



Performance-enhanced KNN algorithm-based heart disease prediction with the help of optimum parameters

Hidayet Takcı*^{ID}

Computer Engineering Dept., Engineering Faculty, Sivas Cumhuriyet University, 58140, Sivas, Türkiye

Highlights:

- Heart disease detection
- Machine learning based diagnosis
- Genetic algorithms

Keywords:

- Heart disease detection
- K nearest neighbor algorithm
- Minkowski distances

Article Info:

Research Article
Received: 31.07.2021
Accepted: 11.02.2022

DOI:

10.17341/gazimmfd.977127

Correspondence:

Author: Hidayet Takcı
e-mail:
htakci@cumhuriyet.edu.tr
phone: +90 532 663 5315

Graphical/Tabular Abstract

In this study, the KNN algorithm was used for the diagnosis of heart diseases. In order to increase the performance of the algorithm, parameter adjustment has been made and in this context; Manhattan, Euclidean and Chebyshev distance measurements, Uniform and Distance weighting methods and neighbor numbers between 1...15 have been tested on UCI Statlog and Cleveland datasets. Figure A summarizes the study.

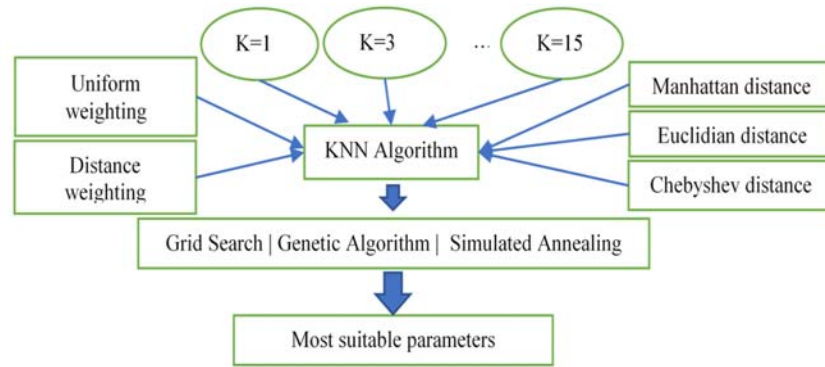


Figure A. The proposed method architecture

Purpose:

The aim of this study is to diagnose heart disease with high performance with the help of K nearest neighbor algorithm supported by different parameters and meta-heuristic methods. Within the scope of the study, parameter optimization was made.

Theory and Methods:

The KNN algorithm was preferred for the diagnosis of heart disease. In the study, the performance of the KNN algorithm is discussed together with the number of neighbors, weight method and distance measurements. Metaheuristic methods were also used in the study aiming to obtain the most appropriate KNN parameters. Among the metaheuristic methods, genetic algorithms and simulated annealing were preferred.

Results:

The aim of the proposed method was to increase the classification success by parameter optimization. We achieved this aim in the study and achieved a success above the literature average. The obtained classification accuracy was measured as 90.11%. Parameter optimization, genetic algorithms and simulated annealing techniques have increased their accuracy in heart disease classification.

Conclusion:

Although the KNN algorithm is a traditional method, it has been quite successful in diagnosing heart diseases. Parameter determination and optimization studies have had a great impact on the success of the algorithm. In our study, especially genetic algorithms significantly increased the classification success.



Optimum parametreler yardımıyla performansı artırılmış KNN algoritması tabanlı kalp hastalığı tahmini

Hidayet Takcı*^{ID}

Sivas Cumhuriyet Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 58140, Sivas, Türkiye

Ö N E Ç I K A N L A R

- Kalp hastalığı tespiti
- Makine öğrenimi tabanlı tanılama
- Genetik algoritmalar

Makale Bilgileri

Araştırma Makalesi
Geliş: 31.07.2021
Kabul: 11.02.2022

DOI:

10.17341/gazimmfd.977127

Anahtar Kelimeler:

Kalp hastalığı teşhisi,
makine öğrenmesi,
K en yakın komşu sınıflayıcısı

ÖZ

Kalp hastalıkları ölüm oranı yüksek hastalıklar grubundadır. Hastalığın teşhisinde bugüne kadar klinik yöntemler ve makine öğrenmesi yöntemleri sıklıkla kullanılmıştır. Bu çalışmada teşhis için KNN algoritması kullanılmıştır. Algoritmanın performansını artırmak için parametre ayarlama yapılmış ve bu kapsamda; Manhattan, Euclidean ve Chebyshev uzaklık ölçümleri, Uniform ve Distance ağırlıklandırma yöntemleri ve 1...15 arasındaki komşu sayıları UCI Statlog ve Cleveland veri kümeleri üzerinde denenmiştir. Statlog veri kümesi için en yüksek sınıflandırma doğruluğu %67,90 olup bu değer komşu sayısı 5, uzaklık yöntemi Euclidean ve ağırlıklandırma Distance ile elde edilmiştir. Aynı veri seti üzerinde genetik algoritmalar da çalıştırılmış ve komşu sayısı 5, uzaklık yöntemi Euclidean ve ağırlıklandırma Distance için %88,88 doğruluk elde edilmiştir. Cleveland veri kümesi için elde edilen maksimum sınıflandırma doğruluğu optimizasyon öncesi %71,42 iken optimizasyon sonrası %90,11 olarak ölçülmüştür. Cleveland veri kümesi için en yüksek sınıflandırma doğruluğu veren parametreler ise; komşu sayısı 3, uzaklık yöntemi Manhattan ve ağırlıklandırma Uniform olmuştur.

Performance-enhanced KNN algorithm-based heart disease prediction with the help of optimum parameters

H I G H L I G H T S

- Heart disease detection
- Machine learning based diagnosis
- Genetic algorithms

Article Info

Research Article
Received: 31.07.2021
Accepted: 11.02.2022

DOI:

10.17341/gazimmfd.977127

Keywords:

Heart disease diagnosis,
machine learning,
K nearest neighbor classifier

ABSTRACT

Heart diseases are diseases with a high mortality rate. Clinical methods and machine learning methods have been used frequently in the diagnosis of the disease. In this study, the KNN algorithm was used for diagnosis. In order to increase the performance of the algorithm, parameter adjustment has been made and in this context; Manhattan, Euclidean and Chebyshev distance measurements, Uniform and Distance weighting methods and neighbor numbers between 1...15 have been tested on UCI Statlog and Cleveland datasets. The highest classification accuracy for the Statlog dataset is 67.90%, which is obtained with the number of neighbors = 5, the distance method = Euclidean, and the weighting = Distance. Genetic algorithms were also run on the same data set and 88.88% accuracy was obtained for the number of neighbors = 5, distance method = Euclidean and weighting = Distance. While the maximum classification accuracy obtained for the Cleveland dataset was 71.42% before optimization, it was measured as 90.11% after optimization. The parameters that give the highest classification accuracy for the Cleveland dataset are; number of neighbors = 3, distance method = Manhattan and weighting = Uniform.

