

Makale / Research Paper

Kardiyovasküler Hastalıkların Belirlenmesi için Yapay Zekâ Yöntemleriyle Kural Tabanlı Teşhis Algoritmalarının Geliştirilmesi

Nilay TAÇYILDIZ^{1a}, Buse Nur KARAMAN^{1a}, Zeynep Bağdatlı^{1a}, Sude ÇİĞNİTAŞ^{1a}, Muhammed Kürşad UÇAR^{1a}

^{1a}Sakarya Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği, Sakarya/TÜRKİYE
buse.karaman@ogr.sakarya.edu.tr

Received/Geliş: 24.06.2022

Accepted/Kabul: 07.09.2022

Özet:

Dünya Sağlık Örgütü (DSÖ) verilerine göre kalp hastalıkları en fazla ölüm oranına sahip hastalıklar arasındadır. Kardiyovasküler hastalıklar olarak bilinen kalp ve damar hastalıkları, damarın iç duvarında plak oluşması ile damarların sertleşmesi sonucu daralması ve kanın akışını zorlaştırması olarak tanımlanır. Hastalığın teşhisi, çeşitli klinik bulguların incelenmesiyle konulmaktadır. Klinik bulgu ve testlerin zaman alması teşhis aşamasını uzatmaktadır. Bu nedenle hastalık teşhis sürecini kolaylaştıracak yeni araç ve yöntemler araştırılmaktadır. Erişime açık paylaşım sitesi Kaggle'dan kalp hastalığı veri kümesi kullanılmıştır. Veri setinde 14 adet özellik bulunmaktadır. Özellikler Eta korelasyon katsayısı yöntemiyle 11'e indirilmiştir. Karar ağaçları algoritmaları yardımıyla kural tabanlı teşhis algoritmaları geliştirilmiştir. Model, 5 eğitim adımı sonucunda en iyi sonuçları vermiştir. Çalışma sonucunda ortalama 94,15 doğruluk oranı, 0,98 duyarlılık, 0,91 özgüllük sonuçları elde edilmiştir. Sonuç: Model, performans sonucuna göre KVH teşhisi için yapay zekâ yöntemleriyle geliştirilip, yüksek doğruluk oranına sahiptir. Bu algoritmanın klinikte hekim tarafından kullanılıp, KVH hastalığı erken teşhisinde yardımcı olacağı düşünülmektedir.

Anahtar Kelimeler: Kardiyovasküler hastalık, Eta korelasyon katsayısı, Karar ağaçları

Development of Rule-Based Diagnostic Algorithms with Artificial Intelligence Methods for the Determination of Cardiovascular Diseases

Abstract: According to the World Health Organization (WHO) data, heart diseases are among the diseases with the highest mortality rate. Cardiovascular diseases, known as cardiovascular diseases, are defined as the formation of plaque on the inner wall of the vessel, the hardening of the vessels, the narrowing of the vessel and making the blood flow difficult. The diagnosis of the disease is made by examining various clinical findings. The clinical findings and tests take time, prolonging the diagnostic phase. For this reason, new tools and methods are being researched to facilitate the disease diagnosis process. Materials and Methods: Heart disease dataset from Kaggle, a public sharing site, was used in the study. There are 14 features in the dataset. The features were selected with the Eta correlation coefficient and reduced to 11. Rule-based diagnostic algorithms have been developed with the help of decision tree algorithms. Results: As a result of the study, rule-based algorithms were developed at approximately 5 levels, with an average accuracy rate of 94.15, sensitivity of 0.98, and specificity of 0.91. Conclusion: According to the model performances, it has a high accuracy rate developed with artificial intelligence methods for the diagnosis of CVD, and it is thought that it can be used as a rule-based diagnostic algorithm by the clinician.

Keywords: Cardiovascular disease, Eta correlation coefficient, Decision trees

Bu makaleye atf yapmak için

Taçyıldız, N., Karaman B.N., Bağdatlı Z. Çiğnitaş, S., Uçar, M. K., "Kardiyovasküler Hastalıkların Belirlenmesi için Yapay Zekâ Yöntemleriyle Kural Tabanlı Teşhis Algoritmalarının Geliştirilmesi", El-Cezerî Fen ve Mühendislik Dergisi 2022, 9(4); 1265-1273.

How to cite this article

Taçyıldız, N., Karaman B.N., Bağdatlı Z. Çiğnitaş, S., Uçar, M. K., "Development of Rule-Based Diagnostic Algorithms with Artificial Intelligence Methods for the Determination of Cardiovascular Diseases", El-Cezerî Journal of Science and Engineering, 2022, 9(4); 1265-1273.

