



# Kahramanmaraş Sütçü İmam University

## Journal of Engineering Sciences



Geliş Tarihi : 04.10.2024  
Kabul Tarihi : 12.12.2024

Received Date : 04.10.2024  
Accepted Date : 12.12.2024

### GENETİK ALGORİTMA TEMELLİ ÖZİNTELİK SEÇİMİNİN ANEMİ HASTALIĞININ TESPİTİNE ETKİSİ

#### THE EFFECT OF GENETIC ALGORITHM BASED FEATURE SELECTION ON DETECTION OF ANAEMIA DISEASE

Mehtap ÖKLÜ<sup>1</sup> (ORCID: 0000-0002-8833-2231)

Hasan BADEM<sup>2\*</sup> (ORCID: 0000-0002-4262-8774)

<sup>1</sup> Kahramanmaraş Sütçü İmam Üniversitesi, Bilişim Sistemleri Bölümü, Kahramanmaraş, Türkiye

<sup>2\*</sup> Kahramanmaraş Sütçü İmam Üniversitesi, Bilgisayar Mühendisliği Bölümü, Kahramanmaraş, Türkiye

\*Sorumlu Yazar / Corresponding Author: Hasan BADEM, [hbadem@ksu.edu.tr](mailto:hbadem@ksu.edu.tr)

#### ÖZET

Vücuttaki oksijen ihtiyacının farklı sebeplerle karşılanamaması durumunda ortaya çıkan anemi, 2023'de Dünya Sağlık Örgütü 500 milyondan fazla kişide görüldüğünü rapor etmiştir. Ayrıca, anemi dünyada en sık görülen kan hastalığıdır. Bu hastalığın en önemli önlemlerinden biri erken teşhistir. Literatürde teşhis konusunda hızlı ve başarılı sonuçların elde edilebilmesi için makine öğrenmesi modelleri önerilmektedir. Ancak makine öğrenmesi modelleri arzu edilen düzeyde etkin sonuçlar veremeyebilir. Optimizasyon algoritmaları ile gerçekleştirilen öznelik seçimi, makine öğrenmesi modellerinin başarı oranlarını arttırabilmektedir. Bu çalışmada hastaların tahlil sonuçları üzerinden gerçekleştirilen anemi tespitinin başarı oranını, öznelik seçici olarak genetik algoritma ile artırılması amaçlanmıştır. K-en yakın komşu, naive bayes, karar ağaçları, lojistik regresyon ve destek vektör makinesi makine öğrenmesi yöntemlerinin sınıflandırma başarımını artırmak için genetik algoritma ile öznelik seçimi gerçekleştirilmiştir. Elde edilen deneysel sonuçlarda önerilen yöntem ile öznelik seçimi yapılmadan elde edilen sonuçlara göre daha yüksek doğruluk oranları elde edilmiştir.

**Anahtar Kelimeler:** Anemi, makine öğrenmesi, genetik algoritma, öznelik seçimi

#### ABSTRACT

Anaemia, which occurs when the body's oxygen needs cannot be met for various reasons, has been reported by the World Health Organization to be seen in more than 500 million people in 2023. Additionally, the anemia is the most common blood disease in the world. One of the most important precautions for this disease is early diagnosis. In the literature, machine learning models have been proposed for achieving rapid and accurate diagnostic results. However, machine learning models may not always provide the desired level. Feature selection utilized with optimization algorithms can enhance the accuracy rates of machine learning models. In this study, improving the accuracy of anemia detection has been aimed on patients' test results through a genetic algorithm as a feature selection operator. Genetic algorithm has been performed as feature selection to improve the classification performance of machine learning methods including k-nearest neighbors, naïve bayes, decision trees, logistic regression, and support vector machine. In the experimental results, higher accuracy rates have been obtained by the proposed method compared with obtained results without feature selection.

**Keywords:** Anaemia, machine learning, genetic algorithm, feature selection

ToCite: BADEM, H., & ÖKLÜ, M., (2025). GENETİK ALGORİTMA TEMELLİ ÖZİNTELİK SEÇİMİNİN ANEMİ HASTALIĞININ TESPİTİNE ETKİSİ. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 28(1), 309-321.

## GİRİŞ

Anemi, diğer adıyla kansızlık, vücuttaki oksijen ihtiyacı duyan dokulara yeterli oksijen taşıyabilecek kırmızı kan hücrelerinin yetersizliği ile ortaya çıkan bir sağlık sorunudur. Dünya genelinde en sık görülen kan hastalığıdır (Vohra vd., 2022). Aneminin en sık görülen türü demir eksikliği anemisidir. Demir anemisinin ortaya çıkma sebebi ise ihtiyaç duyulan demirin besinlerden yeterli seviyede alınamıyor olmasıdır (Yıldız at al., 2021; Seymen, 2014; Cheng at al., 2024).

Dünyada anemi küresel bir sağlık sorunudur. Dünya Sağlık Örgütü (DSÖ) raporlarına göre, 15-49 yaş arası kadınlarda görülme oranı %30 (571 milyon) Gebe kadınlarda %37 ve orta düzey anemiye sahip olan kişi sayısının 1-5 yaş arası çocuklarda görülme oranı ise %40 olduğu rapor edilmiştir (WHO, 2024). Ayrıca, DSÖ'ne göre aneminin önlenmesi, erken müdahale için kritik önem taşımaktadır (WHO, 2014). Anemi hastalığının tespit edilmesi için hemoglobin oranı ve hemoglobinin konsantrasyonu değerlerinin bilinmesi gerekmektedir. Hemoglobinin konsantrasyonunun cinsiyetlere bağlı olarak belirli oranların altında olması gerekmektedir. Bu oranların üstünde kalan bireylerde anemi olduğu kabul edilmektedir (Seymen, 2014; Cheng at al., 2024).

Son zamanlarda karmaşıklaşan dünya problemlerinin çözümü için nümerik yaklaşımla veya sezgisel yaklaşımla birçok farklı türde metod geliştirilmiştir. Gerçek dünya problemlerinin çözümünde yapay zeka yaklaşımları oldukça etkin sonuçlar üretmektedir (Alam vd., 2020; Lambora vd., 2019; Badem, 2019). Probleme özgü getirilen yapay zeka yaklaşımları tek başına arzu edilen başarıyı göstermekte sınırlı kalabilmektedir (Shukla vd., 2015; Gizzi vd., 2022). Yapay zeka modellerinin başarı oranlarını arttırmak için hiperparametre optimizasyonu gerçekleştirilmektedir. Hiperparametre optimizasyonu için yenilikçi yapay zeka destekli olan optimizasyon algoritmaları ortaya çıkmıştır (Badem vd., 2017; Lambora vd., 2019; Khari vd., 2023).

Optimal modelleri elde edilen makine öğrenmesi yöntemlerinin başarımı uzman tasarımı özneliklerden oldukça etkilenmektedir. Bu nedenle uzman tasarımı özneliklerden problemi temsil kabiliyeti yüksek olanların araştırılması da sınıflandırma başarımını artırabilmektedir. Özneliklerin seçilmesi için küresel arama yeteneğine sahip optimizasyon algoritmaları kullanılabilir. Literatürde Genetik Algoritma (Genetic Algorithm, GA) küresel arama ve yerel tüketme yeteneği açısından en etkin algoritmalar arasındadır. GA'nın bu üstün özellikleri sayesinde literatürde pek çok mühendislik probleminde kullanıldığı rapor edilmektedir (Alp ve Soygazi, 2024; Pertiwi vd., 2024).

