

Makine Öğrenmesi Algoritmaları ile Aritmilerin Sınıflandırılması

Ahmet Turan ÖZDEMİR 

Erciyes Üniversitesi Mühendislik Fakültesi Elektrik Elektronik Mühendisliği, KAYSERİ

(Alınış / Received: 11.08.2023, Kabul / Accepted: 26.09.2023, Online Yayınlanma / Published Online: 30.12.2023)

Anahtar Kelimeler

Makine Öğrenmesi,
Bagging Decision Tree,
Random Forest,
Extra Tree,
Gradient Boosting,
Support Vector Machine

Öz: Bu çalışmada elektrokardiyografi (EKG) kayıtlarından aritmi türlerini belirlemek amacıyla modern makine öğrenme algoritmalarının kullanılması incelenmiştir. Amacımız, EKG sinyallerinin daha etkin bir şekilde analiz edilmesiyle aritmi sınıflandırmasında daha üstün sonuçların elde edilmesidir. Çalışmada kullanılan EKG verileri, MIT PhysioNet veri tabanındaki 203, 208, 210 ve 213 numaralı hasta kayıtlarından seçilmiştir. Bu kayıtlar, V, F ve N aritmi türleri ile etiketlenmiştir. Bagging Decision Tree (BDT), Random Forest (RF), Extra Tree (ET), Gradient Boosting (GB) ve Support Vector Machine (SVM) makine öğrenmesi algoritmaları ile yapılan sınıflandırma işlemleri neticesinde en yüksek doğruluk %98,14 doğruluk ile ET algoritması kullanılarak başarılmıştır. Bununla birlikte ET sınıflandırıcısı 2,17 saniyede eğitilmiş ve 0,0269 saniyede cevap üretebilmiştir. Elde edilen sonuçlar, yeni makine öğrenme algoritmalarının EKG aritmi sınıflandırmasında klasik Yapay Sinir Ağları tabanlı sınıflandırıcılara göre daha etkili sonuçlar üretebildiğini göstermektedir. Bu algoritmalar, geleneksel yöntemlere kıyasla daha yüksek doğruluk, hassasiyet ve özgünlük değerleri sunmaktadır.

Arrhythmia Classification using Machine Learning Algorithms

Keywords

Machine Learnings,
Bagging Decision Tree,
Random Forest,
Extra Tree,
Gradient Boosting,
Support Vector Machine

Abstract: In this study, the use of modern machine learning algorithms in order to determine the types of arrhythmias from electrocardiography (ECG) recordings was investigated. Our aim is to obtain superior results in arrhythmia classification by analyzing ECG signals more effectively. ECG data used in the study were selected from patient records 203, 208, 210 and 213 in the MIT PhysioNet database. These recordings are labeled with arrhythmia types V (Ventricular), F (Fusion), and N (Normal). As a result of classification processes made with Bagging Decision Tree (BDT), Random Forest (RF), Extra Tree (ET), Gradient Boosting (GB) and Support Vector Machine (SVM) machine learning algorithms, the ET algorithm has the highest accuracy with 98.14% accuracy. ET classifier was trained in 2.17 seconds and was able to produce a response in 0.0269 seconds. The results show that new machine learning algorithms can produce more effective results in ECG arrhythmia classification than classical Artificial Neural Network-based classifiers. These algorithms offer higher accuracy, sensitivity and specificity values compared to traditional methods.

*İlgili Yazar, email: aturan@erciyes.edu.tr

1. Giriş

Gelişmiş ve gelişmekte olan ülkelerde, yetişkin nüfusta çeşitli organlarda işlev bozukluklarına yol açan temel etken olan kalp-damar hastalıkları, en önde gelen ölüm nedenlerinden biri olarak bilinmektedir. Dünya Sağlık Örgütü'nün raporlarına göre her yıl, kalp hastalıklarına bağlı nedenlerle hayatını kaybeden insan sayısı 17,1 milyonu bulmaktadır. Bu ölümlerin büyük bir kısmı, kalpte meydana gelen elektriksel anormallikler olarak adlandırılan aritmilere bağlıdır. Erken teşhis edilen tehlikeli kalp rahatsızlıkları, kalp hastalıklarının etkili bir şekilde tedavi edilmesi ve ani kalp ölümlerinin önlenmesi açısından büyük bir öneme sahiptir. Bu nedenle ElektroKardiyogram (EKG) sinyallerini değerlendirerek aritmileri tespit eden otomatik sınıflandırıcılar, erken teşhis sürecinde kritik bir rol oynamaktadır.