1.Giriş

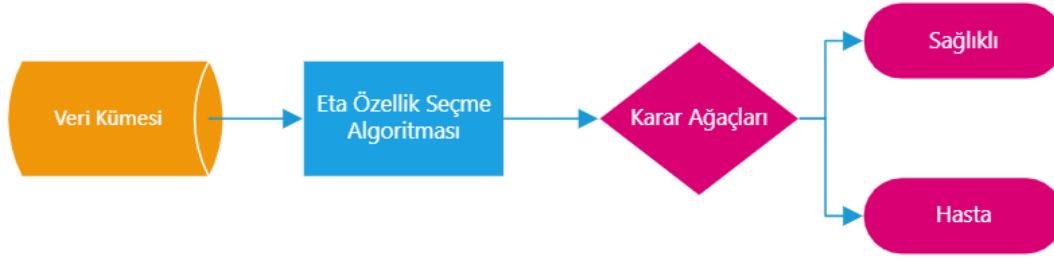
Günümüz dünyasında teknolojide alınan büyük adımlar yapay zekâ alanında da büyük etkilere sebep olmuştur. Yapay zekâ, bir bilgisayarın veya bilgisayar ile kontrolü yapılan makinenin, insanlara ait özellikler olan, probleme çözüm yolu bulma, anlama, bir mana çıkartma, genelleme yapma ve geçmişteki deneyimlerinden öğrenme gibi yüksek mantık gerektiren süreçlere ilişkin fonksiyonları yerine getirme kabiliyeti olarak bilim ve fen camiasında tanımlanmıştır [1]. Yapay zekâyı insana ait becerilerin makineye aktarılması ile ilgilenen bir çalışma alanı olarak tanımlayabiliriz. Yapay zekaya bu becerileri kazandırırken temel olarak kullandığımız yöntem algoritmalarıdır. Algoritma, herhangi bir problemi tanımlamada tanımlama da kullanılan ve sıralı kademelerden meydana gelen mimariye denmektedir. Algoritmalar, programlama tasarımında kullanılan yapılardır ve bir program geliştirmek için gereken adımlar kendi dilimizde anlatılmaktadır [2]. Algoritmalar bir sorunun nasıl çözüleceğini veya belirli bir hedefe nasıl ulaşacağını açıklamanın bir yoludur. Bu denklemden de görebileceğiniz gibi, algoritma sonuç değil, sonuca giden yoldur [2]. Birçok yapay zekâ algoritması mevcuttur fakat hem karmaşık veri setlerinde kullanılabilirlikleri için hem de insanın muhakeme yeteneğine en yakın algoritma olduğundan Karar Ağaçları Sınıflandırma algoritması önemli bir yapay zekâ algoritmasıdır. Karar ağacı sınıflandırması, veri kümelerini sınıflandırmak için etkili ve hızlı bir metottur [3]. Karar ağaçları iki ya da daha fazla veri arasındaki bağlantı ve kümelendirme problemlerinde kullanılan ağaç tabanlı algoritmalarından biridir. Karar ağacı algoritması, bir sürü basit testi mantıksal olarak birleştiren sıralı bir modeldir.; her test, sayısal bir özelliğin bir dönüş eşiği veya nominal bir özelliğin bir dizi tahmini değerlerin karşılaştırılması olarak ilerler. Bu tür sembolik sınıflandırıcılar, anlaşılabilirlik açısından sinir ağları gibi “kara kutu” modellerine göre bir avantaja sahiptir [4]. Burada Yapay Sinir Ağlarında olduğu gibi işlemler bir kara kutu da gerçekleşmez, adımlar kişi tarafından rahatlıkla gözlemlenebilir. Karar ağaçları çok sayıda veri içeren bir olayı bazı teknikler kullanarak alt bölümlere ayıran özellikte bir algoritma olduğu için çok tercih edilmekte, özellikle veri bilimindeki kullanım oranı yadsınamaz şekilde yaygındır. Çünkü Karar Ağaçları, sınıflandırıcının sahip olduğu özellikleri net bir şekilde gösteren, bu özelliklerin olasılıklarına bağlı durumları belirli bir düzen ,sekinde sıralayarak ağaç şeklinde gösteren bir algoritmadır. Karar ağaçları bu alt bölümlere ayırma işlemini gerçekleştirmesine olanak sağlayan Karar Ağacının kendine has temel yapısıdır. Bunlar temel olarak kök, düğüm ve yapraklardır. Kök ve düğüm ve yaprakların yapısı bir şekil ortaya koyduğu için sayısal, anlaşılması zor bir matematik denklemden daha kolay bir şekilde yorumlanır.