Anemi hastalığı teşhisi makine öğrenmesi yöntemleri ile gerçekleştirilmesi etkin sonuçlar verebilir. Fakat, uzman tasarımı öznelikler makine öğrenmesi yöntemlerinin başarımını doğrudan etkilemektedir. Bu çalışmada anemi tespiti için GA temelli öznelik seçimi ile K-En Yakın Komşu (K-Nearest Neighbors, KNN), Naive Bayes (NB), Karar Ağaçları (Decision Trees, DT), Lojistik Regresyon (Logistic Regression, LR) ve Destek Vektör Makinesi (Support Vector Machine, SVM) makine öğrenmesi yöntemlerinin performanslarının artırılması amaçlanmıştır. Bu amaç doğrultusunda anemi hastalığının tespitinde öznelik seçiminin başarı oranına olan etkisinin değerlendirilmesi sağlanmıştır. Bu çalışmanın literatüre katkısı:

- Anemi hastalığının tespitinde makine öğrenmesi yöntemlerinin başarımı karşılaştırmalı olarak sunulmuştur.
- Anemi hastalığının tespitinde ilk defa GA tabanlı bir öznelik seçim operatörü uygulanmış ve geleneksel makine öğrenmesi yöntemlerinin başarımına etkisi değerlendirilmiştir.
- Anemi hastalığının tespitinde karar destek sistemlerinin geliştirilebilmesi için yapay zekâ temelli bir yöntem tanımlanmıştır.

Çalışmanın ikinci kısmında literatür taraması, üçüncü kısmında materyal ve metod, dördüncü kısmında deneysel sonuç ve beşinci kısmında ise tartışma bölümlerine yer verilmiştir.

## LİTERATÜR ÖZETİ

DSÖ, insanlarda en sık görülen kan hastalığının anemi olduğunu rapor etmektedir. Kandaki alyuvarların az olması ve oksijen taşıma kapasitesinin az olması anemiye tetikleyen etkenlerdendir (WHO, 2008). İnsanların yaşam kalitesini olumsuz yönde etkileyen bu hastalığın erken teşhisi oldukça önemlidir. Ancak dünya nüfusunun artışı, hastalardaki yaş ve engel durumu öncelikleri gibi zorluklar erken teşhisi geciktirmektedir. Tuba Karagül Yıldız ve arkadaşlarının yaptığı çalışmada aneminin tespit edilmesini sağlayacak bir sistem önerilmiştir. Sistem, Yapay Sinir Ağları (Neural Networks, NN), SVM, NB ve Karar Ağacı olmak üzere dört farklı modelden oluşturulmuştur. Çalışma için hemogram verileri, yaş, cinsiyet, kronik rahatsızlıkların olduğu 25 farklı öznelikten oluşan ve 1663 örnek

bulunduran bir veri seti kullanılmıştır. Veri seti Türkiye'deki bir üniversite hastanesindeki hastaların dosyaları incelenerek toplanmıştır. Oluşturulan modelde tüm verilere ek olarak, bazı özellik seçme teknikleriyle oluşturulan sekiz farklı veri seti kullanılmıştır. Çalışma sonucunda en yüksek doğruluğu %85,6 ile Karar Ağaçları vermiştir (Yıldız vd., 2021).

Çocuklarda sık görülen anemi, Bangladeş'te ciddi bir sağlık sorunu olarak baş göstermektedir. Anemi ile alakalı olan çevresel faktörlerin tespit edilmesinin yanı sıra, bu faktörler göz önünde bulundurularak anemi hastalığına yakalanma olasılığının tahmin edilmesi de oldukça önemlidir. Anemiye yakalanma ihtimalinin tahmini; toplum hizmetleri, sağlık hizmetleri ve kaynak planlaması için de oldukça önemli bir konudur. JR Khan ve arkadaşlarının yaptığı çalışmada Bangladeş Demografi ve Sağlık Araştırması tarafından 2011'de sunulan veriler kullanılarak 5 yaş altı çocuklardaki anemi ihtimalinin tahmin edilmesi amacıyla Doğrusal Diskriminant Analizi, sınıflandırma/regresyon ağaçları, KNN, SVM, Rastgele Orman (Random Forest, RF) ve LR gibi makine öğrenmesi algoritmaları kullanılmıştır. Algoritmaların değerlendirmeleri doğruluk, duyarlılık, özgüllük ve eğri altındaki alan (AUC) metrikleri göz önünde bulundurularak yapılmıştır. RF algoritmasının %70,73 duyarlılık, %66,41 özgüllük ve 0,6857 AUC ile en başarılı sonucu elde ettiği tespit edilmiştir. Anemi tahmininin öncelikli odak noktası olduğu durumlarda klasik regresyon tekniklerine ek olarak makine öğrenmesi yöntemlerinin de dikkate alınabileceği sonucuna varılmıştır (Khan vd., 2019).

Aneminin fiziksel teşhisi, çok sayıda semptomun varlığı nedeniyle zaman alıcı ve yoğun bir işlemdir. Aneminin en hızlı ve en kolay yoldan teşhisi tam kan sayımı ile sağlanmaktadır. Ancak bu yöntem ile farklı anemi türleri doğrudan tanımlanamamaktadır. Bu durumdan dolayı hastadaki anemi tipine yönelik standardı oluşturabilmek için ileri seviye testler gerekmektedir. Bu testler maliyetli olduğu için küçük ölçekte sağlık hizmeti sunan ortamlarda nadir bulunmaktadır. Birden fazla kırmızı kan hücresi formülü ve farklı optimal kesim değerlerine sahip indekslerin varlığına rağmen beta talasemi, demir eksikliği anemisi, hemogloblin E ve kombinasyon anemileri arasında ayırım yapmak oldukça güçtür. Bu durumun önüne geçebilmek Saputra ve arkadaşları tarafından yapılan çalışmada tanımlama sürecini hızlandırmak için daha kesin ve otomatikleştirilmiş bir tahmin modeli önerilmiştir. Veri seti olarak Yogyakarta kentindeki bir laboratuvarından elde edilen geçmiş veriler kullanılmıştır. Uç Öğrenme Makinesi kullanılarak geliştirilen model üzerinde hata matrisi ve dört sınıfı temsil eden 190 veri kullanılarak performans ölçümü izlenmiştir. Model %99,21 doğruluk, %98,44 duyarlılık, %99,30 kesinlik göstermiştir (Saputra vd., 2023).

