

Kalp Yetmezliği Hastalarının Sağ Kalım Tahmini: Sınıflandırmaya Dayalı Makine Öğrenmesi Algoritmalarının Bir Uygulaması

Sinem BOZKURT KESER¹, Kemal KESKİN²

¹ Eskişehir Osmangazi Üniversitesi, Bilgisayar Mühendisliği Bölümü, 26040, Eskişehir, Türkiye

² Eskişehir Osmangazi Üniversitesi, Elektrik – Elektronik Mühendisliği Bölümü, 26040, Eskişehir, Türkiye

e-posta: sbozkurt@ogu.edu.tr

ORCID ID: <https://orcid.org/0000-0002-8013-6922>

Sorumlu yazar e-posta: kkeskin@ogu.edu.tr

ORCID ID: <https://orcid.org/0000-0002-3969-2396>

Geliş Tarihi: 06.12.2021

Kabul Tarihi: 30.03.2023

Öz

Anahtar kelimeler

Kalp yetmezliği;
Makine öğrenmesi;
Rastgele Orman;
Aşırı Gradyan
Arttırma; Yapay Sinir
Ağları

Kardiyo-vasküler hastalıklar dünya genelinde en çok ölüme sebep olan hastalıklar arasında yer almaktadır. Bir kardiyo-vasküler hastalık olan kalp yetmezliği, kalbin vücudun ihtiyaç duyduğu kanı pompalayamaması durumudur. Ülkemizde sıklıkla görülen bu hastalığın sonucu olarak ölümler yaşanmaktadır. Bu çalışmada kalp yetmezliğe sahip hastaların sağ kalım veya ölüm durumlarının tahmin edilmesi için makine öğrenmesi tabanlı bir yaklaşım önerilmektedir. Üç farklı sınıflandırma algoritması kullanılarak önerilen yöntemin etkinliği değerlendirilmektedir. Gerçekleştirilen deneylerde, Yapay Sinir Ağı algoritması ile en yüksek doğruluk değerine (86.67%) ulaşılmıştır. Önerilen yöntem, ölüm riskinin yüksek olduğu kalp yetmezliği hastalarına daha etkin ve uygun tedavi planlarının hazırlanması açısından yol gösterici olacaktır.

Prediction of Survival of Heart Failure Patients: An Application of Classification-Based Machine Learning Algorithms

Abstract

Keywords

Heart failure;
Machine learning;
Random Forest;
Extreme Gradient
Boosting; Artificial
Neural Networks

Cardio-vascular diseases are among the diseases that cause the most deaths worldwide. Heart failure, a cardiovascular disease, is a condition in which the heart cannot pump the blood that the body needs. Deaths occur as a result of this disease, which is frequently seen in our country. In this study, a machine learning-based approach is proposed to predict survival or death of patients with heart failure. The effectiveness of the proposed method is evaluated using three different classification algorithms. In the experiments performed, the highest accuracy values (86.67%) was achieved with the Artificial Neural Network algorithm. The proposed method will guide the preparation of more effective and appropriate treatment plans for heart failure patients with a high risk of death.

© Afyon Kocatepe Üniversitesi

1. Giriş

Dünya Sağlık Örgütü verilerine göre, dünya genelindeki ölümlerin %31 gibi büyük bir oranla en başta gelen sebebi kardiyo-vasküler hastalıklardır ve her yıl yaklaşık olarak 17,9 milyon bireyin yaşamı bu hastalıklar nedeniyle sonlanmaktadır. Kalp yetmezliği, koroner kalp hastalığı, serebrovasküler hastalık, romatizmal kalp hastalığı sıklıkla karşılaşılan kardiyo-vasküler hastalıklardır (İnt. Kyn. 1). Kalp yetmezliği, kalbin çeşitli sebeplerle vücudun

ihtiyaç duyduğu miktardaki kanı pompalayamaması durumudur. Türkiye’de ilk olarak 2012 yılında erişkinlerin kalp yetmezliği yaygınlığının tespit edilmesine yönelik yapılan çalışmada, yaklaşık 2,5 milyon bireyin kalp yetmezliğine sahip olduğu ve yaygınlık oranınının 45 yaş üstü bireyler için %2,9 olduğu belirtilmiştir (Degertekin vd. 2012). Birçok ülkede olduğu gibi ülkemizde de yaygın olarak görülen kalp yetmezliği, ölüm sebeplerinin başında gelmektedir. Bu sebeple, kalp yetmezliğinin ölümle

sonulanmadan nce tespit edilmesi byk nem tařıtmaktadır. Her ne kadar yapılan klinik uygulamalarla kalp yetmezliđi ile iliřkili olaylar tespit edilebilse de, genel olarak yksek dođruluk seviyesine ulařılamamıřtır (Buchan *et al.* 2019). Bu dođrultuda, uzmanların karar verme srecini hızlandırmada ve daha yksek dođruluk seviyelere eriřimin sađlanması yapay zeka teknolojilerine bařvurulmaktadır.