Literatüre kardiyovasküler hastalıklar olarak geçmiş, kalp ve kalbin damarlarını ilgilendiren kalp hastalıkları, damarın iç duvarında plak oluşması ile damarların sertleşmesiyle damarın daralması ve kanın akışını zorlaştırıcı bir hastalık olarak tanımlanır. Kalp içerisinden geçen kan ile beslenmektedir. Atar damarlara genel olarak Koroner arter adı verilir. Koroner arterler, aorttan ayrılıp kalbe tekrar giren iki ana koludur. Koroner Arter Hastalığı, arterlerin daralması sonucu kan akışının belirli oranda durması sonucu ortaya çıkan ve literatüre ateroskleroz olarak giren bir hastalık türüdür [5]. Dünya üzerindeki ölüm oranları incelendiğinde, kalp krizi yüzünden her yıl sayısız insanın yaşamı son bulmaktadır. Dünya Sağlık Örgütünden elde edilen verilerine göre dünya nüfusunun her yıl yüzde 31'e yaklaşık kısmı kalp hastalığından dolayı hayatını kaybetmektedir [6]. Bu oran ülkemiz de TÜİK verilerine göre nüfusa oranla yüzde 25,8'i oluşturmaktadır. KVH hastalığının analizinde birçok belirleyici etmen vardır. Bunlara yaş, cinsiyet, ırk ve çevresel faktörler örnek olarak verilebilir. Çevresel faktörler incelendiğinde sigara ve tütün kullanımı, sağlıksız beslenme biçimi ve pasif hareket ön plana çıkmaktadır. Bu alışkanlıkların değiştirilmesi ve sağlıklı bir yaşam tarzını benimsenmesi hastalığa yönelik risk durumlarını azaltır [6]. KVH hastalığının önüne geçilmesi için erken tanı ve tedavi büyük oranda önem arz etmektedir. Bu nedenle hastalığın erken teşhisi için çevresel risk faktörlerinin iyi değerlendirilmesi gerekmektedir. Bu çalışmada

yapay zekâ algoritmalarından biri de olan karar ağaçları kullanılmıştır. KVH hastalığını tespit etme ve erken tanı konulabilmesi için tıp bilimine yardımcı olmak hedeflenmiştir.

2. Yöntem ve Materyal

Model şekil 1’de belirtildiği gibi modelleneyecektir. İlk olarak veriler elde edilecek, bir sonraki adımda ise çalışmada yapay zekanın hastalığı tespit etme başarı oranı arttırması için Eta Özellik Seçme algoritması kullanılacak olup ve en iyi özellikleri seçme işlemi yapılacaktır. Son olarak Kardiyovasküler hastalık tanısı için makine öğrenmesi metotlarından biri olan karar ağaçları kullanılacaktır.



Şekil 1. Çalışma akış diyagramı

2.1. Veri Seti

Dünya’da sık rastlanan ve yüksek ölüm yüzdesine sahip, kalp hastalığını önceden teşhis edebilmek amacıyla internet tabanlı açık erişim sitesi olan Kaggle veri tabanı üzerinden veri grubu elde edilmiştir. Veri kümesi toplamda 499 tane sağlıklı, 526 tane kalp hastası olan toplam 1025 adet veriden oluşmaktadır. Veri kümesinde her bir hastadan farklı veriler alınıp toplamda 1025 hasta için 14 öznitelik kullanılmıştır. Alınan veri setindeki 14 özellik her bir kişiye tıbbi tahliller uygulanarak ve bazı ölçümler yapılarak elde edilmiştir. Veri setimiz yaş, cinsiyet, göğüs ağrı tipi(angina), dinlenme kan basıncı(tansiyon), kandaki serum kolesterol miktarı(kolesterol), açlık kan şekeri(diyabet), dinlenme elektrokardiyografik sonuçları (EKG), ulaşılan maksimum kalp hızı gibi özellikleri içermektedir. Özellikler tablo 1’de verilmiştir.

2.2. Özellik Seçme

İnternet tabanından elde ettiğimiz veri setinde gerçekleştirilen tıbbi çalışmada doğru teşhis yapabilmek için veri setinden elde edilen özelliklerden yararlanılarak veri setini iyi analiz etmek önemlidir. Özellik seçimi bir diğer adıyla değişken seçimi seçilen algoritmaya göre veri kümesindeki x adet özellikli arasından en iyi n adet özellik seçme işlemidir [7]. Özellik seçimi veri setindeki önemli özellikleri seçerek veri miktarını azaltılmayı amaçlar. Özellik sayısının en aza indirgenmesi algoritmamıza birçok kolaylık sağlamaktadır. Öznitelik seçme işlemi seçilen algoritmanın işlem hızını arttırır. En genel anlamıyla özellik seçme işlemi uygulanmak istenen modelin başarısını arttırır. Bu amaçla veri setinde hazır olarak elde etmiş olduğumuz 14 Adet özellikten yapay zekanın başarı oranını arttırmak amacıyla ETA algoritması kullanılarak en iyi 10 özellik seçilmiştir. ETA özellik seçme algoritması korelasyon katsayısı hesaplanması işlemidir [7]. Bir değişkeni ifade etmektedir. Sadece sınıf etiketleri ile sayısal değişkenler arasında kullanılabilir. Özelliğin değeri evet veya hayır şeklinde ise bu algoritma kullanılamamaktadır. Bu yöntemin amacı ise az özellik kullanarak daha yüksek performans elde edilmektedir. Seçilen özellikler sonraki aşamalarda makine öğrenmesi metotlarından biri olan karar ağaçları ile sınıflandırmaya tabi tutulmuştur. En iyiden en kötüye seçilen özelliklerin sırası tablo 2’de sunulmuştur.