Özellikle çocukları ve hamile kadınları etkileyen anemi, önemli küresel halk sağlığı sorunlarından biridir. Vücuttaki alyuvar hücrelerinin seviyesinin azalması, yapısının bozulması veya hemogloblin seviyesinin normal eşliğinden düşük olması durumunda anemi ortaya çıkmaktadır. P Appiahene ve arkadaşlarının gerçekleştirdiği çalışmada Evrişimli Sinir Ağı, KNN, NB, SVM ve DT yöntemleri kullanılarak aneminin tespit edilmesi amaçlanmıştır. En yüksek başarılı sonucu %99,96 doğruluk ile NB algoritması sağlarken en düşük başarılı sonucu ise %96,34 ile SVM algoritması sağlamıştır (Appiahene vd., 2023).

Aneminin teşhisi ve tedavisi için birçok kan testi, radyolojik görüntü ve tahlile ihtiyaç duyulur. Bu ölçümler kendi başına yeterli gelmez, sonuçların yorumlanması ve teşhisin koyulabilmesi için doktor kararı gerekir. Tüm bunlara rağmen koyulan teşhisler kesin olamaz, hata payı içerebilir. Yağmur ve arkadaşlarının gerçekleştirdiği çalışmada Kaggle'dan elde edilen bir veri seti ile anemi teşhisinin sınıflandırma problemi ile çözülmesi amaçlanmıştır. Bu doğrultuda NN yöntemlerinden Öğrenmeli Vektör Kuantalama (Learning Vector Quantization , LVQ), Rekabetçi Katman Sinir Ağı, Örüntü Tanıma Yapay Sinir Ağı (Pattern Recognition Artificial Neural Network , PRNN) ve Kendiliğinden Organize Olan Harita kullanılarak başarı oranları hesaplanmıştır. En iyi başarı performansı %100 hassasiyet, %99,88 doğruluk, %99,86 özgüllük, %99,43 kesinlik ile PRNN algoritmasından elde edilmiştir. PRNN'yi takiben LVQ algoritması ise en başarılı ikinci model olmuştur (Yağmur vd., 2023).

Sari ve arkadaşlarının yaptığı çalışmada Parçacık Sürü Optimizasyonu (Particle Swarm Optimization, PSO) ve SVM kümeleri birleştirilerek biyolojik veriler dikkate alınmış ve patolojik verilerin tahmin edilmesi amaçlanmıştır. 539 kişiden oluşan veri seti 5 kümeye bölünmüş ve her bir deneğin değerleri bağımsız olarak alınan ilgili biyomedikal değişkenlere göre kaydedilmiştir. Hibrit PSO-SVM yönteminin, özellikle kümelenebilir hastalık türlerinin teşhisinde oldukça etkili olduğu gözlenmiştir (Arshed vd., 2023).

## MATERYAL VE METOT

Bu çalışmada, anemi hastalığının teşhisinde genetik algoritma temelli öz nitelik seçiminin makine öğrenmesi yöntemleri üzerinden etkisinin değerlendirilmesi amaçlanmıştır. Bu amaç için kullanılan veri seti, kullanılan makine öğrenmesi yöntemleri, genetik algoritma ve önerilen yöntem bu bölümde sunulmuştur.

### Veri Seti

Çalışma kapsamında “Complete Blood Count Anemia Diagnosis” veri seti (Vohra vd., 2022) kullanılmıştır. Bu veri seti, anemi şüphesi taşıyan hastaların kan testi değerlerini içermektedir. İçerdiği öznitelikler Tablo 1’de sunulmuştur. Ayrıca, modelleme de kullanılan veri örneği Tablo 2’de yer almaktadır.

**Tablo 1.** Kan Örneği Parametreleri

| Kısaltma | Açıklama                                 |
|----------|--|
| Age      | Yaş                                      |
| Sex      | Cinsiyet                                 |
| RBC      | Kırmızı Kan Hücre Sayımı                 |
| PCV      | Paketli Hücre Hacmi                      |
| MCV      | Ortalama Hücre Hacmi                     |
| MCH      | Ortalama Hücre Hemoglobini               |
| MCHC     | Ortalama Hücre Hemoglobin Konsantrasyonu |
| RDW      | Kırmızı Hücre Dağıtım Genişliği          |
| TLC      | Beyaz Kan Hücre Sayımı                   |
| PLT      | Trombosit Sayımı                         |
| HGB      | Hemoglobin                               |

Bu veri setinde, her bir örneğin anemi olup olmadığı bilgisine yer verilmemiştir. Etiketleme DSÖ tarafından belirlenen anemi eşik değeri (De Benoist vd., 2008) üzerinden gerçekleştirilmiştir. DSÖ tarafından düzenlenen anemi eşik değerleri Tablo 3’te rapor edilmiştir. Tablo 3 incelendiğinde anemi teşhisi yaş ve cinsiyet durumlarına göre değişebilmektedir. Hastaların tanılarını içeren anemi sütunu da 0 (sağlıklı) ve 1 (anemi) olacak şekilde oluşturulmuştur. Veri setinde yer alan parametrelerin farklı birimlerden oluşmasından dolayı veri setinin normalize edilmesi için Z-skor normalizasyon yöntemi kullanılmıştır.

### Makine Öğrenmesi Yöntemleri

#### K-En Yakın Komşu

KNN yöntemi, gözetimli (denetimli) öğrenme metotlarından olup sınıflandırma işlemlerinde kullanılır (Canbay & Öklü, 2023; Zhang & Li, 2021). Sınıflandırma çalışması yaparken elimizdeki veriler hakkında kısıtlı ön bilgiye sahip olduğunda en çok tercih edilen makine öğrenmesi yöntemi KNN'dir (Canbay & Öklü, 2023; Zhang & Li, 2021). Bu algoritmanın performansını etkileyen parametreler; uzaklık ölçütü, komşu sayısı ( $k$ ) ve ağırlıklandırma yöntemidir (Canbay & Öklü, 2023; Taşcı & Onan, 2016). Çalışmada  $k$  değeri olarak 5, uzaklık ölçütü olarak da Manhattan uzaklığı kullanılmıştır ( $n$  boyutlu düzlemdeki iki konum arasındaki farkların, mutlak değerlerinin toplamı). X-Y konumları arasındaki Manhattan uzaklığı:  $P=(x_1, x_2, \dots, x_n)$  ve  $Q=(y_1, y_2, \dots, y_n)$  olmak üzere, Eşitlik 1’e göre hesaplanır (Taşcı & Onan, 2016; Peterson, 2009).

$$\sum_{i=1}^n |x_i - y_i| \quad (1)$$

#### Karar Ağaçları

DT, veri setini yinelemeli olarak küçük alt bölümlere ayıran bir sınıflandırma yöntemidir. DT, ele aldığı problemi böl ve yönet tekniğiyle çözmeye çalışır. Her bir karar ağacı; kök düğüm, dizi dâhili bölüm ve dizi uç düğümünden (yaprak) oluşur (Canbay & Öklü, 2023). Veri setindeki öznitelikleri mantıksal sıraya uygun dallara ayırarak ağaç yapısı oluşturur. Oluşturulan ağaçta; veri setindeki öznitelikler düğümleri, tahmin edilecek olan sütundaki değerler ise yaprakları temsil eder. Veri setine eklenen her yeni veri, mevcut öznitelik değerleri baz alınarak ağaçtaki uygun

sınıfa dahil edilir. Verinin doğru sınıfa dâhil edilebilmesi için, eklenecek olan verinin ağaç üzerinde gezdirilmesi sağlanır. Bu sayede en uygun sınıf tespit edilir.