Son yıllarda birok alanda etkinliđi kanıtlanmış bir yapay zeka teknolojisi olan makine đrenmesi ve veri madenciliđi, gnmzde zellikle karmařık sađlık problemlerinde, tanı, hastalık izleme, tedavi planı ve hastaların sonucunun tahmini iin sıklıkla kullanılmaktadır. Dřk sol ventrikl ejeksiyon fraksiyonuna sahip kalp yetmezliđi hastalarında ani kardiyak lmlerin tahmin oranını ykseltmek iin yapılan alıřmada, implante edilebilen kardiyoverter defibrilatrden yararlanma olasılıđı en yksek olan kalp yetmezliđi hastalarını belirlemeye ynelik yeni stratejiler geliřtirilmiřtir (Meng *et al.* 2019). Ejeksiyon fraksiyonuna bađlı kalp yetmezliđi yařayan hastalardaki hastaneye yatıř ve lm oranını tespit etmeye ynelik model geliřtirmek ve dođrulamak iin geleneksel lojistik regresyon ynteminin yanında Rastgele Orman ve Destek Vektr Makineleri (SVM, Support Vector Machines) yaklařımları da kullanıldı. Geliřtirilen modelin zamanla deđiřen takip verileri olmadan temel hasta zelliklerini kullanması ne ıkan eksikliđidir (Angraal *et al.* 2020). Uygun ve yerinde mdahale edebilmek iin kalp yetmezliđi hastalarının takibinin yapılması nemlidir. 47 znitelige sahip bir veri seti kullanılarak kalp yetmezliđi hastalarının 30 gnlk yeniden kabul veya lmn tahmin etmek iin ok katmanlı algılayıcı tabanlı bir model geliřtirildi. alıřmada spesifik klinik verilerin kullanılmaması ve sadece 65 yař ve st hastalara yer verilmesi zayıflık olarak ortaya ıkmaktadır (Awan *et al.* 2019). 11 farklı laboratuvar deđiřkeni iin 970 korunmuř ejeksiyon fraksiyonuna bađlı kalp yetmezliđi olan hastadan veriler alındı. Ayrıca 290 hasta da dođrulama grubu olarak kayıt edildi. Yaklařık 5 yıl sren takip sırasında, tm nedenlere bađlı lmler birincil son nokta olarak kullanılırken, bileřik son noktalar ise ikincil son nokta olarak seildi.

Komorbidite yk, kardiyak anormallikler ve uzun vadeli prognoz ile karakterize edilen 3 farklı fenogrup bir gdmsz đrenme yaklařımı olan kmeleme stratejisi ile ayrıřtırıldı. Fenogrup 1 en dřk riske sahip iken, fenogrup 3 en yksek riskli olarak bulundu (Gu *et al.* 2021, Segar *et al.* 2020). Benzer olarak, bir bařka alıřmada atrial fibrilasyon, anemi ve bbrek hastalıklarını da gz nnde bulundurarak 6 fenogrup oluřturuldu (Hedman *et al.* 2020). Pakistan'daki 299 kalp yetmezliđi hastası iin Qlattice adında sembolik regresyon yntemine dayanan yeni bir model ile minimal bir matematiksel dnřm seti tanımlandı ve bu set Cox modelinde tahmin iin kullanıldı (Wilstup and Cave, 2021). Topluluk ađaları makine đrenimi tekniklerini kullanarak geliřtirilen modelde, Ařırı Gradyan Arttırma (XGBoost, Extreme Gradient Boosting) diđer topluluk ađaları yntemlerine gre daha iyi bir performans gstermiřtir (Moreno-Sanchez, 2020). Bir diđer alıřmanın sonuları, kalp yetmezliđine sahip hastaların sadece serum kreatinin ve ejeksiyon fraksiyon zelliklerine dayanarak yapılacak sađ kalım tahminininin, 11 adet znitelikten oluřan veri setinin kullanılmasının tm veri seti kullanılmasının yerine daha iyi performans gstereceđini ortaya koydu (Chicco and Jurman, 2020).