1. Giriş (Introduction)

Kalp hastalıkları tüm dünyada meydana gelen ölüm vakalarının başlıca sebepleri arasında yer alırlar. Hastalık kontrol ve önleme merkezi tarafından yayımlanan bir rapora göre [1] her yıl yaklaşık 600.000 kişi kalp hastalıkları nedeniyle hayatını kaybetmektedir. Kalp hastalıkları, üzerinde çokça çalışılması gereken bir konu olup kalp kaynaklı ölüm oranlarını düşürmenin bir yolu hastalığın doğru ve zamanında teşhis edilmesidir. Kalp hastalıkları sıklıkla aşırı yorgunluk hissi, mide bulantısı, boyun veya sırt ağrısı, nefes darlığı ve benzeri belirtilerle kendini gösterir. Kalp hastalıklarının kesin teşhisi için röntgen, elektrokardiyogram, ekokardiyogram, fiziksel egzersiz, kan testleri ve anjiyo ile desteklenmiş doktor meyanesine ihtiyaç vardır. Bununla birlikte özellikle klinik verilerin artışı kalp hastalıklarının uzmanlar tarafından teşhisini her geçen gün daha da zorlaştırmaktadır [2]. Artan klinik veriler ve bu veriler arasında ilişki kurmanın karmaşıklığı gibi sorunlar nedeniyle teşhiste makine öğrenmesi kullanımı artık kuvvetli bir seçenek haline almıştır [3]. Makine öğrenmesi bakış açısıyla kalp hastalıklarının teşhisi bir denetimli öğrenme görevidir. Makine öğrenmesine dayalı teşhis sayesinde geçmiş tecrübeler ışığında daha hassas değerlendirmeye yapma olanağı artmıştır. Tarihsel veriler ile eğitilen teşhis modelleri kişilerin kalp hastası olup olmadığı hakkında yüksek doğruluklu tahminler üretebilmektedir.

Bugüne kadar hastalığın teşhisinde Naive Bayes [4] ve Destek Vektör Makineleri [5] başta olmak üzere çok sayıda makine öğrenmesi algoritması kullanılmıştır. Karayılan ve Kılıç [6] kalp hastalıklarının teşhisi için geri yayılım algoritmasını kullanan bir yapay sinir ağı mimarisi önermişlerdir. UCI kalp hastalığı veri kümesi üzerinde yapılan ilgili çalışmada %95 doğruluk elde edilmiştir. Atkov vd. [7] kalp hastalıklarına etki eden genetik ve genetik olmayan faktörleri bir arada inceleyerek yapay sinir ağlarına dayalı bir model önermişlerdir. Ağ mimarisi ve girdi faktörleri değiştirilerek yapılan deneylerde %64 - %94 arasında doğru teşhis oranları elde edilmiştir. En iyi doğruluk değeri genetik ve genetik olmayan faktörleri kullanan, iki gizli katmanı olan bir mimari ile elde edilmiştir. Chitra vd. [8] optimize edilmiş bir yapay sinir ağı mimarisi yardımıyla kalp hastalığı olasılığını tahmin eden akıllı bir sistem tasarlamışlardır. Çalışmada, genetik algoritmalar ile desteklenen ileri beslemeli bir ağ yardımıyla %85,3 doğruluk, %83,5 hassasiyet ve %86,4 kesinlik elde edilmiştir. Yazid vd. [9] hastalık teşhisi için yapay sinir ağlarını kullanmış ve parametre ayarlama yaparak performans artışı hedeflemişlerdir. Yapılan çalışmada Cleveland veri kümesi üzerinde bir tahmin modeli ortaya konmuş ve %90,9 doğruluk oranı elde edilmiştir.

Alanla ilgili bazı çalışmalarda birden fazla makine öğrenme algoritması sırayla kullanılmış ve algoritmalar başarı açısından kıyaslanmıştır. Bu kapsamda yapılan bir çalışmada [10] altı farklı makine öğrenmesi algoritması 12 farklı veri kümesi üzerinde çalıştırılmıştır. Kalp hastalıkları veri kümesinde elde edilen doğruluk oranı %59,72 olarak ölçülmüştür. Çalışma bulgularından birisi yapay sinir ağlarının küçük ve orta büyüklükteki veri kümelerinde daha yüksek doğruluklar vermesi ve KNN algoritmasının daha kısa sürede sonuç üretmesidir. Mohan vd. [11], teşhis için doğrusal bir modele dayalı hibrit rastgele orman algoritması kullanan bir sistem önermişlerdir. Bu çalışma, çok sayıda sınıflayıcı ve çeşitli özellik alt kümeleri üzerinde uygulanmıştır. Önerilen rastgele orman algoritması ile %88,7 teşhis doğruluğu elde edildiği raporlanmıştır. Pawlovsky [12] tarafından yapılan çalışmada KNN algoritması kullanılarak mesafeye dayalı bir sistem geliştirilmiş ve teşhiste %85 doğruluk oranı elde etmiştir.

Kalp hastalıklarının teşhisinde tekli sınıflayıcılar kadar çoklu sınıflayıcılar da kullanılmıştır. Bu kapsamda Shouman vd. UCI kalp

hastalığı veri kümesi üzerinde yaptıkları bir çalışmada [13] çoklu sınıflayıcıların kullanımının teşhis başarısını artırdığını raporlamışlardır. Abushariah vd. [14] Matlab yardımıyla MLP ve ANFIS algoritmalarını teşhis için bir arada kullanmışlardır. Çalışma sonucuna göre MLP algoritması ANFIS algoritmasının sonucunu iyileştirmiştir. Cleveland kalp krizi veri kümesi üzerinde elde edilen doğruluk %87,04 olarak ölçülmüştür. Malav vd. UCI kalp hastalığı veri kümesi üzerinde yaptıkları çalışmada [15] kalp hastalıkları ile ilgili 14 adet özelliği farklı şekillerde gruplayarak çalıştırmışlardır. K-ortalama ve geri yayılım algoritmasının birlikte kullanıldığı çalışmada elde edilen sınıflandırma doğruluğu %97 olarak ölçülmüştür.

Özellik seçiminin teşhis başarısı üzerindeki etkisini görmek üzere de çeşitli çalışmalar yapılmıştır. Alzahani vd. [16] kalp hastalıkları veri kümesi üzerinde özellik seçimi yapmış ve özellik seçimi yöntemlerinin teşhis başarısını artırdığını raporlamışlardır. Dolatabadi vd. [17] EKG sinyalleri üzerinde temel bileşen analizi ile özellik seçimi yapmış ve koroner arter hastalarının otomatik teşhisi için bir yöntem ortaya koymuşlardır. Çalışmada Destek Vektör Makinesi kullanılmış ve sınıflandırma doğruluğu %99,2 olarak elde edilmiştir.

Kalp hastalıklarının teşhisinde klasik makine öğrenmesi algoritmalarına ek olarak derin öğrenme algoritmaları da kullanılmıştır [18]. Bu çalışmalardan birinde koroner arter hastalığı derin öğrenme algoritmaları yardımıyla teşhis edilebilmiştir [19]. Yapılan çalışmada, bir derin oto kodlayıcı yeniden yapılandırma hatasını tahmin etmek için kullanılmış ve çalışma sonunda %86,34 doğruluk, %91,37 hassasiyet ve %82,90 kesinlik değeri elde edilmiştir. Wiharto vd. tarafından yapılan çalışmada [20] koroner arter hastalığının teşhisi için bir istihbarat sistemi önerilmiştir. Performans artışı maksadıyla verideki anormalliklerin giderilmesi ve yeniden örnekleme çalışılmıştır. Yeniden örnekleme verileri üzerinde K-yıldız algoritması kullanılmış ve sonuçlar AUC başta olmak üzere çeşitli performans metrikleriyle sunulmuştur. Elde edilen en yüksek AUC değeri %87,5 olarak ölçülmüştür. Geçmiş çalışmalardan görüleceği üzere makine öğrenmesi algoritmaları kalp hastalığı teşhisinde yoğun olarak kullanılmıştır. Bununla birlikte kalp kaynaklı ölümlerin sıklığı nedeniyle daha yüksek doğruluğa sahip yöntemlere hala ihtiyaç vardır. Bu amaca yönelik olarak çalışmamızda kalp hastalıklarının teşhisi için KNN algoritması optimize edilmeye çalışılmıştır. Algoritmanın performansını etkileyen parametreler uzaklık yöntemi, komşu sayısı ve ağırlıklandırma. Performans artışı için çalışmamızda farklı uzaklık yöntemleri, komşu sayıları ve ağırlıklandırma yöntemleri ile deneyler yapılmış ve en uygun parametreler elde edilmiştir. Parametre belirlemede ızgara arama, genetik algoritmalar ve benzetilmiş tavlama yöntemleri kullanılmıştır.