Aritmi, düzensiz, çok yavaş veya çok hızlı kalp atışları ile karakterize edilen bir durumdur. Bu durum, inme veya kalp krizi riskini artırabilen önemli bir sağlık sorunudur [1]. Elektrokardiyogram (EKG) sinyalleri, aritmi teşhisi için yaygın bir şekilde kullanılır çünkü kalp ritmindeki anormallikleri tespit etmedeki etkinlikleri kanıtlanmıştır [2].

Geçtiğimiz yirmi yıl boyunca, EKG işaretleri içerisindeki aritmileri tespit etmeye yönelik birçok çalışma gerçekleştirilmiştir. Özellikle yüksek tahmin yetenekleri ile öne çıkan Yapay Sinir Ağları (YSA), tıbbi tanı sistemlerinde en sık kullanılan yöntem olmuştur. YSA biyolojik sinir ağlarının yapısından esinlenilerek geliştirilmiş ve paralel mimarileri sayesinde hesaplamalara büyük hız kazandırmıştır [3,4].

Derin öğrenme teknikleri de EKG işaretlerinde yer alan aritmilerin sınıflandırılmasında etkili bir şekilde kullanılmıştır [5,6]. Bu teknikler EKG sinyallerinden otomatik olarak öznelik çıkarma yetenekleri sayesinde popülerlik kazanmıştır [7]. Derin öğrenme modelleri normal kalp atışları (N), prematüre ventriküler kontraksiyonlar (PVC), atriyal fibrilasyon ve ventriküler taşikardi gibi farklı aritmi tiplerini ayırt etmek için kullanılmıştır [8]. Farklı aritmi türlerini sınıflandırmak için çeşitli derin öğrenme yaklaşımları önerilmiştir. Örneğin araştırmacılar farklı aritmi türlerine göre EKG sinyallerini sınıflandırmak için evrimsel sinir ağları (CNN) kullanmışlardır [8]. CNN'ler, geleneksel sınıflandırma yöntemleriyle birleştirilerek sınıflandırma doğruluğunu artırmak için başarılı bir şekilde kullanılmıştır [6]. Ayrıca, üretken rakip ağlar (GAN), veri artırma amacıyla aritmi sınıflandırmasında başarılı bir şekilde kullanılmıştır [6].

Aritmi sınıflandırmasının doğruluğunu artırmak için öznelik çıkarma, model birleştirme ve dikkat mekanizmaları gibi farklı teknikler kullanılmıştır [6]. Öznelik çıkarma yöntemleri, ilgili öznelikleri tanımlayarak sınıflandırma modellerinin doğruluğunu artırma amacıyla kullanılmıştır [4]. Model birleştirme teknikleri, farklı modellerin güçlerini birleştirerek genel sınıflandırma performansını artırmak için kullanılmıştır. Dikkat mekanizmaları ise global özelliklerin ilişkisini artırmak için kullanılmıştır [6].

Derin öğrenme tekniklerinin aritmi sınıflandırmasında iyi sonuçlar verdiği bilinmekle birlikte destek vektör makineleri (SVM) ve rastgele ormanlar (RF) gibi diğer makine öğrenme algoritmalarının da aynı amaçlar için sıklıkla kullanıldığı unutulmamalıdır. Bahse konu algoritmalar, öznelik seçimi stratejileri ve EKG verilerinin ön işleme tabi tutulması gibi uygulamaların tümü aritmi sınıflandırmada kullanılmıştır [9]. Sonuç olarak, derin öğrenme teknikleri özellikle EKG sinyallerindeki aritmilerin sınıflandırılmasında etkili olmuştur.

Bu çalışmada, MIT-BIH aritmi veri kümesi üzerinde gelişmiş makine öğrenme algoritmalarının, aritmi tiplerini sınıflandırmadaki etkinliği incelenmiştir. Bu araştırmalar, tıbbi teşhis süreçlerine yardımcı olmak için geliştirilen makine öğrenme tabanlı yöntemlerin başarımları hakkında bilgi vermekte ve pratik uygulamalarda kullanım ihtimalleri değerlendirilmektedir. Bu çalışma gelecekteki araştırmalar için de rehberlik sağlayarak aritmi sınıflandırma başarımlarını artırmak ve klinik uygulamalarda bu tür sınıflandırıcıları daha etkili bir şekilde kullanmak için zemin oluşturmaktadır.