**Tablo 2.** Örnek Veri Seti

| Age | Sex | RBC  | PCV  | MCV  | MCH  | MCHC | RDW  | TLC   | PLT   | HGB  | Anemi |
|-----|-----|------|------|------|------|------|------|-------|-------|------|-------|
| 28  | 0   | 5,66 | 34,0 | 60,1 | 17,0 | 28,2 | 20,0 | 11,10 | 128,3 | 9,6  | 1     |
| 41  | 0   | 4,78 | 44,5 | 93,1 | 28,9 | 31,0 | 13,0 | 7,02  | 419,0 | 13,8 | 0     |
| 40  | 1   | 4,65 | 41,6 | 89,5 | 28,8 | 32,2 | 13,0 | 8,09  | 325,0 | 13,4 | 0     |
| 76  | 0   | 4,24 | 36,7 | 86,6 | 26,7 | 30,8 | 14,9 | 13,41 | 254,0 | 11,3 | 1     |
| 20  | 1   | 4,14 | 36,9 | 89,1 | 27,8 | 31,2 | 13,2 | 4,75  | 196,0 | 11,5 | 1     |
| 24  | 0   | 4,29 | 40,1 | 93,5 | 29,6 | 31,7 | 14,5 | 13,96 | 233,0 | 12,7 | 0     |
| 28  | 1   | 4,98 | 42,3 | 84,9 | 24,9 | 29,3 | 16,2 | 9,33  | 213,0 | 12,4 | 1     |
| 14  | 0   | 4,97 | 43,8 | 88,1 | 28,0 | 31,7 | 15,2 | 3,92  | 229,0 | 13,9 | 0     |
| 16  | 0   | 4,16 | 38,7 | 93,0 | 28,8 | 31,0 | 17,9 | 5,77  | 211,0 | 12,0 | 0     |
| 62  | 0   | 5,25 | 45,6 | 86,9 | 25,3 | 29,2 | 15,6 | 10,68 | 151,0 | 13,3 | 0     |

**Tablo 3.** Anemi Teşhisinde Kullanılan Hemoglobin Eşik Değerleri

| Kısaltma                | Açıklama |
|-------------------------|----------|
| Çocuk (5 yaştan küçük)  | 110      |
| Çocuk (5-12 yaş arası)  | 115      |
| Çocuk (12-15 yaş arası) | 120      |
| Kadın (15 yaş ve üstü)  | 120      |
| Hamile kadın            | 110      |
| Erkek (15 yaş ve üstü)  | 130      |

### Lojistik Regresyon

LR, istatistiksel bir sınıflandırma yöntemidir. Parametrik olan LR, en çok kullanılan sınıflandırma yöntemlerinden biridir. Bir veya birden çok değişkene bağımlı olan değişkenlerin modellenmesinde kullanılır (Canbay & Öklü, 2023).

### Naive Bayes

NB, olasılık tabanlı sınıflandırma algoritmalarından birisidir. Bayes teoreminin uygulandığı bu algoritma özellikle denetimli öğrenme problemlerinde başarılı sonuçlar üretmektedir. Algoritma probleme uyarlanırken veri setindeki öznelilikler arasında şartlı bağımsız, sınıflar üretilirken şartlı bağımlı yaklaşım izler. Veri setinde bulunan sınıfların Gaussian dağılımını takip ettiğini varsayarak hatalı tahmin etme oranını azaltmayı amaçlar (Canbay & Öklü, 2023; Narayan, 2021).

### Destek Vektör Makinesi

SVM sınıflandırma problemlerinde sıkça tercih edilen denetimli bir makine öğrenmesi modelidir. SVM, veri setindeki sınıflar arasındaki en yüksek mesafeli hiper düzlemi bulmayı amaçlar (Canbay & Öklü, 2023; Narayan, 2021). Veri setindeki sınıfların örneklerinin arasındaki mesafelere göre sınıfları ayırt etmeyi amaçlar. Ayırt etmeyi düzgün bir şekilde yapabilmek için hiper düzlem(ler) oluşturur. Sınıflandırmadaki hata oranının minimuma indirilmesi için hiper düzlemler arasındaki mesafenin olabilecek maksimum uzaklıkta olması hedeflenir.

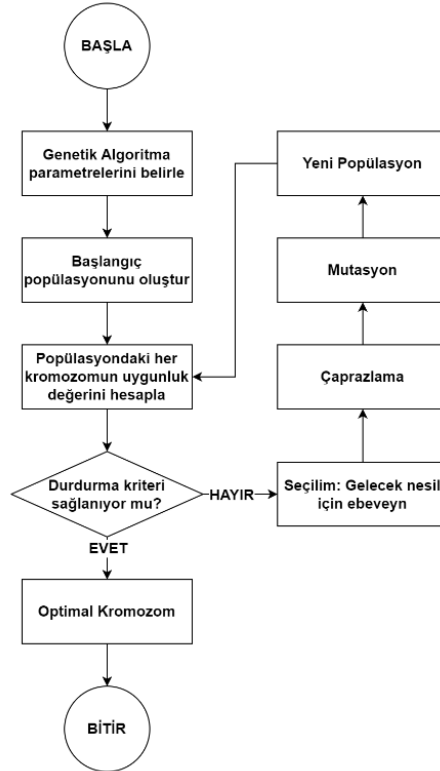
### 10-Kat Çapraz Doğrulama

Bu veri setini 10 eşit parçaya böler ve her parçanın hem eğitim hem de test seti olarak kullanılmasını sağlayarak model performansını değerlendirir. 10 parçanın 9'unu eğitim 1'ini test olarak kullanır. Bu işlemi her 10 parçanın da ayrı ayrı test olarak kullanılmasını sağlayacak şekilde 10 kez tekrar eder (Anguita vd., 2012). Bu şekilde 10-kat

çapraz doğrulama (Cross Validation- CV) tamamlanmış olur. 10-Kat Çapraz Doğrulama yöntemini kullanarak modelin doğrulanması sağlanır.

### Genetik Algoritma

GA, en yaygın kullanılan optimizasyon algoritmalarından biridir (Türkkahraman & Karabulut, 2023; Kaleli, 2023; Acar vd., 2023; Eke, 2022; Weile & Michielssen, 1997; Holland, 1992). Evrim teorisine dayanan GA, evrimin temel süreçlerini baz alan bir çalışma sistemi üzerine geliştirilmiştir. Bu süreçler; seçilim, çaprazlama ve mutasyondur. Üç ana süreçten geçen veriler, süreç boyunca uğradığı etkiler sonucunda bir skora sahip olur. Yüksek skora sahip olan veriler saklanırken düşük skora sahip olan veriler yok edilir. Bu eleme yöntemi, “Güçlü olan hayatta kalır.” felsefesini baz almaktadır (Acar vd., 2023; Eke, 2022; Lambora vd., 2019). GA'nın akış şeması (Gülmez, 2023) Şekil 1'de sunulmuştur.