2. Deneysel Metot (Experimental Method)

KNN algoritması başta örüntü tanıma olmak üzere yapay zeka alanında yoğun olarak kullanılan uzaklık tabanlı bir sınıflayıcıdır. KNN tabanlı sınıflandırmada test örnekleri ile eğitim örnekleri arasındaki uzaklıklar hesap edilerek test örneğine en yakın K adet örnek seçilir. Ardından seçilen K adet örneğin sınıf bilgilerinden çoğunluk oylaması ile test örneğinin sınıfı elde edilir [21]. Bu çalışmada test örneğine yakın eğitim örneklerini bulmakta kullanılan uzaklık yöntemleri ilk parametre, test örneğinin komşu sayısı ikinci parametre ve ağırlıklandırma üçüncü parametre olarak ele alınmıştır. Yapılan deneyler neticesinde en yüksek sınıflandırma doğruluğu veren parametreler ızgara arama yöntemine göre elde edilmiştir. Ayrıca, en uygun özellik değerleri ile daha yüksek sınıflandırma doğrulukları elde edebilmek için genetik algoritmalar ve benzetilmiş tavlama teknikleri kullanılmıştır.

2.1. Kullanılacak parametreler (Parameters to use)

KNN algoritmasının performansını artırmak üzere seçilen uzaklık yöntemi, komşu sayısı ve ağırlıklandırma parametreleri bu kısımda açıklanmıştır.

2.1.1. Parametre 1: Uzaklık yöntemi (Parameter 1: Distance method)

Uzaklık yöntemi parametresi için bu çalışmada üç farklı uzaklık ölçümü kullanılmıştır. Bu ölçümler Minkowski uzaklık yönteminin formlarından oluşan ölçümlerdir. X ve C, n-boyutlu uzayda iki vektör olmak üzere bu iki vektör arasındaki Minkowski uzaklığı Eş. 1'de görüldüğü gibidir.

n : özellik adedi
 x_i : X vektörünün i^{th} özellik değeri
 c_i : C vektörünün i^{th} özellik değeri
 p : minkowski derecesi

$$\left(\sum_{i=1}^n |x_i - c_i|^p \right)^{1/p} \quad (1)$$

Minkowski eşitliğinde yer alan p değeri eşitliğin derecesini verir. Eşitlikte p=1 iken yöntem Manhattan (man) uzaklığına dönüşmektedir. Manhattan uzaklığı, X ve C vektörlerine ait boyutlar arasındaki farkların mutlak değerleri toplamına eşittir. Eşitlikte p=2 olduğunda uzaklık yöntemi Euclidian (euc) uzaklığına dönüşür. Euclidian uzaklığı boyutlar arası farkların kareleri toplamının karekökü olarak bilinir. Eşitliğin derecesi p= ∞ olduğunda ise uzaklık yöntemi Chebyshev (cheb) yöntemine dönüşür. Chebyshev yöntemi boyutlar arasındaki farkların maksimumunu verir [22]. Uzaklık yöntemleri için minimum değer 0 iken maksimum değer için bir sınır vermek mümkün değildir. $X = (x_1, x_2, \dots, x_n)$ ve $C = (c_1, c_2, \dots, c_n)$ R^n olmak üzere uzaklık yöntemlerine ait Eş. 2, Eş. 3 ve Eş. 4. aşağıda verildiği gibidir.

$$\text{Manhattan} = \sum_{i=1}^n |x_i - c_i| \quad (2)$$

$$\text{Euclidian} = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (3)$$

$$\text{Chebyshev} = \max(|x_i - c_i|) \quad (4)$$

KNN sınıflayıcı deneylerinde her üç uzaklık ölçümüne göre de deneyler yapılarak uzaklık parametresinin en uygun değeri tespit edilecektir.

2.1.2. Parametre 2: Komşu sayısı (Parameter 2: Number of neighbors)

KNN algoritmasının doğasına uygun olarak bir test örneğini sınıflandırabilmek için ona en yakın K adet eğitim örneğinin seçimine ihtiyaç vardır. K değerinin ne olması gerektiğine dair kesin bir görüş olmamakla birlikte komşu sayısının maksimum değeri konusunda bir öneri komşu sayısının eğitim örneklerinin karekökü kadar olmasıdır [21]. Bu nedenle çalışmamızda maksimum komşu sayısı eğitim örneklerinin kareköküne yakın bir değer olarak 15 şeklinde seçilmiştir. Ayrıca, kalp hastalıklarının teşhisi iki sınıflı bir problem ve KNN tabanlı sınıflandırma çoğunluk oylamasına dayalı bir algoritma olduğu için komşu sayısı değeri tek değerlerden seçilmiştir. Böylece 1..15 arasındaki tek sayılar (1, 3, 5, 7, 9, 11, 13, 15) bu çalışmada komşu sayısı olarak kullanılacaktır.

454

2.1.3. Parametre 3: Ağırlıklandırma (Parameter 3: Weighting)

KNN algoritması çoğunluk oylamasına dayalı bir sınıflayıcı olmasına rağmen kimi zaman yakın komşuların ne derece yakın olduklarını bilmeye ihtiyaç vardır. KNN algoritmasında eğer her komşu eşit ağırlıkla ele alınacaksa Uniform ağırlıklandırma her komşu aradaki mesafeye dayalı olarak ağırlıklandırılacaksa da Distance ağırlıklandırma seçilir. Mesafeye dayalı ağırlıklandırmada ağırlık faktörü $1/d$ ya da $1/d^2$ şeklinde verilmektedir [23]. Bu çalışmada ağırlıklandırma parametresi olarak Uniform ve Distance değerleri kullanılmıştır. Hangi yöntemin sınıflandırmaya daha fazla katkı sağladığı deneyler aşamasında tespit edilecektir.

2.2. Izgara arama ve metasezgisel yöntemler (Grid search and metaheuristics)

KNN algoritmasının performansını artırabilmek için en uygun parametreleri tespit eden izgara arama ve özellik değerlerini güncelleyen genetik algoritmalar ile benzetilmiş tavlama yöntemleri kullanılacaktır. Izgara arama yöntemi olası bütün parametre değerlerinin denenmesi ve en yüksek sonucu veren eşleşmenin bulunması amacıyla kullanılır. Izgara arama yöntemi ardından KNN sınıflayıcı performansını artırmak için optimizasyon teknikleri kullanılacaktır. Optimizasyon teknikleri, sıklıkla esin kaynağını doğadan alan ve bir problemin en iyi çözümünü bulmaya odaklanmış arama yöntemleridir. Örneğin, robotun bir hedefe giderken kullanacağı yol karıncaların yiyecek ararken kullandığı yöntemden esinlenir [24]. Bu çalışmada parametre ayarlama ve KNN sınıflayıcı performansını artırmak için genetik algoritmalar ve benzetilmiş tavlama gibi meta sezgisel teknikler kullanılmıştır. İlgili tekniklerin geçmiş çalışmalarda da bu alanda yoğun kullanım örnekleri bulunmaktadır [25]. Genetik algoritmalar evrime dayalı ve doğal seçilimi kendisine örnek alan bir arama yöntemidir. Çaprazlama, yeniden bir araya getirme, mutasyon, uygunluk ve en uygunun seçilmesi gibi adımlardan oluşur [26]. Genetik algoritmalar daha önce de KNN algoritmasını optimize etmek için kullanılmıştır [27]. Benzetilmiş tavlama algoritması da optimizasyon problemleri için tasarlanmış olasılığa dayalı bir algoritmadır [28]. En iyi çözümü en kısa zamanda bulmaya odaklanmış bir yöntemdir.