Bu alıřmada, laboratuvar kayıtlarına dayanan, 13 znitelik ve 299 rneklem ile oluřturulan bir veri kmesi, kalp yetmezliđi hastalarının sađ kalma veya lm durumlarının tahmini iin kullanılmıřtır. Tahmin iin makine đrenmesi alanında popler olan, Rastgele Orman, Yapay Sinir Ađları (YSA – Artificial Neural Network) ve Ařırı Gradyan Arttırma (AGA – Extreme Gradient Boosting) yntemleri kullanılmıřtır. Yapılan deneylerde elde edilen sonular analiz edildiđinde literatrde aynı veri kmesini kullanarak yapılan benzer alıřmalara kıyasla daha iyi performans deđerlerine ulařılmıřtır (Chicco and Jurman, 2020; Erdas ve Olcer, 2020).

2. Materyal ve Metot

Bu alıřmada, kalp yetmezliđine bađlı olarak hastaların hayatta kalma veya lme durumları tahmin edilmektedir. Bu amala, hastaların tıbbi kayıtlarını ieren bir veri seti üzerinde makine

öğrenmesi algoritmalarının performansları analiz edilmektedir. Çalışmanın takip eden alt bölümlerinde, veri seti, veri ön işleme, makine öğrenmesi algoritmaları ve performans ölçütleri anlatılmaktadır.

2.1 Veri Seti

Kalbin vücuda yeteri kadar kan pompalamaması sonucu görülen kalp yetmezliği sigara kullanımı, yüksek tansiyon, diyabet, stres gibi faktörler ile daha hızlı ilerlemektedir. Bu hastalığa bağlı hayatta kalma analizi için bir model geliştirmek üzere hastaların belirli bir süre takip edilmesi ve hasta öldüğünde hasta ile ilgili verilerin kayıt altına alınması gerekmektedir. Bu amaçla, Kaliforniya Üniversitesi - Irvine Makine Öğrenmesi Deposu'nda paylaşılan kalp yetmezliği klinik veri seti kullanılmıştır (Aktaş Potur ve Erginel, 2021, Aydın, 2021, Chicco and

Jurman, 2020, Erdas ve Olcer, 2020). Bu veri seti, Nisan ve Aralık 2015 tarihleri arasında 285 gün takip edilen ve New York Kalp Cemiyeti'nin (NYHA) konjestif kalp yetersizliği sınıflamasına göre Sınıf 3 veya 4 kalp yetmezliği olarak sınıflandırılan 299 hastanın verilerini içermektedir. Veri setinde 105 kadın ve 194 erkek hasta bulunmaktadır. Hematokrit seviyeleri, %36'dan düşük olan hastaların anemi olduğu kabul edilmiştir. Veri setindeki hastaların 96'sı (%32.1) belli bir takip süresinden sonra ölmüştür. Bu hastalardan on üç farklı parametre toplanarak veri seti oluşturulmuştur. Bu parametrelerin bazıları sadece ikili (boolean) veri iken bazıları sürekli (continuous) değerlerdir. Toplanan veriler, kısa açıklamalar ve değer aralıkları ile ilgili ayrıntılar Çizelge 1'de sunulmaktadır.

Çizelge 1. Veri setinde kullanılan parametreler, açıklamaları ve değer aralıkları

Öznitelik	Açıklama	Ölçüm Tipi	Değer Aralığı
Yaş	Hasta yaşı	yıl	[40,...,95]
Anemi	Hemoglobin düşüklüğü	ikili	0, 1
Kreatin fosfokinaz	Kandaki CPK enzimi seviyesi	meg/L*	[23,...,7861]
Diyabet	Kandaki glukoz miktarı	ikili	0, 1
Ejeksiyon Fraksiyonu	Kalbin her atmasından sonra kalbe gelen kanın geri pompalandığı miktar	yüzde	[14,...,80]
Yüksek tansiyon	Hasta hipertansiyona sahip olup/olmadığı	ikili	0, 1
PLT	Kandaki trombosit miktarı	kiloplatelets/mL	[25.01,...,850.00]
Cinsiyet	Kadın veya Erkek	ikili	0, 1
Serum kreatinin	Kandaki serum kreatinin miktarı	mg/dL*	[0.50,...,9.40]
Serum sodyum	Kandaki serum sodyum miktarı	mEq/L*	[114,...,148]
Sigara kullanımı	Hastanın sigara kullanıp /kullanmadığı	ikili	0, 1
Zaman	Takip dönemi	gün	[4,...,285]
Ölüm olayı (hedef)	Takip döneminde hasta ölmüşse	ikili	0, 1