Tablo 1. Öznitelikler ve Açıklamaları

Özellik Numarası	Özellik	Değer	Açıklama
1	age	Sayısal Değer	Hastanın Yaşı
2	sex	1: Erkek; 0: Kadın	Hastanın Cinsiyeti
3	cp	1: Tipik Anjina; 2: Atipik Anjina 3: Anjinal Olmayan Ağrı 4: Asemptomatik	Göğüs Ağrısı Tipi
4	trestbps	Sayısal Değer (140mm/Hg)	Dinlenme Kan Basıncı (Tansiyon)(mm/Hg)
5	chol	Sayısal Değer (289mg/dl)	Kandaki Serum Kolesterol Miktarı (mg/dl)
6	fb	1: Doğru, 0: Yanlış	Açlık Kan Şekeri >120 mg/dl
7	restecg	0: Normal, 1: ST-T'ye Sahip 2: Hipertrofi	Dinlenme Elektrokardiyografik Sonuçları
8	thalach	140,173	Ulaşılan Max Kalp Hızı
9	exang	1: Evet, 0: hayır	Egzersize Bağlı Anjina
10	oldpeak	Sayısal Değer	Egzersize Bağlı ST Depresyonu
11	slope	1: Yukarı Eğimli, 2: Düz 3: Aşağı Eğimli	Pik Egzersiz ST Segmentinin Eğimi
12	ca	0-3	Floroskopi ile Renklendirilen Damar Sayısı (0-3)
13	thal	0: Normal, 1: Sabit Kusur 2: Geri Döndürülebilir Kusur	Talasemi
14	target	0: <%50 Çap Daralması 1: > %50 Çap Daralması	Kalp Hastalığı Teşhisi (Anjiyografik Hastalık Durumu)

Tablo 2. Seçilen Özelliklerin Sıralanması ve Korelasyon Katsayısı

Özellik Numarası	Korelasyon İlişik Katsayısı
1	0,2293
2	0,2795
3	0,4349
4	0,1388
5	0,1
6	0,0412
7	0,1345
8	0,4229
9	0,438
10	0,4397
11	0,3455

2.3 Karar Ağaçları

Karar ağaçları birden çok dizin arasındaki bağlantı ve kümelenme problemlerinde kullanılan ağaç tabanlı algoritmaların biridir. Karar ağacı algoritması, bir sürü basit testi mantıksal olarak birleştiren sıralı bir modeldir.; her test, sayısal bir özelliğin bir dönüş eşiği veya nominal bir özelliğin bir dizi tahmini değerlerin karşılaştırılması olarak ilerler [4]. Karar ağaçları çok sayıda veri içeren bir olayı bazı teknikler kullanarak alt bölümlere ayıran özellikte bir algoritma olmasından dolayı sıklıkla kullanılmaktadır özellikle veri bilimindeki kullanım oranı yadsınamaz şekilde yaygındır. Çünkü Karar Ağaçları, sınıflandırıcıya ait öznitelikleri açık bir şekilde gösteren, bu özelliklerin olasılıklarına bağlı durumları belirli bir düzen şeklinde sıralayarak ağaç şeklinde gösteren bir algoritmadır. Karar ağaçları bu alt bölümlere ayırma işlemini gerçekleştirmesine olanak sağlayan Karar ağacının kendine has temel yapısıdır. Bunlar temel olarak kök, düğüm ve yapraklardır. Kök ve düğüm ve yaprakların yapısı bir şekil ortaya koyduğu için sayısal anlaşılması zor bir matematik denkleminde daha kolay bir şekilde yorumlanır. Karar ağacı sınıflandırma algoritmasında her bir özellik bir sınıfa yerleştirilinceye kadar düğüm oluşturma olayı devam eder. Karar Ağaçları tamda isminden de anlaşılacağı üzere bir probleme çözüm uygulamak için topladığımız veri setini alarak belli sınıflar (ağaçlar) oluşturur. Karar ağacı algoritmalarının ilk basamaklarına kök (root veya root node) denir. Her bir veri kökün yapısındaki şarta göre “Evet” veya “Hayır” olarak sınıflandırılır. Kök seçimi oldukça önemlidir çünkü kökün veri setimizi olabildiğince iyi açıklamasını isteriz. Kök hücrelerinin altında bulunan kısma düğümler (interval nodes veya nodes) denir [8]. Düğüm sayısının artması karmaşıklığı artırır. Her bir veri düğümler kullanılarak kategorilere ayrılır yani sınıflandırılır. Karar ağaçları sınıflandırma algoritmasında kök seçimi oldukça önemlidir. Çünkü kökün veri setimizi olabildiğince iyi açıklamasını isteriz. Bu sebepten ötürü kök seçimi büyük önem arz etmektedir. Kök seçiminde kullanılan birçok yöntem mevcuttur. Bunlardan başlıca kullanılan yöntemler şunlardır: Gini index, Entropi, DKM kriterleri, Bilgi kazancı. Bu modelde ise çok sık kullanılan ve anlaşılabilirliği kolay olan Gini ve Entropiyi kök seçiminde yöntem olarak kullandık. Karar ağaçları sınıflandırma algoritması verilen sınıflara ait tüm özellikleri derinlemesine inceldiği için çok büyük veri setine sahip çalışmalarda algoritma şişip ağaç yapısı karmaşıklaşabilir. Bunun sonucunda hem ortaya çıkan algoritmanın uygulanabilirliği konusunda hem de karar ağaçlarında karar verme mekanizması gereği yapı çok karmaşıklaştığı için ezberleme(overfitting) olayının gerçekleşme olasılığı çok yüksek olur [9]. Bunun önüne geçmek için budama işlemi gerçekleştirilir. Bu çalışmanın son aşamasında sınıflandırma algoritmalarından bir tanesi olan Topluluk Karar Ağaçları kullanılmıştır. Sınıflandırma yapılırken modelin doğruluk oranını ve performansının yorumlanmasında çeşitli yaklaşımlar kullanılmaktadır. Bu modelin başarı oranı analizinde şu yaklaşımlar ele alınmıştır; doğruluk, duyarlılık, özgüllük, kesinlik, F-ölçümü, KAPPA ve AUC Performans değerlendirme kriterleri kullanılmıştır [10]. Bu modelde;