3. Sonuçlar ve Tartışmalar (Results and Discussions)

Bu çalışmada uzaklık ölçümü, komşu sayısı ve ağırlıklandırmanın sınıflandırma başarısına etkisi incelenmiş ve sınıflandırma sonuçları doğruluğa dayalı olarak karşılaştırılmıştır. Ayrıca model değerlendirme 10-kat çapraz doğrulama yöntemine göre yerine getirilmiştir.

3.1. Veriseti (Dataset)

Deneylerimizde UCI Statlog ve Cleveland kalp hastalıkları veri kümesi kullanılmıştır [29]. UCI Statlog veri kümesi her biri 13 özellik değeri ile sunulan 270 örnek içermektedir. Cleveland veri kümesi de Statlog veri kümesine benzer özellikler içerir. Statlog veri kümesinden farklı olarak Cleveland veri kümesinde tahmin değişkeni ile birlikte 14 özellik ve toplam 303 kayıt yer alır. Her iki veri kümesinde de yer alan değişkenler hastalık belirtileri ve teşhis aşamasında ölçülen değerler ile ilişkilidir. Veri kümesinde yer alan tahminci değişkenler ikili sayı veya sürekli sayı türündedir. Hem Statlog veri kümesine hem de Cleveland veri kümesine ait özellikler Tablo 1'de verilmiştir.

3.2. Doğruluğa dayalı karşılaştırmalar (Accuracy based comparisons)

Seçilen parametrelerin sınıflandırma doğruluğuna etkisini görmek için öncelikle farklı uzaklık yöntemleri farklı komşu sayıları ve

ağırlıklandırma yöntemleri ile deneyler yapılmış ve sonuçlar raporlanmıştır. Sonuçları karşılaştırmak ve daha başarılı olan seçimi elde edebilmek için model değerlendirme metriklerinden doğruluk ölçümü kullanılmıştır. Doğruluk, tahmin edilen sınıf ile gerçek sınıfın uyumunu oransal olarak veren bir ölçümdür. Örneğin, modelin kalp hastalığı riski var şeklinde tahmin ettiği vakanın gerçekten riskli olması ile modelin risk yok şeklinde tahmin ettiği vakanın gerçekten risksiz olması sayısının bütün vakalara oranı doğruluk olarak sunulur. Doğruluk, Tablo 2’de verilen karışıklık matrisine dayalı olarak tarif edilebilir.

Çalışmanın yöntemine uygun olarak her bir parametre değeri için deneyler yapılmış ve sonuçlar elde edilmiştir. Ayrıca, optimizasyon yöntemleri ile elde edilen sınıflandırma sonuçları da sunulmuştur. Doğruluğa dayalı ölçümlerde, önce Statlog veri kümesi için sonra da Cleveland veri kümesi için değerler aşağıdaki tablolarda yer

almaktadır. Deneyler sırasında öncelikle üç farklı parametre değerinin en uygun değerleri ızgara arama yöntemine göre bulunmaya çalışılmıştır. Yönteme göre komşu sayısı, ağırlık fonksiyonu ve uzaklık yöntemi belirlenen değer aralıklarında test edilmiş ve en iyi sonucu veren parametreler elde edilmiştir. Sınıflandırma doğruluğu komşu sayısı ile ilişkili olmakla birlikte komşu sayısı artarken doğruluk artar veya azalırken doğruluk azalır şeklinde bir şey söylemek her zaman mümkün değildir. Bununla birlikte Tablo 3’de görüleceği üzere en yüksek sınıflandırma doğruluğunu veren komşu sayısı 11 olarak tespit edilmiştir. Uzaklık yöntemleri açısından sonuçlara bakıldığında optimizasyon öncesi chebyshev yöntemi euclidean ve manhattan yöntemlerinden daha yüksek doğruluk vermiştir. Ayrıca uniform ağırlıklandırma distance ağırlıklandırmadan daha yüksek sınıflandırma doğruluğu vermiştir. Elde edilen en yüksek sınıflandırma doğruluğu %67,90 olarak ölçülmüştür.

Tablo 1. UCI Statlog ve Heart Disease veri kümesi (UCI Statlog and Heart Disease dataset)

UCI Statlog	UCI Heart Disease (Cleveland)
Yaş	Yaş
Cinsiyet	Cinsiyet
Göğüs ağrısı tipi (chest pain type) (4 farklı değer)	Göğüs ağrısı tipi (chest pain type) (4 farklı değer)
Dinlenme sırasındaki kan basıncı	Dinlenme sırasındaki kan basıncı
Mg/dl cinsinden serum kolesterolü (serum cholesterol in mg/dl)	Mg/dl cinsinden serum kolesterolü
Açlık kan şekeri > 120 mg/dl	Açlık kan şekeri > 120 mg/dl
Dinlenme sırasındaki elektrokardiyografik sonuçlar	Dinlenme sırasındaki elektrokardiyografik sonuçlar
Ulaşılan maksimum kalp hızı	Ulaşılan maksimum kalp hızı
Egzersize bağlı angina	Egzersize bağlı angina
Dinlenmeye göre egzersizin neden olduğu STEMI depresyonu	Dinlenmeye göre egzersizin neden olduğu STEMI depresyonu
Zirve egzersiz STEMI segmentinin eğimi	Zirve egzersiz STEMI segmentinin eğimi
Floroskopi ile renklendirilen büyük damar sayısı (0-3)	Floroskopi ile renklendirilen büyük damar sayısı (0-3)
Talyum Stres Testi Sonucu (thal) 3 = normal; 6 = kalıcı hasar; 7 = geri döndürülebilir hasar	Talyum Stres Testi Sonucu (thal) 3 = normal; 6 = kalıcı hasar; 7 = geri döndürülebilir hasar
	Kalp hastalığı teşhisi (najiyoğrafik hastalık statüsü status)

Tablo 2. Karışıklık matrisi (Confusion matrix)

		Tahmin edilen sınıf	
		Risk var	Risk yok
Gerçek Sınıf	Risk var	TP (True Positive)	FN (False Negative)
	Risk yok	FP (False Positive)	TN (True Negative)
$Doğruluk = (TP+TN)/(TP+FN+FP+TN)$			

Tablo 3. Statlog veri kümesi için sınıflandırma doğrulukları (Classification accuracies for the Statlog dataset)

K	Izgara arama			Genetik algoritma			Benzetilmiş tavlama		
	Man.	Euc.	Cheb.	Man.	Euc.	Cheb.	Man.	Euc.	Cheb.
Ağırlıklandırma yöntemi: Uniform									
1	60,49	49,38	54,32	81,48	75,30	80,25	71,43	62,96	50,61
3	66,66	60,49	60,49	79,01	69,13	72,83	66,11	74,07	54,32
5	60,49	60,49	64,19	79,01	67,90	80,25	66,62	61,72	50,61
7	64,19	62,96	64,19	80,25	59,25	66,66	58,17	45,67	59,25
9	64,19	64,19	64,19	82,72	65,43	80,25	75,30	70,37	51,85
11	65,43	61,72	67,90	85,18	80,25	74,07	65,43	81,48	60,49
13	64,19	64,19	60,49	83,95	65,43	83,95	75,30	64,19	74,07
15	66,66	62,96	60,49	83,95	77,77	79,01	77,77	59,25	59,25
Ağırlıklandırma yöntemi: Distance									
1	60,49	49,38	54,32	82,71	80,25	69,11	62,96	55,55	59,25
3	66,66	60,49	60,49	83,95	77,77	72,83	59,25	62,96	64,19
5	59,25	62,96	64,19	75,30	88,88	77,77	58,02	70,37	51,85
7	64,19	64,19	65,43	70,37	86,41	67,90	67,90	61,72	45,67
9	64,19	64,19	65,43	62,96	80,25	46,91	62,96	70,37	56,79
11	65,43	60,49	64,19	80,25	62,96	72,83	71,60	59,25	59,25
13	64,19	61,72	60,49	81,48	60,49	70,37	66,66	76,54	55,55
15	64,19	62,96	61,72	75,30	86,41	80,25	69,13	67,90	64,19