2. Materyal ve Metot

Bu bölümün kapsamında, veri formasyonu ve ön işleme adımları ile birlikte gerçekleştirilen makine öğrenme algoritmalarının türleri ayrıntılı bir şekilde açıklanacaktır. Ayrıca geliştirilen sınıflandırma modellerinin performanslarının değerlendirmesi için kullanılan ölçütler ve metrikler de detaylı bir şekilde tartışılacaktır. Böylece araştırmanın genel yapısı ile birlikte içeriği hakkında daha kapsamlı bir anlayışa sahip olunabilecektir.

2.1. Veri Formasyonu

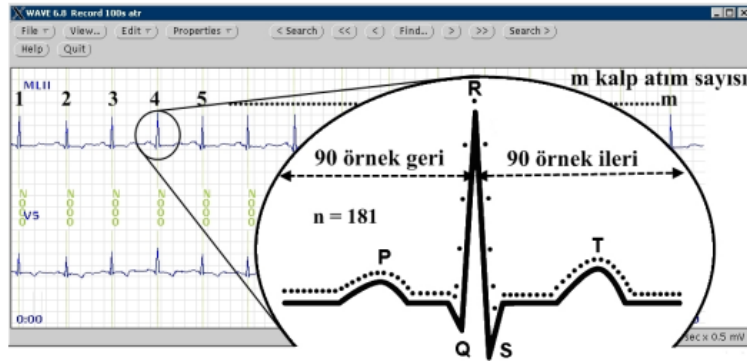
Bu araştırmada, MIT-BIH EKG veri tabanından elde edilen ve ham veri olarak adlandırılan 205, 208, 210 ve 213 numaralı hastaların kayıtları kullanılmıştır. Bu hastalara ait EKG kayıtlarının seçilmesinin amacı, normal (N), PVC (V) ve füzyon (F) tipi atımların üç sınıftan zengin bir şekilde bulunmasını sağlamaktır. Ham veri seti, 9.217 normal, 1.477 PVC ve 755 füzyon atımından oluşmaktadır (Tablo 1).

Tablo 1. Çalışmada kullanılan ham dosyaların aritmi içerikleri.

AAMI Sınıflandırması	N	V	F
MIT-BIH Aritmi Tipleri	N	PVC	F
Ham Veri Seti*	9.217	1.477	755
205	2.571	71	11
208	1.585	992	372
210	2.421	194	10
213	2.640	220	362

*205, 208, 210 ve 213 numaralı hasta kayıtlardan oluşan ham veri setidir.

205, 208, 210 ve 213 numaralı kayıtlardan her bir kalp atımı, R dalgaları etrafında 90 örnek öncesi ve 90 örnek sonrasında olmak üzere toplam 181 örnekleme noktası kullanılarak örneklendirilmiştir. m , kayıttaki kalp atım sayısını, n ise örnek sayısını temsil etmektedir. Bir kaydın süresi 30 dakika olduğu için m değeri sabit değildir, çünkü kalp atım sayısı kişiden kişiye değişebilmektedir. Bu çalışmada tüm kalp atımları 181 örnekleme noktasıyla temsil edilmekte olup, n sabit ve değeri 181'dir (Şekil 1).



Şekil 1. EKG işaret işleme adımları ve izlenen yollar [3].

2.2. Veri Ön İşleme

Veri seti üzerinde standardizasyon işlemi, veri değerlerinin farklı özneliklerin birbirleriyle karşılaştırılabilir ve işlenebilir hale getirilmesi amacıyla kullanılan bir ön işleme adımdır. Bu adım, veri değerlerini ortalaması sıfır ve standart sapması bir olarak dönüştürerek veri setinin dağılımının standart hale gelmesini sağlar.

Standardizasyon, veri setindeki her bir özellik için aşağıdaki formül kullanılarak uygulanır:

$$Z = \frac{X_i - \bar{X}}{\sigma_x} \quad (1)$$

Burada:

- Z , standartlaştırılmış değerdir.
- X_i , i . orijinal veri değeridir.
- \bar{X} , özelliğin ortalamasıdır.
- σ_x , özelliğin standart sapmasıdır.

Bu formül, her bir veri değerini, ortalamadan çıkararak verinin merkezlenmesini sağlar ve ardından standart sapmaya bölerek verinin ölçeğini dönüştürür. Sonuç olarak, veri setindeki her bir değişkenin standart bir dağılıma sahip olduğu ve aynı ölçekte olduğu veride standardizasyon elde edilir.