*meg/L: microgram per liter, mL: microliter, mEq/L: milliequivalents per litre

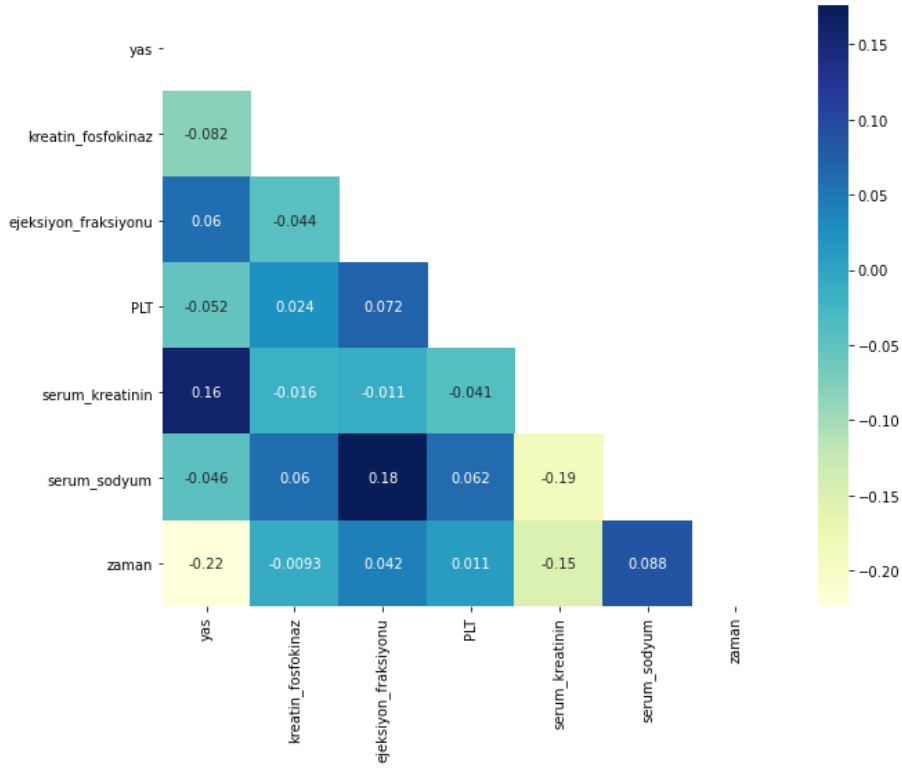
2.2 Veri Ön İşleme

Veri setinde kategorik ve sürekli öznitelikler bulunmaktadır. Bu özniteliklerin birbirleriyle ilişkilerini tespit etmek üzere Şekil 1 ile gösterilen ilişki haritası oluşturulmuştur.

Kas dokusunun hasar görmesine bağlı olarak kreatinin fosfokinaz enzim seviyesinin kandaki değeri yüksek çıkmaktadır. Bu durum, kalp yetmezliğinin göstergeleri arasında sayılmaktadır. Ejeksiyon fraksiyonu, kasılma sırasında sol ventrikülün yüzde olarak ne kadar kan pompalandığını gösterirken serum kreatinini ise böbreklerin fonksiyonelliğini tespit etmede kullanılan kas metabolizmasının ürettiği bir atıktır. Yüksek serum kreatinini, böbreklerin düzgün

çalışmadığına işaret etmektedir. Serum sodyum testi ile kandaki sodyum seviyesi belirlenmektedir. Düşük sodyum seviyeleri kalp yetmezliği belirtileri arasındadır (Chicco and Jurman, 2020).