Sınıflayıcı doğruluğunu artırmak için optimizasyon teknikleri de kullanılmıştır. Tablo 3'de görüleceği üzere optimizasyon teknikleri sınıflandırma doğruluklarına pozitif yönde etki etmiştir. Statlog veri kümesinde optimizasyon öncesi en yüksek sınıflandırma doğruluğu %67,90 iken optimizasyon sonrası bu değer %88,88 olarak ölçülmüştür. Optimizasyon öncesi en başarılı komşu sayısı 11 iken optimizasyon sonrası komşu sayısı 5 olarak tespit edilmiştir. Uzaklık yöntemi euclidean ve ağırlıklandırma yöntemi distance olarak tespit edilmiştir. Genetik algoritmalar yöntemi ile sınıflandırma doğruluğu optimizasyon öncesine kıyasla %20,98 oranında artmıştır. Benzetilmiş tavlama yöntemi de sınıflandırma doğruluğunu artmıştır. Benzetilmiş tavlama yöntemi sonrası elde edilen maksimum sınıflandırma doğruluğu %81,48 olarak ölçülmüştür. Böylece optimizasyon tekniklerinin sınıflandırma doğruluğuna belirgin şekilde katkı sağladığı görülmüştür.

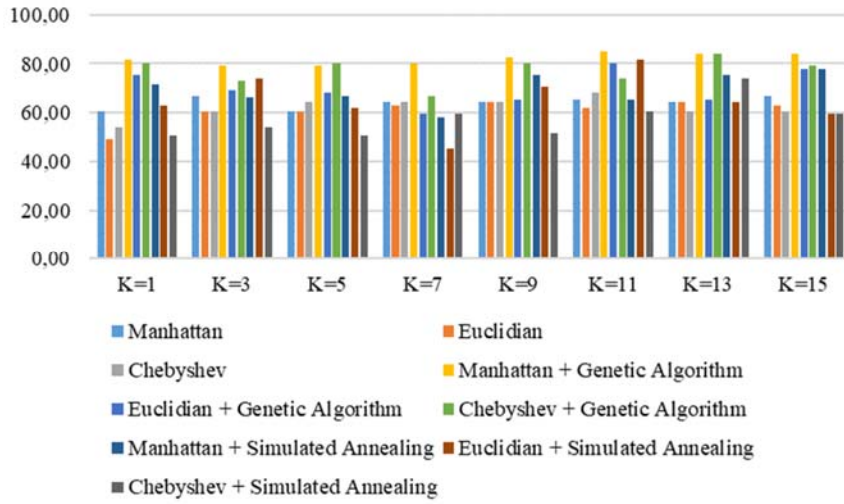
Şekil 1'de görüleceği üzere en yüksek sınıflandırma doğruluğu genetik algoritmalarla desteklenmiş ve manhattan uzaklığını kullanan (Manhattan + Genetic Algorithm) sınıflayıcı ile elde edilmiştir. Onu

genetik algoritmalarla optimize edilen ve chebyshev uzaklığını kullanan (Chebyshev + Genetic Algorithm) sınıflayıcısı takip etmiştir. Benzetilmiş tavlama yöntemiyle optimize edilen sınıflayıcıları onları takip etmiştir.

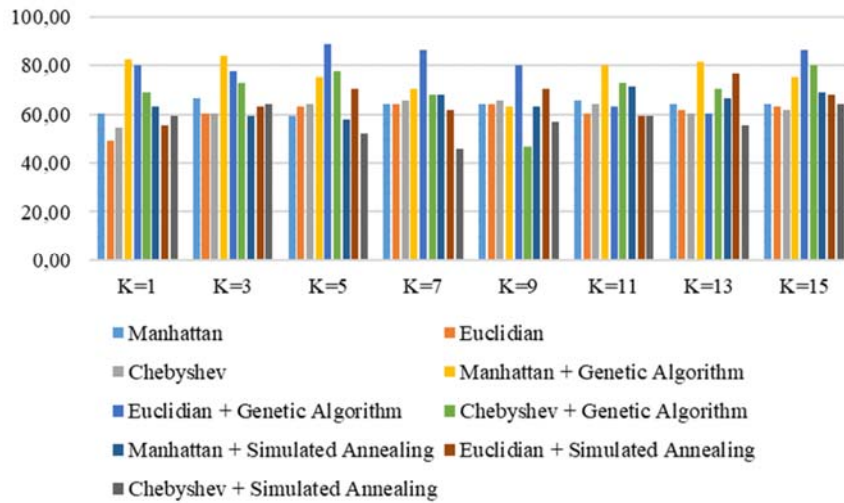
Şekil 2'de distance ağırlıklandırma ile ağırlıklandırılmış veri kümesi üzerinde elde edilen değerler görülmektedir. En yüksek sınıflandırma doğruluğu genetik algoritmalarla desteklenmiş euclidean uzaklık yöntemini kullanan sınıflayıcı (Euclidian + Genetic Algorithm) ile edilmiştir. Onu takip eden sınıflayıcı genetik algoritmalarla desteklenmiş ve manhattan uzaklığını kullanan (Manhattan + Genetic Algorithm) sınıflayıcı olmuştur.

Statlog veri kümesi için yapılan çalışmalar Cleveland veri kümesi için de yapılmıştır. Cleveland veri kümesinde sınıflandırma doğruluğunu artıran parametrelerin neler olduğu tespit edilmiştir.

Cleveland veri kümesi üzerinde yapılan deneylerde optimizasyon öncesi elde edilen en yüksek sınıflandırma doğruluğu %71,42 olarak



Şekil 1. Uniform ağırlıklandırma ile elde edilen sınıflandırma doğrulukları (Classification accuracies achieved with uniform weighting)



Şekil 2. Distance ağırlıklandırma ile elde edilen sınıflandırma doğrulukları (Classification accuracies achieved by distance weighting)

elde edilmiştir. Bu değeri veren komşu sayısı 3 ve 9, uzaklık yöntemi manhattan ve ağırlıklandırma yöntemi uniform ağırlıklandırmadır. Optimizasyon teknikleri uygulandıktan sonra Genetik algoritma destekli KNN %90,11 doğruluk verirken benzetilmiş tavlama destekli KNN %76,54 doğruluk vermiştir. Optimizasyon yardımıyla elde edilen sınıflandırma doğruluğundaki artış %18,69 olarak ölçülmüştür.

Tablo 4 Cleveland veri kümesi üzerinde yapılan çalışma sonuçlarını vermektedir. Cleveland veri kümesinde elde edilen sonuçlar genel olarak Statlog veri kümesinden daha yüksektir. Şekil 3'den görüleceği üzere genetik algoritmalar ile optimize edilen ve manhattan uzaklığı kullanan sınıflayıcı en yüksek değeri vermiştir. Onu genetik algoritmalar ile desteklenen euclidean uzaklığına dayalı sınıflayıcı

takip etmiştir. Benzetilmiş tavlama yöntemi de başlangıç durumuna göre sınıflandırma doğruluklarını artırmıştır.