Standardizasyon, veri setinde yer alan büyük ölçekli değerlerin veya aykırı değerlerin sınıflandırma algoritması performansını olumsuz olarak etkilemesini önler. Aynı zamanda, çeşitli özelliklerin farklı ölçeğe sahip olması durumunda, bu farklılığın modele yanlılık veya büyük hatalar getirmesini engeller. Bu nedenle standardizasyon, veri seti üzerinde olası sınıflandırma veya tahmin hatalarını azaltır ve model performansını iyileştirir.

Bu çalışmada, MIT-BIH aritmi veri kümesi üzerinde standardizasyon yöntemi kullanılarak özellik değerlerinin standartlaştırıldığı bir ön işleme adımı gerçekleştirilmiştir. Bu adım, aritmi sınıflandırma algoritmalarının veri setine daha iyi uygulanabilmesini sağlayarak daha sağlıklı sonuçlar elde edilmesine katkıda bulunmuştur.

2.3. Makine Öğrenme Algoritmaları

Bu çalışmada, MIT-BIH aritmi veri kümesinde yer alan çeşitli aritmi tiplerini doğru bir şekilde sınıflandırmak amacıyla Bagging Decision Tree Classifier (BDT), Random Forest Classifier (RF), Extra Tree Classifier (ET), Gradient Boosting Classifier (GB) ve Support Vector Machine (SVM) gibi beş farklı güçlü makine öğrenmesi algoritması kullanılmış ve performansları ayrıntılı bir şekilde tartışılmıştır.

2.3.1. Bagging Decision Tree Classifier (BDT)

Bagging Decision Tree Classifier (BDT), bagging ve karar ağaçlarının prensiplerini birleştiren bir ensemble öğrenme algoritmasıdır. Bagging, eğitim verisinin farklı alt kümeleri üzerinde birden fazla model eğitme ve tahminlerini çoğunluk oylamasıyla birleştirme yöntemlerini içeren bir tekniktir. Karar ağaçları, girdi verisinin özelliklerine dayalı olarak bir dizi ikili kararın izlenmesi yoluyla tahminler yapan ağaç tabanlı modellerdir.

BDT'de, karar ağaçları eğitim verisinin farklı bootstrap örnekleri üzerinde eğitilir. Her bir ağaç bağımsız olarak eğitilir ve tüm ağaçların tahminleri çoğunluk oylamasıyla birleştirilerek final tahmini yapılır. BDT'deki rastgelelik, bootstrap örnekleri için veri örnekleme ve karar ağacının her bir düğümünde özelliklerin rastgele seçiminden gelir [10]. Bu rastgelelik, aşırı uyumlamayı azaltmaya ve modelin genelleme performansını artırmaya yardımcı olur.

2.3.2. Random Forest Classifier (RF)

Random Forest Classifier (RF), BDT'ye yakın ilişkili bir diğer ensemble öğrenme algoritmasıdır. RF algoritması da karar ağaçlarını temel sınıflandırıcı olarak kullanır, ancak ek bir seviyede rastgelelik kullanır. RF'de her bir karar ağacı rastgele bir özellik alt kümesi üzerinde eğitilir, bu da ağaçlar arasında daha fazla çeşitlilik sağlar. Final tahmini, BDT'ye benzer şekilde tüm ağaçların tahminlerinin çoğunluk oylamasıyla birleştirilmesiyle yapılır [10]. RF'in yüksek boyutlu verilerle başa çıkmada etkili olduğu ve sınıflandırma görevlerinde yüksek doğruluk elde ettiği gösterilmiştir [10].

2.3.3. Extra Tree Classifier (ET)

Extra Tree Classifier (ET), karar ağacı algoritmasının başka bir türevidir ve ek bir rastgelelik getirir. ET'de her bir özelliğin optimal ayırım noktasını bulmak yerine rastgele bir ayırım noktası seçilir. Bu rastgelelik, ağaçlar arasındaki çeşitliliği artırarak aşırı uyumlamayı azaltmaya yardımcı olur. ET'nin gürültülü ve yüksek boyutlu verilerle başa çıkmada etkili olduğu gösterilmiştir [11].