- TP, KVH sahip bireyin gerçekten hasta olması
- FP, KVH sahip bireyin KVH olarak etiketlenmesi
- TN, KVH olmayan bireyin gerçekten hasta olmaması
- FN, KVH olan bireyin KVH değil olarak etiketlenmesi

Anlamlarına gelmektedir.

Duyarlılık (Sensitivity); Duyarlılık kriteri, aslında kalp hastalığından muzdarip olan ve pozitif olduğu tahmin edilen vakaların sayısını gösterir.

$$Duyarlılık = \frac{TP}{(TP + FN)} \quad (1)$$

Özgüllük (Specificity); Sınıflandırıcının etkilenmemiş örnekleri tahmin etmedeki etkinliği, özgüllük kriterine bağlı belirlenir. Denklem 2’de özgüllük tahmininin matematiksel ifadesi verilmiştir.

$$\text{Özgüllük} = \frac{TN}{(TN + FP)} \quad (2)$$

Kesinlik (Precision); Kesinlik kriteri ile kalp hastası olarak tahmin edilen örnek sayısının gerçekten pozitif olup olmadığını belirlemek için kullanılır. Denklem 3’te kesinlik tahmininin matematiksel ifadesi verilmiştir.

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} \quad (3)$$

Doğruluk (Accuracy); Doğruluk bir çalışmanın sonucunda elde edilen başarı oranıdır. Tüm tahminlerin doğru tahminlere oranlanması ile bulunur. Denklem 4 ve 5’te doğruluk ve hata oranı ifadelerinin sayısal olarak gösterimi verilmiştir.

$$\text{Doğruluk} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$\text{Hata Oranı}(ERR) = 1 - ACC \quad (5)$$

F-ölçüsü(F-Measure); F-ölçüsü kesinlik ve duyarlılık değerlendirmelerinin harmonik ortalamasını göstermektedir. Denklem 6’da F-ölçüsü tahmininin matematiksel ifadesi verilmiştir.