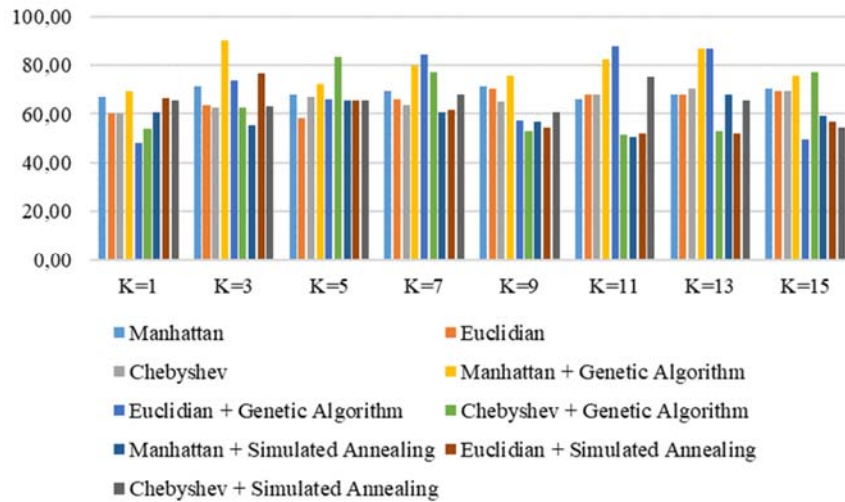
Şekil 4, distance ağırlıklandırma ile ağırlıklandırılmış verilerin doğruluklarını sunmaktadır. En yüksek sınıflandırma doğruluğunu genetik algoritmalarla desteklenen chebyshev uzaklık yöntemi vermiştir. Onu genetik algoritmalarla optimize edilen ve euclidean uzaklığı kullanan sınıflayıcı takip etmiştir.

3.3. Geçmiş çalışmalarla karşılaştırma (Comparison with past studies)

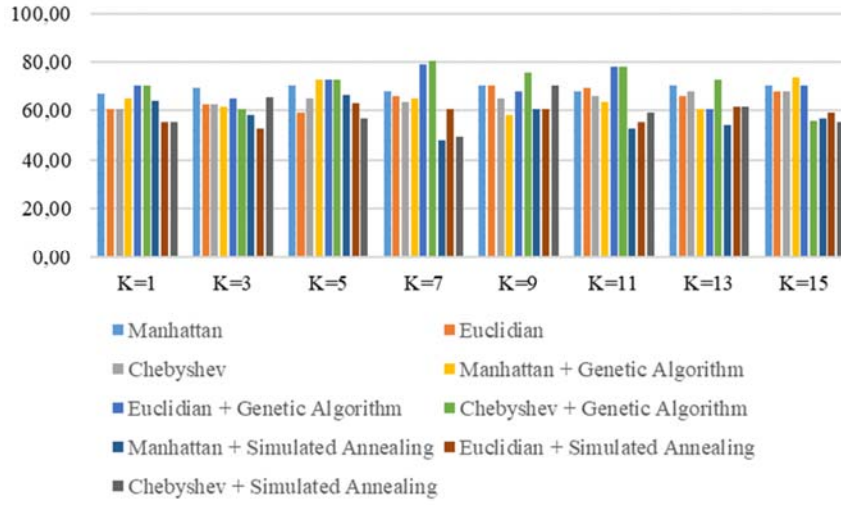
Bugüne kadar başta kalp krizi olmak üzere çok sayıda kalp hastalığının teşhisinde makine öğrenmesi algoritmaları kullanılmıştır.

Tablo 4. Cleveland veri kümesi için sınıflandırma doğrulukları (Classification accuracies for the Cleveland dataset)

K	Izgara arama			Genetik algoritma			Benzetilmiş tavlama		
	Man.	Euc.	Cheb.	Man.	Euc.	Cheb.	Man.	Euc.	Cheb.
Ağırlıklandırma yöntemi: Uniform									
1	67,03	60,43	60,44	69,23	48,35	53,84	60,49	66,66	65,43
3	71,42	63,73	62,63	90,11	73,62	62,63	55,55	76,54	62,96
5	68,13	58,24	67,03	72,52	65,93	83,51	65,43	65,43	65,43
7	69,23	65,93	63,74	80,22	84,61	76,92	60,49	61,73	67,90
9	71,42	70,32	64,83	75,82	57,14	52,74	56,79	54,32	60,49
11	65,93	68,13	68,13	82,41	87,91	51,65	50,61	51,85	75,30
13	68,13	68,13	70,33	86,81	86,81	52,74	67,90	51,85	65,43
15	70,33	69,23	69,23	75,82	49,45	76,92	59,26	56,79	54,32
Ağırlıklandırma yöntemi: Distance									
1	67,03	60,43	60,43	64,83	70,33	70,33	64,19	55,55	55,55
3	69,23	62,64	62,64	61,53	64,83	60,43	58,02	53,08	65,43
5	70,32	59,34	64,83	72,52	72,52	72,52	66,66	62,96	56,79
7	68,13	65,93	63,74	64,83	79,12	80,22	48,12	60,49	49,38
9	70,32	70,32	64,83	58,24	68,13	75,82	60,49	60,49	70,37
11	68,13	69,23	65,93	63,73	78,02	78,02	53,08	55,55	59,25
13	70,32	65,93	68,13	60,43	60,44	72,52	54,32	61,73	61,72
15	70,32	68,13	68,13	73,62	70,33	56,04	56,79	59,25	55,55



Şekil 3. Uniform ağırlıklandırma ile elde edilen sınıflandırma doğrulukları (Classification accuracies achieved with uniform weighting)



Şekil 4. Distance ağırlıklandırma yöntemi için sınıflandırma doğrulukları
(Classification accuracies for the distance weighting method)

Bu çalışmaların çoğunda Cleveland Kalp Hastalığı veri kümesi ve Statlog Kalp hastalığı veri kümesi kullanılmıştır. Bu bölümde özetlenen çalışmalar da ilgili veri kümelerini kullanan çalışmalardan seçilmiştir. 2016 yılında Suganya vd. [30] kalp hastalığını tahmin etmek için CART sınıflandırıcısının kullanımını önermiştir. CART sınıflandırıcı algoritmasının kalp hastalığı tahmin doğruluğu %83 olarak elde edilmiştir. Dangare vd.[31] kalp hastalığı tahmini için Karar Ağaçları, Naive Bayes ve Sinir Ağları algoritmalarını kullanmıştır. Statlog ve Cleveland veri kümeleri üzerinde yapılan çalışmada Karar ağaçları %96,66, Naive Bayes %94,44 ve Sinir ağları %99,25 doğruluk vermiştir. Sigara kullanımı ve obezite özelliklerinin eklenmesi ile doğruluk oranları Sinir ağlarında %100, karar ağaçlarında %99,62 ve Naive Bayes algoritmasında %90,74 olmuştur.

2015 yılında Abdar vd.[32] kalp hastalıkları riskini tahmin etmek için C5.0, Neural Network, SVM ve KNN yöntemlerini kullanmışlardır. C5.0, Sinir Ağları, SVM ve KNN'nin doğruluğu sırasıyla %93,02, %89,4, %86,05 ve %80,23'tür. C5.0 algoritması diğerlerine kıyasla daha iyi performans göstermiştir. Takcı tarafından UCI Statlog veri kümesi üzerinde yapılan çalışmada [33] kalp krizi tahmini için en uygun sınıflayıcı ve özellik seçim yönteminin tespiti hedeflenmiştir. 12 adet sınıflayıcı performans açısından kıyaslandıktan sonra özellik seçiminin doğru tahmine etkisi ölçülmüştür. Elde edilen değerlere göre doğrusal çekirdek kullanan destek vektör makinesi en başarılı sınıflayıcı ve ReliefF algoritması en başarılı özellik seçim yöntemi olmuştur. Sınıflayıcılardan elde edilen maksimum doğruluk %84,81 olarak ölçülmüştür. Dwivedi [34] kalp krizi tahmini için çok sayıda makine öğrenmesi algoritmasını UCI statlog kalp hastalığı veri kümesi üzerinde test etmiştir. Yapılan çalışmada; yapay sinir ağları, destek vektör makineleri, lojistik regresyon, KNN, sınıflandırma ağacı ve naive bayes algoritmalarını kullanılmış ve en iyi sonucu %85 doğruluk ile lojistik regresyon algoritması vermiştir.