2.3.4. Gradient Boosting Classifier (GB)

Gradient Boosting Classifier (GB), zayıf öğrencilerin genellikle karar ağaçlarının birleştirilmesiyle güçlü bir öğrenci oluşturmasını sağlayan bir boosting algoritmasıdır. BDT ve RF'dan farklı olarak, boosting temel sınıflandırıcıları ardışık olarak eğitir ve her bir sonraki sınıflandırıcı önceki sınıflandırıcılar tarafından yanlış sınıflandırılan örnekler üzerinde odaklanır. Bu tekrarlayan süreç, modelin genel performansını iyileştirmeye yardımcı olur. GB'nin dengesiz verilerle başa çıkmada etkili olduğu ve yüksek doğruluk verdiği bilinmektedir [10].

2.3.5. Support Vector Machine (SVM)

Destek Vektör Makinesi (SVM), sınıflandırma uygulamalarında yaygın olarak kullanılan bir denetimli öğrenme algoritmasıdır. SVM farklı sınıflara ait veri noktalarını maksimum marj ile ayıran optimal hiperdüzlemi bulur. SVM farklı çekirdek işlevlerini kullanarak hem doğrusal olarak ayrılabilir hem de doğrusal olmayan olarak ayrılabilen verilerle başa çıkabilir. SVM'in yüksek boyutlu verilerle başa çıkmada etkili olduğu ve sınıflandırma görevlerinde yüksek doğruluk elde ettiği gösterilmiştir [12].

Özetlemek gerekirse BDT, RF, ET, GB ve SVM sınıflandırma görevleri için kullanılan güçlü makine öğrenme algoritmalarıdır. BDT ve RF karar ağaçlarını temel sınıflandırıcı olarak kullanırken, BDT çoğunluk oylaması, RF ise rastgele özellik seçimi kullanır. ET ek bir rastgelelik getiren bir karar ağacı algoritması varyantıdır. GB zayıf öğrencileri birleştirerek güçlü bir öğrenci oluşturan bir boosting algoritmasıdır. SVM ise veri noktalarını ayıran optimal hiperdüzlemi bulan bir denetimli öğrenme algoritmasıdır. Bu algoritmalar geniş bir alanda incelenmiş ve çeşitli araştırmalarla etkinliği gösterilmiştir [10–12].

2.4. Performans Metrikleri

Çalışma kapsamında model eğitimini optimize etmek ve genelleme yeteneklerini değerlendirmek amacıyla k-katlamalı çapraz doğrulama tekniği kullanılmıştır. Bu yöntem, veri setini k alt küme halinde böler, ardından k kez model eğitimi ve testi gerçekleştirir. Her iterasyonda veri alt kümelerinden biri test verisi olarak ayrılırken diğer k-1 alt küme eğitim verisi olarak kullanılır. Bu yaklaşım sayesinde farklı veri alt kümeleri üzerinde model performansı değerlendirilir ve elde edilen sonuçların ortalaması alınarak daha güvenilir bir değerlendirme elde edilir [13-18]. Bu özel çalışmada k değeri 10 olarak belirlenmiştir. k-katlamalı çapraz doğrulama sonrasında elde edilen modeller, doğruluk oranı ve karmaşıklık matrisi gibi metrikler kullanılarak ayrıntılı bir şekilde değerlendirilmiştir.

Doğruluk (Accuracy): Sınıflandırma modellerinin doğru tahmin ettiği örneklerin, toplam tahmin sayısına oranını ifade eden bir ölçüdür. Basitçe söylemek gerekirse, doğru şekilde sınıflandırılan örneklerin toplam örnek sayısına oranını ifade eder. Ancak, dengesiz sınıf dağılımları durumunda doğruluk tek başına yeterli bir performans ölçütü olmayabilir.

$$Accuracy(Acc) = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \quad (2)$$

Karmaşıklık Matrisi (Confusion Matrix): Bu matris sınıflandırma modellerinin performansını değerlendirmek için kullanılan bir tablodur. Her satır gerçek sınıfları temsil ederken, her sütun tahmin edilen sınıfları temsil eder. Dört temel terimi içerir: Doğru Pozitifler (True Positives - T_p), Yanlış Pozitifler (False Positives - F_p), Doğru Negatifler (True Negatives - T_n) ve Yanlış Negatifler (False Negatives - F_n). Bu metrikler, modelinizin sınıfları ne kadar doğru veya yanlış tahmin ettiğini gösterir.

