172

Erdaş, Ç.B., Sümer, E., and Kibaroğlu, S., 2021. Neurodegenerative disease detection and severity prediction using deep learning approaches. Biomedical Signal Processing and Control, 70, 103069. https://doi.org/10.1016/j.bspc.2021.103069

Hoogendijk, E., Afilalo, J., Ensrud, K., Kowal, P., Onder, G., and Fried, L., 2019. Frailty: implications for clinical practice and public health. The Lancet, 394(10206), 1365-1375. https://doi.org/10.1016/s0140-6736(19)31786-6

Kikuchi, H., Inoue, S., Amagasa, S., Fukushima, N., Machida, M., Murayama, H., Fujiwara, T., Chastin, S., Owen, N., & Shobugawa, Y. (2020). Associations of older adults' physical activity and bout-specific sedentary time with frailty status: Compositional analyses from the NEIGE study. Experimental Gerontology. https://doi.org/10.1016/j.exger.2020.111149

Kiranyaz, S., Ince, T., and Gabbouj, M., 2015. Real-Time Patient-Specific ECG Classification by 1D Convolutional Neural Networks. IEEE Transactions on Biomedical Engineering, 63(3), 664-675. https://doi.org/10.1109/TBME.2015.2468589

Leme, D.E.C., and Oliveira, C., 2023. Machine Learning Models to Predict Future Frailty in Community-Dwelling Middle-Aged and Older Adults: The ELSA Cohort Study. The Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 78(11), 2176–2184. https://doi.org/10.1093/gerona/glad127

Mizuguchi, Y., Nakao, M., Nagai, T. (2024). Machine learning–based gait analysis to predict clinical frailty scale in elderly patients with heart failure. *European Heart Journal - Digital Health, 5*(2), 152–162. https://doi.org/10.1093/ehjdh/ztad082

Panhwar, Y., Naghdy, F., Naghdy, G., Stirling, D., and Potter, J., 2019. Assessment of frailty: a survey of quantitative and clinical methods. BMC Biomedical Engineering, 1-20. https://doi.org/10.1186/s42490-019-0007-y

Razjouyan, J., Naik, A. D., Horstman, M. J., Kunik, M. E., Amirmazaheri, M., Zhou, H., Sharafkhaneh, A., & Najafi, B. (2018). Wearable sensors and the assessment of frailty among vulnerable older adults: An observational cohort study. Sensors, 18(5), 1336. https://doi.org/10.3390/s18051336

Rolfson, D., Majumdar, S., Tsuyuki, R., Tahir, A., and Rockwood, K., 2006. Validity and reliability of the Edmonton Frail Scale. Age and Ageing, 35(5), 526-529. https://doi.org/10.1093/ageing/afl041

Sano, A.V.D., Stefanus, A.A., Madyatmadja, E.D., Nindito, H., Purnomo, A., and Sianipar, C.P.M., 2023. Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier. Procedia Computer Science, 227, 482-489. https://doi.org/10.1016/j.procs.2023.10.549

Sokas, D., Butkuvienė, M., Tamulevičiūtė-Prascienė, E., Beigienė, A., Kubilius, R., Petrėnas, A., and Paliakaitė, B., 2022. Wearable-based signals during physical exercises from patients with frailty after open-heart surgery (version 1.0.0). *PhysioNet*. https://doi.org/10.13026/mp8k-7p27

Szczepanowski, R., Uchmanowicz, I., Pasieczna, A.H., Sobecki, J., Katarzyniak, R., Kołaczek, G., Lorkiewicz, W., Kędras, M., Dixit, A., Biegus, J., Wleklik, M., Gobbens, R.J.J., Hill, L., Jaarsma, T., Hussain, A., Barbagallo, M., and Morabito, F.C., 2024. Application of machine learning in predicting frailty syndrome in patients with heart failure. Advances in Clinical and Experimental Medicine, 33(3), 309-315. https://doi.org/10.17219/acem/184040

Wang, Y., Wei, S., Zuo, R., et al., 2024. Automatic and real-time tissue sensing for autonomous intestinal anastomosis using hybrid MLP-DC-CNN classifier-based optical coherence tomography. Biomedical Optics Express, 15(4), 2543-2557. https://doi.org/10.1364/BOE.521652