Veri setinde kayıp veri bulunmamaktadır. Makine öğrenmesi modellerini uygulamadan önce veriler %75 eğitim ve %25 test verisi olarak ayrılmıştır. Bu işlemden sonra, sürekli özniteliklere standartlaştırma işlemi uygulanmıştır. Sürekli özniteliklere standartlaştırma işlemi uygulanmazsa, yüksek varyanslı öznitelikler model uydurma sürecine hâkim olabilir ve modelin beklendiği gibi diğer özniteliklerden (daha düşük varyanslı) doğru bir şekilde öğrenmemesine neden olabilir.



Şekil 1. Sürekli öznitelikler arasındaki ilişki haritası

2.3 Sınıflandırma Algoritmaları

Bu çalışmada, kalp yetmezliğine bağlı olarak hastaların hayatta kalma veya ölme durumları tahmin etmek üzere RO, YSA ve AGA modelleri kullanılmıştır.

RO, 2001 yılında Breiman tarafından geliştirilmiş birçok karar ağacı ile oluşturulmuş bir topluluk algoritmasıdır (Breiman 2001). Karar ağacı algoritmalarından farklı olarak kök düğümü bulma ve düğümleri bölme işlemlerini rastgele olarak gerçekleştirir. Yeniden örnekleme yöntemi ile veri setinden farklı alt kümeler oluşturulur ve her bir alt örneklem için bir karar ağacı modeli oluşturulur. Karar ağaçları, sınıflandırma ve regresyon ağaçları (CART, Classification and Regression Trees) yöntemi kullanılarak üretilir. Sınıflandırma işleminde her ağacın tahmin sonucu alınır ve oy yönteminin sınıflandırıcısının ne olacağına çoğunluk tarafından karar verilir. RO hızlı çalışan sınıflandırma algoritması olup aşırı uyum sorununa karşı dirençlidir. Çok sayıda karar ağacının sınıflandırıcı olarak kullanılması ve her bir sınıflandırıcıdan elde edilen tahminlerden çoğunluk oyu ile genel bir tahmin üretilmesi, tahminlerdeki hata oranını ve yanlışlığı azaltır.

YSA, sınıflandırma, modelleme, tahmin gibi birçok günlük yaşam probleminin çözümünde başarılı bir yöntemdir. Çok Katmanlı Algılayıcı (ÇKA, Multi-layer Perceptron), doğrusal olmayan problemlerin çözümü için sıklıkla kullanılan bir YSA modelidir. En az üç katmandan oluşan giriş ve çıkış katmanları arasında en az bir katman kullanan ileri beslemeli, geri yayımlı bir ağıdır (Haykin 1999). İleri yayılım aşamasında, ağın çıkış ve hata değeri hesaplanırken, geriye doğru yayılma aşamasında hesaplanan hata değerini en aza indirecek şekilde katmanlar arasındaki bağlantı ağırlık değerleri güncellenir. YSA gücü, esnekliği ve kullanım kolaylığı nedeniyle birçok karmaşık en iyileme problemlerinde etkili çözümler üretilmesini sağlamaktadır. Son yıllarda klasik tekniklerle çözülemeyen problemler için sinir ağlarının uygulanmasına ilgi artmış ve YSA birçok tıbbi uygulamada başarıyla kullanılmaktadır.

AGA, Karar Ağacı'nı temel alan ölçeklenebilir, taşınabilir ve hesaplamalı olarak derlenmiş bir topluluk öğrenme algoritmasıdır Gradyan Arttırma Algoritması'nı kullanan bu yöntem Düzenli Arttırma yöntemi olarak da adlandırılır ve paralel işleme işlevselliği sağlar. Gradyan Arttırma algoritması en iyileme problemini, adım yönünü belirleme ve adım boyutunu bulmak üzere iki temel adımda çözmeye çalışır. AGA ise adım boyutunu ve yönünü tek

seferde bulur. AGA, dizi modellerindeki hataları azaltmayı amaçlar. Bellek ve donanım kaynaklarının en iyi şekilde kullanımını hedefleyen bu algoritma son zamanlarda sınıflandırma ve regresyon problemlerinde sıklıkla tercih edilmektedir (Chen and Guestrin, 2016).