3. Bulgular

Bu çalışmada, elektrokardiyografi (EKG) kayıtlarındaki aritmi tiplerinin sınıflandırılması amacıyla çeşitli makine öğrenme algoritmalarının performansı değerlendirilmiştir. Çalışmada incelenen sınıflar; normal atım (N), erken ventriküler kasılma (V), ve füzyon tipi atım (F) olarak belirlenmiştir. Farklı algoritmaların sınıflandırma yetenekleri doğruluk değerleri, kararlılık, eğitim ve test süreleri, ayrıca karmaşıklık matrisi analizleri temelinde değerlendirilmiştir. Bu bölümde, her bir algoritmanın performansının ayrıntılı bir şekilde incelenmesiyle elde edilen sonuçlar sunulmaktadır. Sonuçlar, aritmi sınıflandırma alanında hangi algoritmanın en iyi performansı sergilediği konusunda değerli bilgiler sunmaktadır.

Makine öğrenme algoritmalarının performansını değerlendirmek amacıyla, analiz sonuçlarını içeren Tablo 2'ye göz atalım. Bu tablo, kullanılan algoritmaların performansını farklı ölçütlere göre özetlemektedir. Algoritmalar arasındaki karşılaştırmalar yapılarak, hangi algoritmanın daha iyi bir sınıflandırma yeteneği gösterdiği ve sınıf tahminlerinin dağılımını anlamamız kolaylaşır.

Tablo 2. Makine öğrenme algoritmalarının başarımlarını ve çalışma süreleri

Algoritma	Doğruluk	Standart Sapma	Eğitim Süresi (s)	Test Süresi (s)
BDT	0,9796	0,0038	9,22	0,0169
ET	0,9814	0,0047	2,17	0,0269
RF	0,9804	0,0035	7,24	0,0271
GB	0,9785	0,0038	95,58	0,0062
SVM	0,9783	0,0044	2,51	0,2653

BDT algoritması yüksek bir doğruluk değeri (0,9796) sergilemektedir. Standart sapmanın düşük olması (0,0038), algoritmanın sonuçlarının kararlı olduğunu göstermektedir. Eğitim süresi (9,22 saniye) biraz uzun olabilir, ancak test süresi oldukça hızlıdır (0,0169 saniye). Karışıklık matrisine bakıldığında (Tablo 3),

Tablo 3. BDT karışıklık matrisi.

	Tahmin N	Tahmin V	Tahmin F
Gerçek N	9174	11	32
Gerçek V	18	1423	36
Gerçek F	110	45	600

- N: Doğru tahmin oranı %99,53 ($9217/9217 = 0,9953$). Bu oldukça yüksek bir başarı oranıdır.
- V: Doğru tahmin oranı %96,34 ($1423/1477 = 0,9634$). Bu da genel olarak iyi bir sonuçtur.
- F: Doğru tahmin oranı %79,47 ($600/755 = 0,7947$). Bu sınıf için diğerlerine göre daha düşük bir tahmin oranı görünmektedir.

ET algoritması en yüksek doğruluk değerlerinden birine sahiptir (0,9814) ve düşük standart sapması (0,0047) sonuçların istikrarlı olduğunu gösterir. Eğitim süresi (2,17 saniye) düşük ve test süresi (0,0269 saniye) kabul edilebilir düzeydedir. Karışıklık matrisi incelendiğinde (Tablo 4),

Tablo 4. ET karışıklık matrisi.

	Tahmin N	Tahmin V	Tahmin F
Gerçek N	9199	2	16
Gerçek V	18	1430	29
Gerçek F	111	44	600

- N: Doğru tahmin oranı % 99,80 ($9199/9217 = 0,9980$). Bu sonuç oldukça yüksek bir başarı düzeyini temsil eder.
- V: Doğru tahmin oranı % 96,83 ($1430/1477 = 0,9683$). Bu da genel olarak iyi bir sonuçtur.
- F: Doğru tahmin oranı %79,47 ($600/755 = 0,7947$). Bagging Classifier ile benzer bir sonuç elde edilmiştir.

RF algoritması yüksek doğruluk değeri (0,9804) ve düşük standart sapma (0,0035) ile iyi bir performans göstermektedir. Eğitim süresi (7,24 saniye) ortalama düzeydedir, ancak test süresi (0,0271 saniye) biraz daha uzundur. Karışıklık matrisi incelendiğinde (Tablo 5),

Tablo 5. RF karışıklık matrisi.

	Tahmin N	Tahmin V	Tahmin F
Gerçek N	9197	3	17
Gerçek V	17	1430	30
Gerçek F	106	46	603

- N: Doğru tahmin oranı %99,78 ($9197/9217 = 0,9978$). ET Classifier gibi yüksek bir başarı oranıdır.
- V: Doğru tahmin oranı %96,87 ($1430/1477 = 0,9687$). Diğer algoritmalara benzer bir başarı düzeyi var.
- F: Doğru tahmin oranı %79,80 ($603/755 = 0,7980$). Diğer iki algoritma gibi bu sınıfta da benzer bir sonuç.