Şekil 1. GA Akış Diyagramı

GA çalışırken bireylerden (kromozomlar) oluşan bir arama uzayı oluşturur. Her bir birey, vektör olarak temsil edilen bir kromozomdur (Gülmez, 2023). Her bir iterasyonda her bir kromozom incelenir ve uygunluk değeri hesaplanır. Uygunluk değeri; yüksek olan kromozomların hayatta kalma ihtimali yüksek (Keklik & Özcan, 2023), düşük olan kromozomların hayatta kalma ihtimali ise düşük olacaktır (Gülmez, 2023). Uygunluk değeri hesaplandıktan sonra elemelerden başarıyla geçen örnekler seçilim aşamasını tamamlamış ve çaprazlama aşamasına geçmiştir. Birden fazla çaprazlama tekniği bulunmaktadır. Bu teknikler sayesinde örnekler arasında veri (gen) değiş tokuşu sağlanarak çeşitliliğin artması sağlanır. Daha sonra önceden belirlenmiş mutasyon oranı kadar kromozom mutasyona uğratılır (İnneci & Badem, 2023). Bu sayede yavru bireyler elde edilmiş ve bir iterasyon tamamlanmış olur. Yavru bireylerin uygunluk derecesi hesaplanarak karşılaştırılır. Başarılı sayılan bireyler sonraki nesle aktarılırken geri kalan bireyler ise elenir. Bu döngü, durdurma kriteri sağlanıncaya kadar devam eder (Wan, 2024; Gülmez, 2023; Lambora vd., 2019).

### Gen

GA'nın en küçük yapı taşıdır. Nesiller arası kalıtsallığı sağlar. GA'da incelenen her bir karar parametresi gen olarak tanımlanmaktadır. Yani incelenen problemde, gen sayısı ile karar değişkeni sayısı eşittir. Genler aracılığıyla taşınan bilgiler, probleme göre ikilik, ondalık, reel ya da veri kombinasyonu olarak kodlanabilir (Keklik & Özcan, 2023; Katoch vd., 2021; Lambora vd., 2019).

### **Kromozom**

Genler bir araya gelerek kromozomları oluştururlar. Her bir kromozom GA'daki bir bireyi temsil etmektedir. Oluşturulan bu bireyler mevcut problemin bir çözümüne karşılık gelmektedir (Keklik & Özcan, 2023; İnneci & Badem, 2023; Katoch vd., 2021; Lambora vd., 2019 ).

### **Popülasyon**

GA'daki çözüm kümeleri olan popülasyon, birden fazla kromozomun bir araya gelmesiyle oluşur. Optimal sonuca ulaşabilmek için popülasyon içerisindeki bazı kromozomlar popülasyon dışına atılarak yeni kromozomlar üretilir (İnneci & Badem, 2023; Katoch vd., 2021; Lambora vd., 2019).

### **Uygunluk Fonksiyonu**

Amaç (uygunluk) fonksiyonu, bir kromozomun çözüm için uyumlu olup olmadığını tespit eder. Bu sayede kromozomlar arasında rekabet oluşur. Her kromozom için uygunluk fonksiyonu çalışır ve bir skor elde edilir. Kromozomun sonraki nesle aktarılma ve üreme için seçilme ihtimalleri bu skora bağlıdır (Ajder, 2023; Gülmez, 2023, Lambora vd., 2019).

### **Rulet Tekerleği**

Rulet tekerleğinde bireylerin seçilme olasılığını, her bir bireyin uygunluk fonksiyonu değeri ile ilişkilendirerek belirler. Bu yöntemde her bir birey, rulet tekerleğinde bir dilimi temsil eder ve dilimlerin büyüklüğü, bireyin uygunluk fonksiyonu değerine bağlıdır. Yani, daha iyi uygunluk değerine sahip bireyler, rulet tekerleğinde daha geniş bir dilimi kaplar ve dolayısıyla seçilme olasılıkları daha yüksektir (Keklik & Özcan, 2023).

Çaprazlama için eşleştirilecek bireyler, bu rulet tekerleği üzerinde rastgele bir şekilde seçilir. Bu, popülasyon içindeki çeşitliliği korurken, daha iyi uygunluk değerine sahip bireylerin de çaprazlama operatörüne dahil edilmesini ve bu sayede başarı oranlarının yüksek değerlerde kalmasını sağlar. Rulet tekerleği sistemi, çaprazlama operatörünün adil ve çeşitli bir şekilde uygulanmasını sağlar. Bu sistem, bireylerin uygunluk değerlerine dayalı olarak seçilme olasılıklarını dengeler. Genellikle GA'nın performansını artırır (Keklik & Özcan, 2023). Bu çalışmada, çaprazlama operatörünün uygulanması için eş seçme sürecinde rulet tekerleği sistemi tercih edilmiş ve bireylerin eşleştirilmesinde bu yöntem kullanılmıştır.

### **Önerilen Yöntem**

GA temelli öznitelik seçim yöntemi ile anemi hastalığının teşhisi için önerilen yöntem Şekil 2'de sunulmuştur. Önerilen yöntemde, veri setinde bulunan öznitelik sayısı kadar parametreye sahip bireyler için bir başlangıç popülasyonu oluşturulur. Her bir parametre [0-1] arası rastgele değerler ile başlatılır. GA'nın maksimum iterasyonuna ulaşana kadar her birey Şekil 3'te sunulan öznitelik seçimine göre uygunluk değeri hesaplanır. Araştırma durduğunda belirlenen en iyi bireyin temsil ettiği seçim vektörüne göre elde edilen alt veri seti ile makine öğrenmesi ile sınıflandırma gerçekleştirilir. Bu süreçte seçim, çaprazlama ve mutasyon işlemi sırası ile aşağıdaki gibi gerçekleştirilir.

**Seçim:** Yüksek uyumluluk skoruna göre seçilen kromozomlar, birbirleri arasında çaprazlanmak üzere ele alınır (Ajder, 2023; Gülmez, 2023). Başarısız sayılan kromozomlar ise yok edilir. Çalışma kapsamında rulet çemberi seçim tekniği kullanılmıştır.

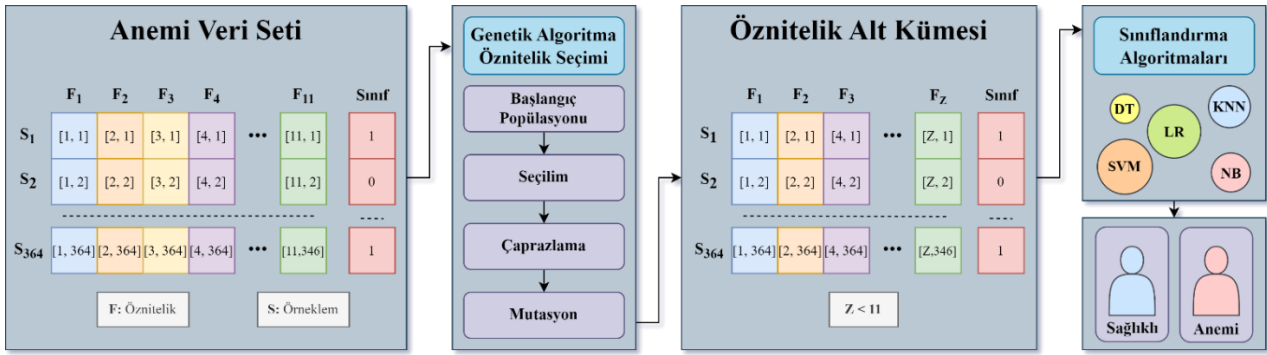
**Çaprazlama:** Seçim aşamasından başarıyla geçen kromozomlar, çeşitlilik oranının artması için birbirleri arasında çaprazlanır. Çaprazlama sonucunda elde edilen veriler, yeni veri olarak havuza eklenir (Ajder, 2023; Acar vd., 2023; Gülmez, 2023). Çalışma kapsamında tek noktali çaprazlama yöntemi kullanılmıştır.