$$F - \text{Ölçüsü} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (6)$$

4. Bulgular ve Tartışma

Çalışmada hedeflenmek istenen hastanın yaşı, hastanın kadın mı erkek mi olduğu bilgisi, göğüs ağrısı tipi, istirahat kan basıncı, kolesterol değeri, açlık kan şekeri, istirahat elektrokardiyografik sonuç, maksimum nabız, egzersize bağlı anjin, egzersiz vb özelliklerin kardiyovasküler hastalıkların teşhisi ile ilişkisini analiz etmek ve kardiyovasküler hastalık teşhisini gerçekleştirmektir. Bu modelin en doğru şekilde teşhis yapabilmesi için makine öğrenmesi algoritması olan karar ağacı algoritmaları kullanılmıştır. Modelin veri analizi işlemleri sonucunda 1025 adet veriden 526 tane kardiyovasküler hasta örneği ve 499 tane kardiyovasküler hasta olmayan örneği veri örneği kullanılmıştır. Çalışmamızda veri setimizden elde ettiğimiz 14 özellik kullanılmıştır. Bu özellikler kullanılarak algoritmanın sağlıklı hasta ayrımı yapması (0,1) hedeflenmiştir. Veri setinin bilgisayar ortamında analiz edilmesi ardından özellik seçme algoritmalarından biri olan Eta özellik seçme algoritmasına tabi tutulmuştur. Özellik seçme işlemi modele ait en iyi özelliklerin seçilmesi ve işlemidir. Özellik seçme ile modelin başarı oranının artırılması amaçlanmıştır. Geliştirilen modelde makine öğrenmesi metodu kullanılarak sınıflandırma işlemi yapılmıştır ve modellerin yeni vaka tahminindeki başarı değerleri araştırılmıştır. Son zamanlarda yaygın olarak kullanılan yapay zekâ algoritmalarından biri olan karar ağaçları ile modelin başarı oranı hesaplanmıştır. Bu adımda age, sex, cp, trestbps, fbs, restecg, thalach, exang ve oldpeak ve diğer özellikler kullanılmıştır. KVH teşhisini koyabilmek için model test veri kümesi ve eğitim veri kümesi olmak üzere 2 gruba ayrılmıştır. Test sınıfının büyüklüğü %20 olarak belirlenmiştir. Karar ağacı algoritma yapısında kriter olarak entropi tercih edilmiştir. Modelin başarı oranı ;özgüllük, duyarlılık, kesinlik, F-ölçüm ve doğruluk metriklerine göre belirlenmiştir. Seçilen farklı özelliklere göre doğruluk oranları incelenmiş ve en yüksek başarı yüzdesine sahip modelin sonuçları baz alınmıştır. Yapılan

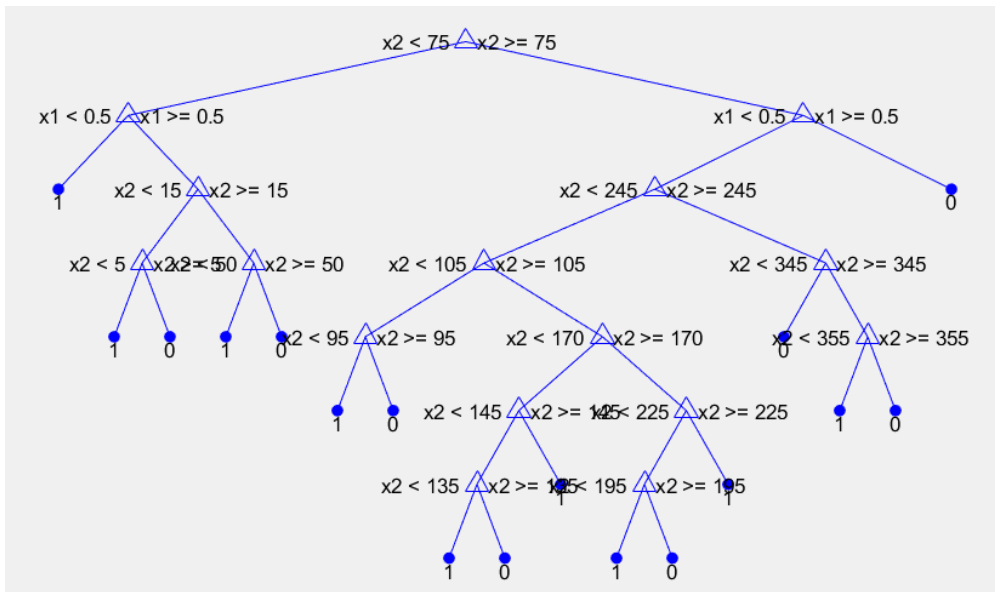
hesaplamalar sonucunda; %95,56 Doğruluk, %97,9 Özgüllük, %93,58 Duyarlılık, %95,7 F-ölçüm değerleri alınmıştır. %91,21 Kappa ve %95,4 AUC değeri ile KVH teşhisi tespit ettiği gözlemlenmiştir.

Merve Esra Taşçı ve Rüya Şamlı veri madenciliği ile kalp hastalığı teşhis etme adlı çalışmada çeşitli yapay zekâ metotlarını uygulamış ve çalışmalarında her metot için yüksek başarı oranı elde etmişlerdir [11]. Özlem Persil Özkan ve diğerlerinin yaptıkları çalışmada bireylerin sağlıklı yaşam biçimini benimsemelerinin yaşam kalitesine önemli katkılar sağlayacağını ve hastalığın tespitini kolaylaştıracağını belirlemişlerdir [12]. Özge Ekrem ve arkadaşları performans değerlendirme metriklerini baz alarak başarı oranını incelemişlerdir ve uyguladıkları Rastgele Orman algoritmasının sonucunda %86,88 doğruluk, %85,71 özgüllük, %87,87 duyarlılık, %87,87 kesinlik ve %87,87 F-ölçüsü oranı ile birlikte kabul edilebilir bir oran elde etmişlerdir [10]. Hamdi Sayın ve diğerleri ise yaptıkları benzer çalışmada tasarladıkları CNN algoritmasında hata oranı %5,3, duyarlılık oranı %94,4 özgüllük %94,6, F değeri %94,4 ve doğruluk oranı %94,7 beş farklı performans değerlendirmişlerdir ve yüksek doğruluğa ulaşmışlardır [13].