GB algoritması yüksek doğruluk değeri (0,9785) ve kabul edilebilir düzeydeki standart sapma (0,0038) ile iyi bir performans sergilemektedir. Eğitim süresi (95,58 saniye) diğer algoritmalara göre çok uzun olmakla birlikte, test süresi (0,0062 saniye) oldukça kısadır. Karışıklık matrisi incelendiğinde (Tablo 6),

Tablo 6. GB karışıklık matrisi.

	Tahmin N	Tahmin V	Tahmin F
Gerçek N	9191	6	20
Gerçek V	25	1418	34
Gerçek F	107	45	603

- N: Doğru tahmin oranı %99,72 ($9191/9217 = 0,9972$). Diğer algoritmalara yakın yüksek bir başarı oranı.
- V: Doğru tahmin oranı %96,04 ($1418/1477 = 0,9604$). Diğer algoritmalara göre biraz daha düşük.
- F: Doğru tahmin oranı %79,80 ($603/755 = 0,7980$). Diğer algoritmalara benzer.

SVM algoritması yüksek doğruluk değeri (0,9783) ve kabul edilebilir düzeydeki standart sapma (0,0044) ile iyi bir sınıflandırma başarısı sergilemektedir. Eğitim süresi (2,51 saniye) ortalama bir değere sahipken, test süresi (0,2653 saniye) diğer algoritmalara göre çok daha uzundur. Karışıklık matrisi incelendiğinde (Tablo 7),

Tablo 7. SVM karışıklık matrisi.

	Tahmin N	Tahmin V	Tahmin F
Gerçek N	9204	6	7
Gerçek V	22	1425	30
Gerçek F	130	45	580

- N: Doğru tahmin oranı %99,86 ($9204/9217 = 0,9986$). Diğer algoritmalar gibi yüksek bir başarı oranı.
- V: Doğru tahmin oranı %96,36 ($1425/1477 = 0,9636$). Diğer algoritmalara benzer bir başarı düzeyi.
- F: Doğru tahmin oranı %76,89 ($580/755 = 0,7689$). Diğer algoritmalara göre daha düşük bir sonuç.

Bulgular değerlendirildiğinde, her bir algoritmanın belirli sınıflarda daha iyi veya daha zayıf performans sergilediği görülmektedir. Ancak birçok yönden ET algoritması ön plana çıkmaktadır. Bu algoritmanın yüksek doğruluk, düşük standart sapma ve hızlı işlem süreleri ile iyi bir performans gösterdiğini söyleyebiliriz. Karışıklık matrisi sonuçları da bu öneriyi desteklemektedir.

4. Tartışma ve Sonuç

Elde edilen bulgular ve sonuçları ile birlikte sunulan beş üstün makine öğrenme algoritması, gelecekteki araştırmalara öncülük etme potansiyeline sahiptir. Bu alandaki çalışmaların, aritmi sınıflandırma performansını daha da artırmak ve klinik uygulamalarda etkin bir şekilde kullanmak için daha fazla geliştirme fırsatları sunabileceği akıldta tutulmalıdır.

Yapılan incelemede, MIT-BIH aritmi veri kümesi kullanılarak, aritmi tiplerini sınıflandırmak amacıyla on farklı makine öğrenme algoritması değerlendirilmiştir. Bu çalışmanın sonucunda, elde edilen veriler detaylı bir şekilde analiz edilmiş ve en yüksek doğruluk performansını gösteren beş özel makine öğrenme algoritması belirlenmiştir.

Geliştirilen yöntem ve hassas veri işleme ile farklı aritmi tiplerinin makine öğrenmesi algoritmaları kullanarak kesin bir şekilde sınıflandırılması başarılmıştır. Temel kaynağı MIT-BIH aritmi veri kümesi olan bu çalışmanın kapsamlı deneyler sonucunda elde edilen sonuçları titizlikle değerlendirilmiştir. Başarım, standart sapma, eğitim ve test sürelerine bakıldığında ET algoritmasının %98,14 doğruluk, 0,0047 standart sapma, 2,17 s eğitim ve 0,0269 s cevap süresiyle oldukça başarılı ve pratik uygulamalarda EKG işaretlerinin sınıflandırılması amacıyla kullanım potansiyeline sahip olduğu görülmektedir. Bu çalışma ile modern makine öğrenmesi tekniklerinin klasik yapay zeka tekniklerine tercih edilebileceği görülmüştür.