2.4 Performans Ölçütleri

Gerçek sınıf etiketleri ile hedef sınıf etiketleri arasındaki ilişkinin gösteriminde karmaşıklık matrisinden yararlanılır (Han and Kamber, 2001). Sınıflandırma işlemi sonucu elde edilen performans değerlerini (doğru pozitif (DP), doğru negatif (DN),

Çizelge 3. Performans Ölçütleri

Ölçüt	Açıklama	Denklem
Doğruluk (ACC)	Doğru olarak sınıflandırılan örneklem sayısının toplam örneklem sayısına oranı	$\frac{DP + DN}{DP + YP + DN + YN}$
Hassasiyet (PREC)	Sınıflandırma algoritmasının doğru tahmin ettiği pozitif örneklem sayısının pozitif olarak tahmin ettiği toplam örneklem sayısına oranı	$\frac{DP}{DP + YP}$
Hatırlama (RC)	Sınıflandırma algoritmasının pozitif örneklemeleri doğru tahmin etme gücü	$\frac{DP + YN}{DN}$
Özgüllük (SP)	Sınıflandırma algoritmasının ne kadar gerçek negatif değeri doğru tahmin ettiğinin ölçüsü	$\frac{DN + YP}{DN}$
F-ölçütü (FS)	Hassasiyet ve hatırlama ölçütlerinin geometrik ortalaması	$\frac{2 * PREC * RC}{PREC + RC}$

Alıcı İşletim Eğrisi (AİE, Receiver Operating Characteristics) ve Eğri Altında Kalan Alan (AKA, Area Under Curve): Performans değerlendirmede kullanılan bir diğer analiz AİE analizidir. Sınıflandırma algoritmasının, hassasiyet ve özgüllük açısından birlikte değerlendirilebilmesine olanak sağlar. Bu eğrinin oluşturduğu koordinat sisteminin x ekseninde (1-özgüllük), y ekseninde ise duyarlılık yer alır (Harrington, 2012; Mitchell, 1997).

3. Deneysel Sonuçları

Bu çalışmada, kalp yetmezliğine bağlı olarak hastaların hayatta kalma veya ölme durumları tahmin edilmektedir. Bu doğrultuda, Python programlama dili ile geliştirilmiş makine öğrenmesi ve veri madenciliği alanında yayınlanmış algoritmalara ait fonksiyonları içeren popüler bir kütüphane olan scikit-learn kütüphanesinden yararlanılmıştır. Deneyler Windows 10 işletim

yanlış pozitif (YP), yanlış negatif (YN)) göstermek için kullanılan bir görselleştirme aracıdır (Çizelge 2).

Çizelge 2. Karmaşıklık matrisi gösterimi

	Tahmin Edilen Sınıf Etiketleri	
	0	1
Gerçek Sınıf Etiketleri	0 DP	1 YN
	1 YP	DN

Çizelge 2 ile verilen karmaşıklık matrisindeki bu değerler kullanılarak sınıflandırma algoritmasının başarımını analiz etmek için Çizelge 3 ile verilen performans ölçütleri hesaplanır.

sisteminde, Python'un 3.8.10 sürümü, scikit-learn kütüphanesinin 0.24.2 sürümü kullanılarak gerçekleştirilmiştir. Tüm deneyler, 8 GB RAM'e sahip (Intel (R) Core(TM) i7-4510U CPU @ 2.00GHz 2.60 GHz) bir bilgisayarda gerçekleştirilmiştir.

Çalışmada ele alınan problem sınıflandırma problemi olarak ele alınıp sınıflandırma aşamasında RO, AGA ve YSA algoritmasından yararlanılmıştır. Sınıflandırma algoritmalarının başarımında hiper-parametrelerin seçimi çok önemlidir. Bu amaçla, rastgele arama algoritması yöntemi ile hiper-parametreler en iyilenmiştir (Bergstra and Bengio, 2012). Algoritmalar için rastgele arama algoritması ile en iyilenen hiper-parametreler Çizelge 4 ile detaylandırılmaktadır. Sınıflandırma algoritmalarının analizi amacıyla oluşturulan karmaşıklık matrisi, Çizelge 5 ile verilmektedir.