Tablo 3. Seçilen Özelliklere Göre Modellerin Performans Değerlendirme Kriterleri

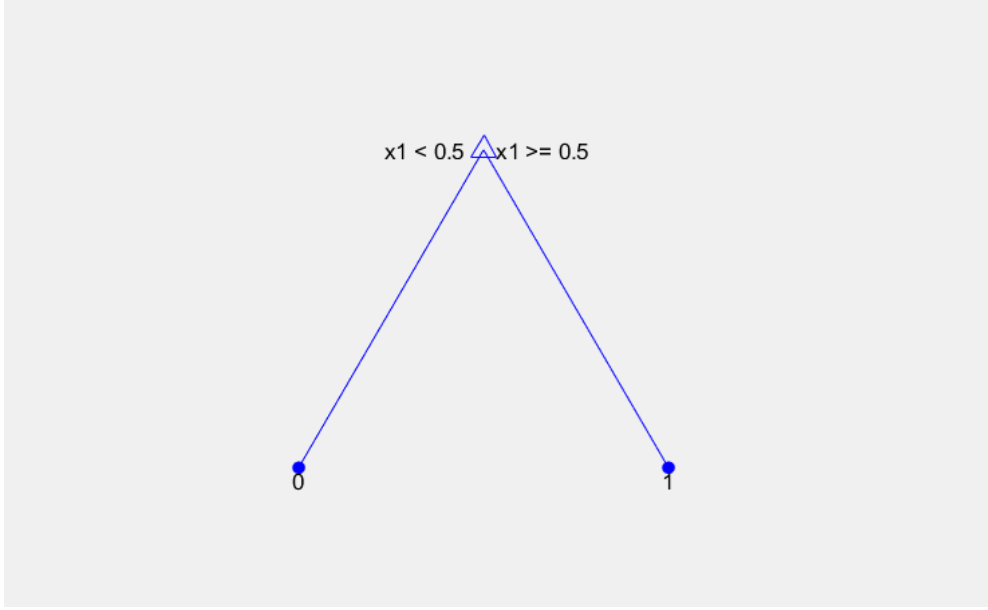
	Doğruluk	Duyarlılık	Özgüllük	F-Ölçüm	Kappa	AUC
Model 1	95,6098	0,9792	0,9358	0,9570	0,9121	0,9575
Model 2	94,1463	0,9792	0,9083	0,9424	0,8830	0,9437
Model 3	94,1463	0,9375	0,9450	0,9412	0,8825	0,9412
Model 4	93,1707	0,9375	0,9266	0,9320	0,8630	0,9321
Model 5	93,1707	0,9375	0,9266	0,9320	0,8630	0,9321
Model 6	89,2683	0,9583	0,8349	0,8923	0,7864	0,8966
Model 7	81,9512	0,7396	0,8899	0,8078	0,6346	0,8147
Model 8	80,4878	0,7708	0,8349	0,8016	0,6072	0,8028
Model 9	77,5610	0,7604	0,7890	0,7744	0,5494	0,7747
Model 10	77,5610	0,7604	0,7890	0,7744	0,5494	0,7747
Model 11	77,5610	0,7604	0,7890	0,7744	0,5494	0,7747

Çalışmada en yüksek doğruluk oranına ait (Model 1) Topluluk Karar Ağacı Algoritması şekil 2’de modellenmiştir.



Şekil 2. Model 1’e ait karar ağacı teşhis algoritması

Hekimler tarafından algoritmanın daha anlaşılır ve kolay karar verilebilir yapıda olması için şekil 2’de verilen modele budama uygulaması yapılmıştır ve şekil 3’te gösterilmiştir.