Raza [35] tarafından yapılan çalışmada çoğunluk oylaması kullanan bir topluluk mimarisi önerilmiştir. Bir vakada kalp hastalığını tahmin etmek için lojistik regresyon, çok katmanlı algılayıcı ve Naive Bayes algoritmaları birleştirilmiştir. Bu birleştirme sonucunda %88,88'lik sınıflandırma doğruluğu elde edilebilmiştir. Bir başka çalışmada [36] mesafeye dayalı dayalı aykırı değer saptama modeli yardımıyla kalp hastalıkları teşhis deneyleri yapılmıştır. Sunulan modelin başarısını ortaya koymak üzere çeşitli performans ölçütleri altında Heart Statlog ve Cleveland kalp hastalığı veri kümesi üzerinde çalışma yapılmış ve Heart Statlog veri kümesi üzerinde maksimum %96,70 doğruluk elde edilmiştir.

edilmiştir. Kardiyovasküler hastalıklar konusunda yapılan bir araştırmada [37] teşhis için bir Uzman Sistem geliştirilmiştir. Amaç, hastalığı teşhis edecek kişiye uygun tavsiyenin verilmesidir. Teşhisi kolaylaştırmak ve kuralları azaltmak için Kaba Küme Teorisi kullanılmıştır. Bu araştırmada, UCI makine öğrenimi deposundan Statlog veri kümesi kullanılmıştır. Sistemin uygulanması için MATLAB kullanılmıştır.

KNN sınıflayıcı performansını artırmak için çalışmamızda yaptığımız gibi parametre iyileştirmeye yönelik çalışmalar da yapılmıştır. O çalışmalardan birinde, kalp hastalıkları veri kümesi üzerinde bir metasezgisel yöntem olan parçacık sürü optimizasyonu (PSO) özellik seçimi için kullanılmıştır [38]. UCI Statlog veri kümesi üzerinde yapılan çalışmada KNN algoritması ile PSO yöntemi birlikte çalıştırılmış ve %100 doğrulukla teşhis yerine getirilebilmiştir. PSO yardımıyla seçilen özellikler 8 tane ve K=1-15 arasındaki değerlerin tamamında %100 doğruluk elde edilmiştir. Çalışmada özelliklere IQR adı verilen bir ağırlık ilavesi de yapılmıştır.

Yine bir başka metasezgisel olan genetik algoritmalar da çok sayıda çalışmada sonuç iyileştirmede kullanılmıştır. Bu çalışmalardan birinde [39] genetik algoritmalar yardımıyla KNN sınıflayıcı performansı iyileştirilmiştir. Genetik algoritmalar sınıflandırmaya etki eden en iyi özellikleri seçmek konusunda destek olmuş ve doğruluk oranları %100 değerine ulaşmıştır.

Bu alandaki geçmiş çalışmaların bir özeti Tablo 5'de görülmektedir. Kalp hastalıklarının teşhisinde yapay sinir ağları, karar ağaçları, destek vektör makineleri, K en yakın komşu algoritması, derin öğrenme algoritmaları ve bulanık sistemler kullanılmıştır. Kimi çalışmalarda tekil olarak kullanılan algoritmalar kimi çalışmalarda da birlikte ele alınmıştır. Özellik seçimi, parametre iyileştirme ve diğer yöntemlerin kullanımı da yapılan çalışmalar arasındadır. Metasezgisel yöntemlerin sınıflandırma başarısını artırdığı ve sıklıkla özellik seçimi amacıyla kullanıldığı görülmüştür. Çalışmamızda KNN algoritması kullanılmış ve sınıflayıcı performansını artırmak için komşu sayısı, uzaklık yöntemi ve ağırlıklandırma parametreleri optimize edilmiştir. Model doğruluğunu artırmak için Genetik Algoritmalar ve Benzetimli Tavlama yöntemlerinden yararlanılmıştır. Çalışmamız hem UCI Statlog veri kümesi üzerinde hem de Cleveland veri kümesi üzerinde yapılmış olup optimizasyon desteği ile elde edilen en yüksek sınıflandırma doğruluğu %90,11 olarak elde edilmiştir.

Tablo 5. Geçmiş çalışmalarla karşılaştırma (Comparison with past studies)

Yazar	Algoritma	Veri kümesi	Maksimum Doğruluk (%)
Yazid vd. [9]	Yapay Sinir Ağları	UCI Cleveland Veri Kümesi	90,90
Suganya vd. [30]	CART sınıflayıcı	UCI Statlog Kalp Hastalığı Veri Kümesi	83,00
Dangare vd. [31]	Karar Ağaçları, Naive Bayes ve Sinir Ağları	UCI Statlog Kalp Hastalığı Veri Kümesi	99,25
Abdar vd. [32]	C5.0, Sinir Ağları, SVM ve KNN	UCI Statlog Kalp Hastalığı Veri Kümesi	93,02
Takcı [33]	12 farklı makine öğrenmesi algoritması	UCI Statlog Kalp Hastalığı Veri Kümesi	84,81
Dwivedi [34]	ANN / SVM / Lojistik regresyon / KNN / Sınıflandırma ağacı / Naive Bayes	UCI Statlog Kalp Hastalığı Veri Kümesi	85,00
Raza [35]	Lojistik regresyon, çok katmanlı algılayıcı ve Naive Bayes	UCI Statlog Kalp Hastalığı Veri Kümesi	88,88
Bhuvaneswari vd. [36]	Gradient Tree Boosting Algorithm	UCI Statlog Kalp Hastalığı Veri Kümesi	96,70
Jabbar [38]	KNN+PSO+IQR	UCI Statlog Kalp Hastalığı Veri Kümesi	100,00
Noor [39]	KNN+GA (Genetik algoritma)	UCI Statlog Kalp Hastalığı Veri Kümesi	100,00
Önerdiğimiz yöntem	KNN+GA	UCI Statlog ve Cleveland Veri Kümesi	90,11

4. Sonuçlar (Conclusions)

Bu çalışmanın amacı kalp hastalığı teşhisinde optimize edilmiş KNN algoritmasını kullanarak sınıflandırma doğruluğunu artırmaktır. Bu kapsamda komşu sayısı, uzaklık yöntemi ve ağırlıklandırma parametreleri ele alınmış ve en yüksek sınıflandırma doğruluğunu veren parametre değerleri izgara arama yöntemine göre tespit edilmiştir. Optimizasyon öncesinde Statlog veri kümesinde en yüksek sınıflandırma doğruluğu veren komşu sayısı=11, uzaklık yöntemi=chebyshev ve ağırlıklandırma yöntemi=Uniform olarak belirlenmiştir. Elde edilen en yüksek sınıflandırma doğruluğu %67,90'dır. Sınıflayıcı optimizasyonu için iki metasezgisel yöntem kullanılmış ve genetik algoritmalar yardımıyla %88,88 sınıflandırma doğruluğu elde edilmiştir. Cleveland veri kümesi için yapılan deneylerde de yine optimizasyon öncesi elde edilen maksimum doğruluk %71,42 iken optimizasyon sonrası elde edilen değer %90,11 olmuştur. En iyi sonucun elde edildiği komşu sayısı 3, ağırlıklandırma yöntemi distance ve uzaklık yöntemi euclidean olarak tespit edilmiştir.