Çizelge 4. Sınıflandırma algoritmaları için seçilen hiper-parametreler

Sınıflandırıcı	Hiper-parametre	Açıklama	Değer
RO	n_estimator	Ağaç sayısı	10
	learning_rate	Öğrenme oranı	0.16
AGA	min_child_weight	Çocuk düğümdeki minimum örnek ağırlığı	10
	max_depth	Ağacın maksimum derinliği	10
	n_estimators	Ağaç sayısı	100
YSA	num_layers	Katman sayısı	2
	units	Katmandaki nöron sayısı	16

epochs	Eğitim tur sayısı	10 ⁴
drop_rate	Seyreltme oranı	0.25
activation	Aktivasyon fonksiyonu	RELU
optimizer	Optimizasyon algoritması	ADAM

Çizelge 5. Sınıflandırma algoritmalarının karmaşıklık matris sonuçları

		Tahmin edilen sınıf etiketi					
		RO		AGA		YSA	
		0	1	0	1	0	1
Gerçek sınıf etiketi	0	47	4	44	7	49	2
	1	10	14	6	18	8	16

Çizelge 5 ile verilen karmaşıklık matrislerinde 0 kalp yetersizliğinin olmadığı ve 1 ise kalp yetersizliğinin olduğu vakaları ifade etmektedir. Çizelge 5'deki sonuçlar incelendiğinde RO ve YSA algoritmalarının kalp yetersizliği olmayan vakaları tespit etmede daha başarılı olduğu AGA algoritmasının ise kalp

yetersizliği olan vakaları tespit etmede daha başarılı olduğu görülmektedir. Çizelge 5 ile verilen karmaşıklık matrisleri kullanılarak hesaplanan performans ölçütleri ve hesaplama süreleri Çizelge 6 ile verilmektedir.

Çizelge 6. Sınıflandırma algoritmalarının sonuçları

	ACC (Eğitim)	ACC (Test)	PREC	RC	FS	AUC	MEZ* (s)	MTS* (s)
RO	%99.10	%81.33	%80.12	%75.25	%66.67	%88.39	0.032	0.002
AGA	%100.00	%82.67	%82.88	%80.64	%80.30	%90.60	0	0.005
YSA	%87.05	%86.67	%86.90	%86.67	%86.08	%89.87	0	0.112

*MEZ: Model Eğitim Zamanı, MTZ: Model Test Zamanı

Çizelge 6 ile verilen sonuçlar incelendiğinde YSA algoritması ile elde edilen performans ölçütü değerlerinin daha iyi olduğu görülmektedir.

Çalışmada önerilen algoritmalar için model eğitim ve test aşamalarında harcanan sürelerinin de birbirlerine oldukça yakın olduğunu söyleyebiliriz.

Çizelge 7. Aynı veri kümesini kullanan çalışmaların sonuçlarının karşılaştırılması

Yazarlar	Yöntem	Doğru sınıflandırma oranı
(Chicco and Jurman, 2020)	Lojistik regresyon	%83.80
(Erdas ve Olcer, 2020)	One Rule	%86.00
Önerdiğimiz yaklaşım	YSA	%86.67

Literatürde aynı veri kümesini kullanarak yapılmış olan çalışmalar incelendiğinde, önerilen yöntem ile elde edilen doğruluk değerinin, diğer çalışmalardan daha iyi olduğu görülmektedir (Çizelge 7).

4. Sonuç ve Tartışma

Bu çalışmada, kalp yetmezliğine bağlı olarak hastaların hayatta kalma veya ölme durumları tahmin edilmektedir. Hastaların tıbbi kayıtlarını içeren gerçek bir veri seti üzerinde RO, AGA ve YSA algoritmalarının performansları analiz edilmektedir. Doğruluk, hassasiyet, hatırlama, f-ölçütü, AİE eğrisi altında kalan alan ve işlem süreleri açısından

sınıflandırıcılar analiz edilmiş ve çalışmada ele alınan problemin çözümünde en uygun modelin RO modeli olduğu tespit edilmiştir. Çalışmada kullanılan algoritmaların karmaşık matrisleri incelendiğinde RO ve YSA algoritmalarının kalp yetersizliği olmayan vakaları tespit etmede daha başarılı olduğu AGA algoritmasının ise kalp yetersizliği olan vakaları tespit etmede daha başarılı olduğu görülmektedir.

Gelecekte yapılacak çalışmalarda, farklı sınıflandırıcıların da performansı bu doğrultuda incelenerek farklı modellerin bir topluluk öğrenme modeli ile bir araya getirilmesi ve daha yüksek

performans değerlerinin elde edilmesi üzerine çalışılacaktır.