Şekil 3. Model 1’e ait karar ağacı budama uygulaması

5. Sonuç ve Öneriler

KVH tanısı konulan bireylerin sağlıklı yaşam standartlarına daha fazla uyum göstermeleri yaşam kalitelerinin artmasına önemli katkılar sağlayacaktır. Sağlıklı yaşam davranışlarının yaşam boyu korunabilmesini sağlamak için hastaların belirli zaman aralıklarıyla izlenmesi önerilmektedir. Yaş, cinsiyet ailenin tıbbi geçmişi değiştirilemez risk faktörlerindedir ancak pasif yaşam, fazla kilo, stres, tütün ve alkol kullanımı gibi faktörler değiştirilebilir risk faktörleri grubunda yer almaktadır. Yine aynı şekilde diyabet, kolesterol yüksekliği, hipertansiyon ilaçla tedavi edilerek kontrol altına alınabilir. Bu risk faktörlerinin önceden tayin edilmesi ve erken tanı konulması KVH hastalığına sahip olan bireylerin yaşamlarını sağlıklı bir şekilde devam ettirmeleri için son derece önem taşımaktadır. Geliştirilen bu model KVH hastalığına sahip bireyleri başarı ile tespit etmiştir. Bu çalışma ile KVH hastalığına sahip bireylerin verileri analiz edilmiş olup literatüre katkı sağlayacağı varsayılmaktadır. Sonuç olarak bu çalışma KVH için klinikte erken teşhis amaçlı kullanılabilir.

Teşekkür

Çalışmada ‘Heart Disease UCI’ veri setini açık kaynak erişimli internet sitesine (kaggle.com) aktaran kişi/kişilere teşekkürlerimizi sunarız.

Yazar(lar)ın Katkıları

BK, bilgisayar ortamında gerçekleştirilen faaliyetlerden ve algoritmanın kodlanmasından sorumludur. NT, bilimsel yayın araştırması ve literatür taramasından sorumludur. BK, NT makale yazımında katkı sağlamışlardır. MKU makalenin organizasyonundan sorumludur. Tüm yazarlar makalenin son halini okudu ve onayladı.

Çıkar Çatışması

Yazarlar herhangi bir çıkar çatışması beyan etmemişlerdir.

Kaynakça

- [1]. Nabiyev V.V., “Yapay Zeka: İnsan-Bilgisayar Etkileşimi”, Seçkin Yayıncılık, 2012.
- [2]. Aytekin A., Sönmez Çakır F., Yücel Y. B., Kulaözü İ., “Algoritmaların Hayatımızdaki Yeri ve Önemi”, *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi*, cilt 5, no. 7, pp. 143-150, 2018.
- [3]. Aitkenhead M., “A Co-Evolving Decision Tree Classification Method”, *Expert Systems with Applications*, cilt 34, no. 1, pp. 18-25, 2008.
- [4]. Kotsiantis S.B., “Decision trees: a recent overview”, *Artificial Intelligence Review*, no. 4, pp. 261-283, 2013.
- [5]. Kasapoğlu E.S., Enç N., “Koronar Arter Hastaları için Bir Rehber”, *Journal of Cardiovascular Nursing*, cilt 8, no. 15, pp. 1-7, 2017.
- [6]. Bulut F., “Determining Heart Attack Risk Ration Through Adaboost/ Adaboost ile Kalp Krizi Risk Tespiti”, *CBÜ Fen Bil.Dergi.*, cilt 12, no. 3, pp. 459-472, 2016.
- [7]. Budak H., “Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım”, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, cilt 22, no. Özel Sayı, pp. 21-31, 2018.
- [8]. Chen Y., Wang T., Wang B., Li Z., “A Survey of Fuzzy Decision Tree Classifier”, *Fuzzy Information And Engineering*, cilt 2, no. 6, pp. 149-159, 2009.
- [9]. Todorovski L., Džeroski S., “Combining Classifiers with Meta Decision Trees», *Machine Learning*, pp. 223-249, 2003.
- [10]. Ekrem Ö., Salman O. K. M., Aksoy B., İnan S. A., “Yapay Zekâ Yöntemleri Kullanılarak Kalp Hastalığının Tespiti”, *Mühendislik Bilimleri ve Tasarım Dergisi*, cilt 8, no. Özel Sayı: : Uluslararası Mühendislikte Yapay Zeka ve Uygulamalı Matematik Konferansı (UMYMK 2020), pp. 241-254, 2020.
- [11]. Taşcı M. E., Şamlı R., “Veri Madenciliği İle Kalp Hastalığı Teşhisi”, *Avrupa Bilim ve Teknoloji Dergisi*, no. Özel Sayı, pp. 88-95, 2020.
- [12]. Persil Özkan Ö., Büyükkunal S. K., Yiğit Z., İnci Z., Şakar F. Ş., Özçelik Ersü D., “Kardiyovasküler Hastalık Tanısı Almış Hastaların Sağlıklı Yaşam Biçimi Davranışlarının Değerlendirilmesi”, *Mersin Univ Sağlık Bilim Dergisi*, cilt 12, no. 1, pp. 22-31, 2019.
- [13]. Sayın H., Salman O. K. M., Aksoy B., Köse U., “EKG Sinyallerini kullanarak Kalp Ritimlerinin Yapay Zekâ ile Sınıflandırılması”, *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, cilt 9, no. 1, pp. 7-15, 2020.