Kaynakça

- [1] Hammad, M., Meshoul, S., Dziwiński, P., Pławiak, P., Elgendy, I. A. 2022. Efficient Lightweight Multimodel Deep Fusion Based on ECG for Arrhythmia Classification. *Sensors*. 22. 9347.
- [2] Wasimuddin, M., Elleithy, K., Abuzneid, A.-S., Faezipour, M., Abuzagheh, O. 2020. Stages-Based ECG Signal Analysis From Traditional Signal Processing to Machine Learning Approaches: A Survey. *IEEE Access*. 8, 177782-177803.
- [3] Özdemir, A.T. Erken Ventriküler Kasılmalarda YSA Tabanlı Bir Sınıflandırıcının FPGA ile Gerçekleştirilmesi. 2010. Doktora Tezi. Erciyes Üniversitesi.
- [4] Özdemir, A.T., Danışman, K. 2011 Fully parallel ANN-based arrhythmia classifier on a single-chip FPGA: FPAAC. *Turk. J. Electr. Eng. Comput. Sci.*, 19(4). 667-687.
- [5] Ma, S., Cui, J., Xiao, W., Liu, L. 2022. Deep Learning-Based Data Augmentation and Model Fusion for Automatic Arrhythmia Identification and Classification Algorithms. *Comput. Intell. Neurosci.* 2022. 1577778.
- [6] Murat, F., Yildirim, O., Talo, M., Baloglu, U.B., Demir, Y., Acharya, U.R. 2020. Application of Deep Learning Techniques for Heartbeats Detection Using ECG Signals-Analysis and Review. *Comput. Biol. Med.* 120. 103726.
- [7] "Electrocardiogram Signal Classification Based on Deep Learning Techniques", 26 Haziran 2023. <https://www.researchsquare.com> (erişim 09 Temmuz 2023).
- [8] Badrinath, R., Navada, A., Narahari, H., Datta, A., Sarojadevi, H. 2019. Deep Learning based Arrhythmia Classification with an ECG Acquisition System. *Int. J. Innov. Technol. Explor. Eng.* 9(2). 3849-3852.
- [9] Jain, R., Betrabet, P. R., Rao, B. A., Reddy, N.V.S. 2022. Classification of Cardiac Arrhythmia using improved Feature Selection methods and Ensemble Classifiers. *J. Phys. Conf. Ser.* 2161(1). 012003.
- [10] Kulkarni, V. Y., Sinha, P. K. 2012. Pruning of Random Forest classifiers: A survey and future directions. *International Conference on Data Science & Engineering (ICDSE)*, 18-20 Temmuz, Cochin, Hindistan, 64-68.
- [11] Bonsignori, V., Guidotti, R., Monreale, A. 2021. Deriving a Single Interpretable Model by Merging Tree-Based Classifiers. *Discovery Science: 24th International Conference*, 11-13 Ekim, Halifax, Kanada, 347-357.
- [12] Dutta, S., Bandyopadhyay, S.K. 2020. Early Breast Cancer Prediction using Artificial Intelligence Methods. *J. Eng. Res. Rep.* 13(2). 48-54.
- [13] Özdemir, A.T., Barshan B. 2014. Detecting falls with wearable sensors using machine learning techniques. *Sensors* 14 (6), 10691-10708.
- [14] Özdemir, A.T. 2016. An analysis on sensor locations of the human body for wearable fall detection devices: Principles and practice. *Sensors* 16 (8), 1161.
- [15] Yildiz, F., Özdemir, A.T. 2019. Prediction of laser-induced thermal damage with artificial neural networks. *Laser Physics* 29 (7), 075205.
- [16] Kavuncuoğlu, E., Uzunhisarcıklı, E., Barshan, B., Özdemir, A.T. 2022. Investigating the Performance of Wearable Motion Sensors on recognizing falls and daily activities via machine learning. *Digital Signal Processing* 126, 103365.
- [17] Mohammednour, A.B, Özdemir A.T. 2020. GNSS positioning accuracy improvement based on surface meteorological parameters using artificial neural networks. *International Journal of Communication Systems* 33 (9), e4373.

- [18] Kavuncuoğlu, E., Yıldız, F., Özdemir, A.T. 2022. Artificial Intelligence (AI) algorithms for evaluation of optical fiber scintillation detector performance. Optik 258, 168791.