**Mutasyon:** Çaprazlama aşamasında elde edilen yeni bireylerle beraber tüm bireyler ele alınır. Probleme özgü olarak belirlenmiş olan mutasyon oranında, rastgele olarak farklı kromozomlar üzerindeki veriler değiştirilir. Çalışmada Sürünme mutasyonu (Creep mutation) tekniği uygulanmıştır. Bu teknikte, her bir bireyin genetik materyali içerisinde rastgele seçilen gen(ler), mevcut değerlerinden bağımsız olarak tamamen rastgele yeni değerler alacak şekilde değiştirilir. Bu işlem, gen havuzunun çeşitliliğini artırmayı ve algoritmanın yerel minimumlara sıkışmasını önlemeyi amaçlar. Mutasyon sonucunda oluşan kromozomlar, mevcut kromozomların yerine kaydedilir ve yeni

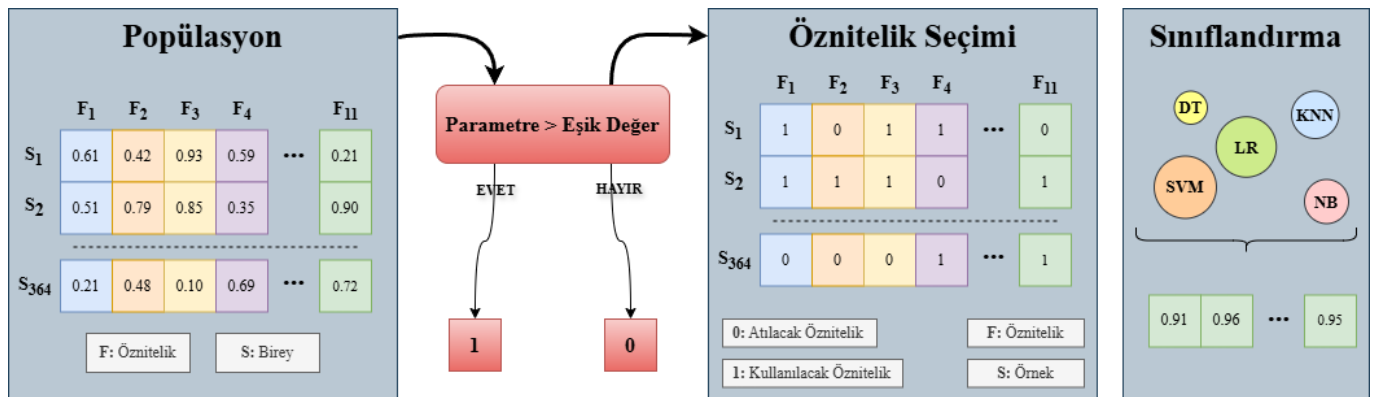
kromozomlar üretilmez. Mutasyon sonucunda uygunluk hedefine ulaşıldığı zaman döngü durur ve en optimal çözüm aranır (Keklik & Özcan, 2023).

**Uygunluk Fonksiyonu:** Makine öğrenmesi ile elde edilen doğruluk skoru, uygunluk fonksiyonu olarak kullanılır. Dolayısıyla Doğruluk değerini maksimize edilmesi amaçlanmıştır

Önerilen yöntemin optimizasyon iş akışı Şekil 2’de, öznelik seçim akışı ise Şekil 3’te sunulmuştur. Popülasyonda seçim vektörü olarak araştırma yapılırken tüm parametreler 0-1 arası reel sayılar olarak tanımlıdır. Seçim işlemi gerçekleştirilirken eşik değer altında kalan öznelikler 0 olarak, üstünde kalanlar ise 1 olarak temsil edilmektedir. Bu sayede veri setinden problemi temsil etme gücü yüksek olanlar belirlenmektedir. Çalışmamızda Eşik değer olarak 0,5 alınmıştır.



Şekil 2. Çalışma Akış Şeması



Şekil 3. Öznelik Seçimi Akış Şeması

## DENEYSEL SONUÇLAR

Önerilen yöntemde KNN, SVM, LR, NB ve DT makine öğrenmesi yöntemleri sınıflandırıcı olarak kullanılmıştır. Her bir yöntem öznelik seçilimi uygulanmadan koşturulmuştur. Ayrıca, bahsedilen makine öğrenmesi yöntemleri için GA ile öznelik seçimi gerçekleştirilerek önerilen yöntemin anemi hastalığı tespitindeki etkisi gözlenmiştir. Önerilen yöntem için kullanılan kontrol parametreleri Tablo 4’te sunulmuştur.

Tablo 4 incelendiğinde, kullanılan yöntemlerin parametre ve değerleri görülmektedir. Genetik algoritmaya özgü kontrol parametreleri sezgisel olarak deneme-yanımla yöntemi ile belirlenmiştir. Popülasyon boyutu 50 olarak belirlenmiştir. Bu değer çeşitliliğin düşük olmamasını sağlarken hesaplama maliyetini de makul seviyede tutmayı sağlamaktadır. Seçilim oranı %20’dir. Bu oran sonraki nesle başarılı bireyler taşınmasını sağlamaktadır. Çaprazlama oranının %20 olarak belirlenmiştir. Bu sayede karışma oranının yüksek seviyelere çıkarak başarılı grupların bozulmasının önüne geçilmesi amaçlanmıştır (Karaboğa 2020). Mutasyon oranı %20 olarak belirlenmiştir. Veri setinde 11 adet parametre bulunmaktadır. Bu nedenle mutasyon etkisinin gözlenebilmesi adına mutasyon oranı literatüre göre yüksek tutulmuştur. Bu sayede mevcut bireylerin değişime uğratılarak algoritmanın daha verimli bir



şekilde çalışmasına olanak sağlayarak dengeli bir araştırma modeli elde edilebilmesi sağlanmıştır. KNN modelinin komşu sayısı 5'tir. SVM modelindeki C parametresi, modelin hata toleransı ve marj genişliği arasındaki dengenin kurulmasını sağlarken LR modelindeki C parametresi ise regularizasyon gücünü kontrol ederek modelin genelleştirme yeteneğinin ayarlanabilmesini sağlar (Mirjalili vd., 2020; Lambora vd., 2019). Tablo 4'teki parametreler tanımlanırken literatürde genel olarak kullanılan oranlar tercih edilmiştir

**Tablo 4.** Simülasyon Parametreleri

| Algoritma | Parametre   | Değer |
|-----------|---|-------|
| GA        | Popülasyon Boyutu                                     | 50    |
|           | Seçilim Oranı   | %20   |
|           | Çaprazlama Oranı                                      | %20   |
|           | Mutasyon Oranı  | %20   |
| KNN       | Komşu Sayısı  | 5     |
| SVM       | C (hata toleransı ve marj genişliği arasındaki denge) | 10    |
| LR        | C (regularizasyon gücü ve genelleştirme yeteneği)     | 10    |

Modelleme sürecinde elde edilen koşma süreleri Tablo 5'te sunulmuştur. Tablo 5 incelendiğinde, Tüm makine öğrenmesi modellerinin GA ile çalışma süreleri görülmektedir. Her bir iterasyonda makine öğrenmesi modellerine giren verilerin aynı olması, rastgelelik durumu ortadan kaldırılarak sağlanmıştır. Bu durumda aynı verilerle çalışan modellerden en hızlı sonuç veren model 104 saniye ile NB olmuştur.