5. Kaynaklar

- Aktaş Potur, E., ve Erginel, N., 2021. Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi. *European Journal of Science and Technology*, **24**, 112–118.
- Angraal, S., Mortazavi, B. J., Gupta, A., Khera, R., Ahmad, T., Desai, N. R., Jacoby, D. L., Masoudi, F. A., Spertus, J. A., and Krumholz, H. M., 2020. Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction. *JACC: Heart Failure*, **8(1)**, 12–21.
- Awan, S. E., Bennamoun, M., Sohel, F., Sanfilippo, F. M., Chow, B. J., and Dwivedi, G., 2019. Feature selection and transformation by machine learning reduce variable numbers and improve prediction for heart failure readmission or death. *PLOS ONE*, **14(6)**, e0218760.
- Aydın, A., 2021. Kalp Yetmezliği Hastalarında Kritik Parametre Seçimi ve Sağkalım Modeli Geliştirilmesi. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi*, 155–162.
- Bergstra, J., and Bengio, Y., 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, **13(2)**.
- Breiman, L., 2001. Random Forest. *Machine Learning*, **45(1)**, 5–32.
- Buchan, T. A., Ross, H. J., McDonald, M., Billia, F., Delgado, D., Duero Posada, J. G., Luk, A., Guyatt, G. H., and Alba, A. C., 2019. Physician Prediction versus Model Predicted Prognosis in Ambulatory Patients with Heart Failure. *The Journal of Heart and Lung Transplantation*, **38(4)**, S381.
- Chen, T., and Guestrin, C., 2016. XGBoost. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Chicco, D., and Jurman, G., 2020. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, **20(1)**, 16.
- Degertekin, M., Erol, C., Ergene, O., Tokgozoglu, L., Aksoy, M., Erol, M. K., Eren, M., Sahin, M., Eroglu, E., Mutlu, B., ve Kozan, O., 2012. Heart Failure Prevalence and Predictors in Turkey (HAPPY) Çalışması. *Türk Kardiyoloji Dernegi Arsivi-Archives of the Turkish Society of Cardiology*, **40(4)**, 298–308.
- Erdas, C. B., and Olcer, D., 2020. A Machine Learning-Based Approach to Detect Survival of Heart Failure Patients. *2020 Medical Technologies Congress (TIPTEKNO)*, 1–4.
- Gu, J., Pan, J., Lin, H., Zhang, J., and Wang, C., 2021. Characteristics, prognosis and treatment response in distinct phenogroups of heart failure with preserved ejection fraction. *International Journal of Cardiology*, **323**, 148–154.
- Han, J., and Kamber, M., 2001. *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Harrington, P., 2012. *Machine Learning in Action*. Manning Publications.
- Haykin, S. S., 1999. Neural Networks: A comprehensive Foundation. In *Prentice-Hall, Inc*, **7458**. Prentice Hall.
- Hedman, Å. K., Hage, C., Sharma, A., Brosnan, M. J., Buckbinder, L., Gan, L.-M., Shah, S. J., Linde, C. M., Donal, E., Daubert, J.-C., Målarstig, A., Ziemek, D., and Lund, L., 2020. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart*, **106(5)**, 342–349.
- Meng, F., Zhang, Z., Hou, X., Qian, Z., Wang, Y., Chen, Y., Wang, Y., Zhou, Y., Chen, Z., Zhang, X., Yang, J., Zhang, J., Guo, J., Li, K., Chen, L., Zhuang, R., Jiang, H., Zhou, W., Tang, S., ... Zou, J., 2019. Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retrospective multicentre registry in China. *BMJ Open*, **9(5)**, e023724.
- Mitchell, T., 1997. *Machine Learning*. McGraw Hill.
- Moreno-Sanchez, P. A., 2020. Development of an Explainable Prediction Model of Heart Failure Survival by Using Ensemble Trees. *2020 IEEE International Conference on Big Data (Big Data)*, 4902–4910.
- Segar, M. W., Patel, K. V., Ayers, C., Basit, M., Tang, W. H.

W., Willett, D., Berry, J., Grodin, J. L., and Pandey, A., 2020. Phenomapping of patients with heart failure with preserved ejection fraction using machine learning-based unsupervised cluster analysis. *European Journal of Heart Failure*, **22(1)**, 148–158.

Wilstup, C., and Cave, C., 2021. Combining symbolic regression with the Cox proportional hazards model improves prediction of heart failure deaths. *MedRxiv*, 2021.01.15.21249874.

İnternet Kaynakları

1-<https://www.who.int/Health-Topics/Cardiovascular-Diseases/>, (07/06/2021).