Yapılan deneyler KNN sınıflayıcıların doğru parametreler ve optimizasyon yardımıyla yüksek sınıflandırma doğruluğu verebileceğini göstermektedir. KNN sınıflayıcılarının performansını artırmada özellikle genetik algoritmaların kullanımının sınıflandırma doğruluğunu artırdığı görülmüştür. Benzetilmiş tavlama yöntemi de sonuçları iyileştirmekle birlikte genetik algoritmalar kadar başarılı olamamıştır. Yapılan çalışmalar neticesinde KNN sınıflayıcılardaki verimlilik düşüşü probleminin doğru parametreler ve optimizasyon teknikleri yardımıyla çözülebileceği görülmüştür. Ayrıca diğer sınıflayıcılar gibi KNN sınıflayıcıların da kalp hastalıklarının teşhisinde önemli bir teşhis yöntemi olduğu görülmüştür.

Kaynaklar (References)

- İnternet: Heart Disease Facts, <https://www.cdc.gov/heartdisease/facts.htm>, 2019.
- Tripoliti E. E., Papadopoulos T. G., Karanasiou G. S., Naka K. K., Fotiadis D. I., Heart Failure: Diagnosis, Severity Estimation and Prediction of Adverse Events Through Machine Learning Techniques, Computational and Structural Biotechnology Journal, 15, 26-47, 2017.
- Solanki Y., A Survey on Risk Assessments of Heart Attack Using Data Mining Approaches, International Journal of Information Engineering and Electronic Business, 11 (4), 2019.
- Rajesh N., Manesha T, Hafeez S., Krishna H., Prediction of Heart Disease Using Machine Learning Algorithms, International Journal of Engineering & Technology, 7 (2.32), 363, 2018.
- Nashif S., Raihan M. R., Islam M. R., & Imam M. H., Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System, World Journal of Engineering and Technology, 6 (4), 2018.
- Karayilan T., Kilic O., Prediction of heart disease using neural network, International Conference on Computer Science and Engineering (UBMK), Antalya, 5-8 Ekim 2017.
- Atkov O. Y., Gorokhova S. G., Sboev A. G., Generozov E. V., Muraseyeva E. V., Moroshkina S. Y., Cherniy N. N., Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters, Journal of Cardiology, 59 (2), 2012.
- Chitra R., Seenivasagam V., Knowledge Discovery from Heart Disease Dataset Using Optimized Neural Network, In: Prasath R., Kathirvalavakumar T. (eds) Mining Intelligence and Knowledge Exploration, Lecture Notes in Computer Science, 8284, Springer, Cham, 2013.
- Yazid M.A., Haikal S., Shukor T., Novi A., Artificial neural network parameter tuning framework for heart disease classification, Proc Electric Eng Comput Sci Inform, 5 (5), 674-679, 2018.
- Karakoyun M., Hacibeyoğlu M., Biyomedikal Veri Kümeleri İle Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatistiksel Olarak Karşılaştırılması, Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 16 (48), 30-42, 2014.
- Mohan S., Chandrasegar T., Gautam S., Effective heart disease prediction using hybrid machine learning techniques, IEEE Access, 7, 81542-81554, 2019.
- Pawlovsky A. P., An ensemble based on distances for a kNN method for heart disease diagnosis, In: 2018 International conference on electronics, information, and communication (ICEIC), IEEE, 1-4, 2018.
- Shouman M., Turner T., Stocker R., Using data mining techniques in heart disease diagnosis and treatment, Japan-Egypt Conference on Electronics, Communications and Computers, 2012.
- Abushariah M. A. M., Alqudah A. A. M., Adwan O. Y., Yousef R. M. M., Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches, Journal of Software Engineering and Applications, 7 (12), 1055-1064, 2014.

15. Malav A., Kadam K., & Kamat P., Prediction of heart disease using k-means and artificial neural network as hybrid approach to improve accuracy, *International Journal of Engineering and Technology*, 9 (4), 3081–3085, 2017.
16. Alzahani S. M., Althopity A., Alghamdi A., Alshehri B., Aljuaid S., An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction, *Lecture Notes on Information Theory*, 2 (4), 2015.
17. Dolatabadi A. D., Khadem S. E. Z., Asl B. M., Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM, *Computer Methods and Programs in Biomedicine*, 138, 117–126, 2017.
18. Miao K. H., Miao H., J., Coronary Heart Disease Diagnosis using Deep Neural Networks, *International Journal of Advanced Computer Science and Applications*, 9 (10), 2018.
19. Amarbayasgalan T., Park K. H., Lee J. Y., Ryu K. H., Reconstruction error based deep neural networks for coronary heart disease risk prediction, *Plos One*, 14 (2), 2019.
20. Wiharto W., Kusnanto H., Herianto H., Intelligence System for Diagnosis Level of Coronary Heart Disease with K-Star Algorithm, *Healthcare Informatics Research*, 22 (1), 30, 2016.
21. Bramer M. A., *Principles of data mining*, Springer, London, 2017.
22. Tan P.-N., Steinbach M., & Kumar V., *Introduction to Data Mining*, Pearson Education, San Francisco, 2006.
23. Doad P. K. and Bartere M. M., A Review : Study of Various Clustering Techniques, *International Journal of Engineering Research & Technology*, 2 (11), 3141-3145, 2013.
24. Cruz-Bernal A., Meta-Heuristic Optimization Techniques and Its Applications in Robotics, *Recent Advances on Meta-Heuristics and Their Application to Real Scenarios*, <https://doi.org/10.5772/54460>, 2013.
25. Akay B., Karaboga D., Parameter tuning for the artificial bee colony algorithm, in: N. T. Nguyen, R. Kowalczyk, S.-M. Chen (Eds.), *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, 608-619, 2009.
26. Weise T., *Global Optimization Algorithms-Theory and Application-Second edition e-book*. <https://pdf-drive.com/global-optimization-algorithms-theory-and-application-free-ebook-2ed/>, 2007.
27. Peterson M. R., Doom T. E., Raymer M. L., GA-Facilitated KNN Classifier Optimization with Varying Similarity Measures. *Proceedings of the 2005 IEEE Congress on Evolutionary Computation*, 3, 2514-2521, 2005.
28. Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., Optimization by Simulated Annealing, *Science, New Series*, 220, 671–680, 1983.
29. Internet: UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml>, 2019.
30. Suganya S. and Tamil Selvi P., A Proficient Heart Disease Prediction Method using Fuzzy-Cart Algorithm, *International Journal of Scientific Engineering and Applied Science*, 2 (1), 1-6, 2016.
31. Dangare S., C., Apte S., Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, *International Journal of Computer Applications*, 47 (10), 44–48. <https://doi.org/10.5120/7228-0076>. 2012.
32. Abdar M., R. Niakan Kalhori S., Sutikno T., Much Ibnu Subroto I. & Arji G., Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases, *International Journal of Electrical and Computer Engineering (IJECE)*, 5 (6), 1569. <https://doi.org/10.11591/ijece.v5i6.pp1569-1576>. 2015.
33. Takci H., Improvement of heart attack prediction by the feature selection methods, *Turkish Journal Of Electrical Engineering & Computer Sciences*, 26, 1–10, 2018.
34. Dwivedi A. K., Performance evaluation of different machine learning techniques for prediction of heart disease, *Neural Computing and Applications*, 29 (10), 2016.
35. Raza K., Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule, in *Healthcare Monitoring Systems*, 179–196, Elsevier, 2019.
36. Bhuvaneeswari R., Sudhakar P., Narmadha R. P., Intelligent Heart Disease Diagnosis System Using Distance Based Outlier Detection With Gradient Tree Boosting Algorithm, *International Journal of Scientific & Technology Research*, 8 (11), 2019.
37. Africa A. D. M., A rough set-based expert system for diagnosing information system communication networks, *International Journal of Information and Communication Technology*, 11 (4), 496, 2017.
38. Jabbar M. A., Prediction of heart disease using k-nearest neighbor and particle swarm optimization, *Biomedical Research*, 28 (9), 4154-4158, 2017.
39. Noor K., Improving The Accuracy Of KNN Classifier For Heart Attack Using Genetic Algorithm, *Journal of University of Kerbala*, 2016.