**Tablo 5.** Model Çalışma Süreleri

| Metot | Süre (sn)  |
|-------|------------|
| DT    | 186        |
| KNN   | 180        |
| LR    | 265        |
| NB    | <b>104</b> |
| SVM   | 171        |

**Tablo 6.** Simülasyon Sonuçları

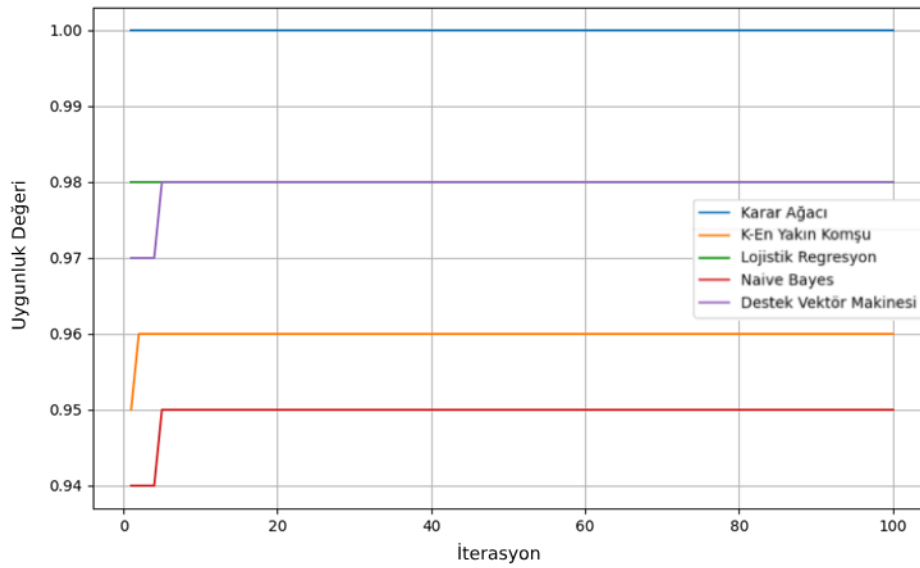
| Algoritmalar | Doğruluk (Accuracy) |         |         |                | Kazanç       |
|--------------|---------------------|---------|---------|----------------|--------------|
|              | Ortalama            | Minimum | Maximum | Standart Sapma |              |
| KNN          | 0,8803              | 0,7188  | 1,0     | 0,0653         | <b>%8,66</b> |
| KNN + GA     | 0,9565              | 0,8485  | 1,0     | 0,0372         |              |
| DT           | 0,9932              | 0,9062  | 1,0     | 0,0149         | %0,34        |
| DT + GA      | <b>0,9966</b>       | 0,9688  | 1,0     | 0,0096         |              |
| LR           | 0,9654              | 0,8750  | 1,0     | 0,0338         | %1,37        |
| LR + GA      | 0,9786              | 0,9062  | 1,0     | 0,0251         |              |
| NB           | 0,9146              | 0,7576  | 1,0     | 0,0517         | %3,59        |
| NB + GA      | 0,9474              | 0,7879  | 1,0     | 0,0411         |              |
| SVM          | 0,9222              | 0,7812  | 1,0     | 0,0527         | %5,88        |
| SVM + GA     | 0,9764              | 0,8788  | 1,0     | 0,0263         |              |

GA ile öznitelik seçimi üzerinden elde edilen doğruluk değerleri ile makine öğrenmesi yöntemlerinin doğruluk değerleri Tablo 6'da karşılaştırmalı olarak sunulmuştur. Tablo 6 incelendiğinde, öznitelik seçimi uygulanmadan makine öğrenmesi modelleri ile 10-Kat Çapraz Doğrulama üzerinden ortalama %88,03 ile %99,32 aralığında doğruluk oranları elde ettiği gözlemlenmiştir. GA desteği olmadan en yüksek skoru %99,32 doğruluk oranıyla DT elde etmiştir. Ardından %96,54 oranıyla LR, ardından %92,22 oranıyla SVM, ardından %91,46 oranıyla NB ve son olarak %88,03 oranıyla KNN takip etmiştir. Önerilen yöntem ile makine öğrenmesi yöntemlerinin başarı oranları ise %0,34-%8,66 aralığında artış gösterdiği görülmektedir. Önerilen yöntem ile en düşük performans artışı %0,34 ile DT

üzerinde etkili olurken, en yüksek performans artışı ise %8,66 ile KNN üzerinde etkili olmuştur. Burada öznelik seçimi olmadan DT ile %99,32 ile oldukça yüksek doğruluk değeri elde edildiğinden performans artışı sağlanabilecek aralığın çok sınırlı kaldığı görülmektedir. Ancak KNN modeli diğer modellere göre daha düşük başarı oranı sağladığından, daha yüksek kazanç oranına sahiptir. Sonuç olarak KNN + GA modeli, GA ile öznelik seçimi sayesinde rakip modeller arasında en yüksek başarıya sahip olduğu görülmektedir.

Tablo 6 dan görüldüğü üzere en başarılı model %99,66 oranıyla önerilen yöntem üzerinden DT+GA olmuştur. Bu modelde öznelik seçimi sonucunda seçilen parametreler **yaş**, **cinsiyet**, **PCV**, **MCHC** ve **hemoglobin** olduğu görülmüştür.

Optimizasyon algoritmalarının performans değerlendirilmesinde, yakınsama grafikleri de oldukça önemlidir. Önerilen yöntemde kullanılan her bir makine öğrenmesi yöntemleri için elde edilen yakınsama grafikleri de Şekil 4'te sunulmuştur. Ayrıca önerilen yöntemin 10 kat çapraz doğrulama üzerinden elde edilen simülasyon sonucunda Keskinlik, Geri çağırma, F1 skor ve Özgüllük metrikleri için sonuçlar da Tablo 7'de yer almaktadır.



Şekil 4. Genetik Algoritma Yakınsama Grafiği

Tablo 7. Simülasyon Metrikleri

| Algoritmalar | Ortalama              |                       |               |                        |
|--------------|-----------------------|-----------------------|---------------|------------------------|
|              | Keskinlik (Precision) | Geri Çağırma (Recall) | F1            | Özgüllük (Specificity) |
| KNN+GA       | 0,9082                | 0,9019                | 0,9033        | 0,8909                 |
| DT+GA        | 0,9286                | <b>0,9301</b>         | <b>0,9278</b> | 0,9159                 |
| LR+GA        | <b>0,9344</b>         | 0,9235                | 0,9265        | 0,9193                 |
| NB+GA        | 0,9001                | 0,8756                | 0,8831        | 0,8887                 |
| SVM+GA       | 0,9297                | 0,9204                | 0,9231        | <b>0,9172</b>          |

Şekil 4, GA'nın çeşitli makine öğrenmesi modelleri üzerindeki yakınsama sürecini göstermektedir. Bu grafikte beş farklı modelin genetik algoritma tarafından optimize edilen uygunluk değerlerinin iterasyon sayısına göre değişimi izlenmektedir. DT modeli, iterasyonların başında neredeyse %100'lük bir uygunluk değerine ulaşarak en yüksek performansı sergilemiş ve bu değeri süreç boyunca korumuştur. LR, %98'e yakın bir uygunluk değeri ile ikinci sırada yer alırken, SVM ise %97-%98 aralığında üçüncü sırada yer almaktadır. KNN modeli %95-%96 düzeylerinde sabitlenmiştir. NB ise diğer modellere kıyasla daha düşük bir performans göstermiş ve %94-%95 seviyesinde kalmıştır. Her modelin belirli bir iterasyon sonrasında uygunluk değerlerinde durağanlaşma gözlemlenmiş, bu da genetik algoritmanın modeller üzerinde optimal çözüme yakınsadığını ve daha fazla gelişim göstermediğini işaret etmektedir. Özellikle DT'nin yüksek performansı, bu modelin genetik algoritma ile en iyi uyumu gösterdiğini ortaya koymaktadır.

Tablo 7 incelendiğinde ise, kesinlik açısından en yüksek değeri LR üzerinden alırken, geri çağırma ve F1-skor için ise en başarılı makine öğrenmesi yöntemi DT olduğu görülmektedir. Aynı zamanda SVM'nin özgüllük metriğinde en başarılı yöntem olduğu görülmektedir. DT'nin doğruluk üzerinde sağlamış olduğu başarıda da dikkate alındığında anemi hastalığı tespitinde daha etkin olabileceği değerlendirilebilir.

## SONUÇ

En yaygın kan hastalıklarından olan aneminin teşhis edilebilmesinde makine öğrenmesinin etkisi oldukça önemli bir araştırma alanıdır. Bu çalışmada makine öğrenmesi ile gerçekleştirilen sınıflandırma başarımı, GA tabanlı öznelik seçim yöntemi ile artırılması amaçlanmıştır. Literatürde yaygın olarak kullanılan KNN, SVM, LR, NB ve DT makine öğrenmesi yöntemleri temel sınıflandırıcı olarak kullanılmıştır. Elde edilen deneysel sonuçlar incelendiğinde; GA öznelik seçimi ile elde edilen sonuçların, seçim olmadan elde edilen sonuçlara göre daha başarılı sonuçlar verdiği gözlemlenmiştir. Başlangıç popülasyonunun, çaprazlama oranının, mutasyon oranının, seçim metotlarının ve makine öğrenmesi modelinin değiştirilmesi ile başarının daha yüksek oranlara çıkabileceği öngörülmektedir. Ayrıca, anemi hastalığının tespitinde literatüre sunulmuş yeni meta sezgisel optimizasyon algoritmalarının başarısının araştırılması daha etkin sonuçların elde edilmesine olanak sağlayacağı öngörülebilir.

## KAYNAKLAR

- Acar, E. B., Karabey, C., & Köse, B. İnsansız Hava Aracı İle Paket Dağıtımında Gezgin Satıcı Probleminin Genetik Ve Parçacık Sürü Optimizasyon Algoritmaları İle Çözümü. Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi, 10(20), 168-181. <https://doi.org/10.54365/adyumbd.1249391>
- Ahmad, A., Alzaidi, K., Sari, M., & Uslu, H. (2023). Prediction of anemia with a particle swarm optimization-based approach. *An International Journal of Optimization and Control: Theories & Applications (IJOCTA)*, 13(2). 10.11121/ijocta.2023.1269
- Ajder, A. (2023). Geleneksel Yaklaşımlar: Parçacık Sürü Optimizasyonu Algoritması. *Teknobilim-2023: Yapay Zeka ve Mühendislik*, 21.
- Alam, T., Qamar, S., Dixit, A., & Benaida, M. (2020). Genetic algorithm: Reviews, implementations, and applications. *arXiv preprint arXiv:2007.12673*. <https://doi.org/10.48550/arXiv.2007.12673>
- Alp, G., & Soygazi, F. (2024, May). Meta-Heuristic Supported Feature Selection in Classification Algorithms for Diabetes Diagnosis. In *2024 32nd Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. 10.1109/SIU61531.2024.10601062
- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012, April). The K'in K-fold Cross Validation. In *ESANN* (Vol. 102, pp. 441-446).
- Appiahene, P., Asare, J. W., Donkoh, E. T., Dimauro, G., & Maglietta, R. (2023). Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. *BioData mining*, 16(1), 2. <https://doi.org/10.1186/s13040-023-00319-z>
- Badem, H., Basturk, A., Caliskan, A., & Yuksel, M. E. (2017). A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited-memory BFGS optimization algorithms. *Neurocomputing*, 266, 506-526. <https://doi.org/10.1016/j.neucom.2017.05.061>
- Badem, H. (2019). Parkinson Hastalığının Ses Sinyalleri Üzerinden Makine Öğrenmesi Teknikleri ile Tanımlanması. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 8(2), 630-637. 10.28948/ngumuh.524658
- Cheng, L., Nie, Y., Wen, H., Li, Y., Zhao, Y., Zhang, Q., ... & Fu, S. (2024). An ensemble machine learning model for predicting one-year mortality in elderly coronary heart disease patients with anemia. *Journal of Big Data*, 11(1), 1-20. <https://doi.org/10.1186/s40537-024-00966-x>
- Eke, İ. (2022). Optimum PID Kazançları Genetik Algoritma İle Hesaplanan Otomatik Gerilim Regülatörü. *International Journal of Engineering Research and Development*, 14(3), 351-361. <https://doi.org/10.29137/umagd.1176936>
- Gizzi, E., Nair, L., Chernova, S., & Sinapov, J. (2022). Creative problem solving in artificially intelligent agents: A survey and framework. *Journal of Artificial Intelligence Research*, 75, 857-911. <https://doi.org/10.1613/jair.1.13864>

- Gülmez, B. (2023). Market zinciri ürün dağıtım problemi için farklı genetik algoritma versiyonları ile çözümü ve karşılaştırması. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 6(1), 180-196.
- Holland, J. H. (1992). Genetic algorithms. *Scientific american*, 267(1), 66-73.
- Inneci, T., & Badem, H. (2023). Detection of Corneal Ulcer Using a Genetic Algorithm-Based Image Selection and Residual Neural Network. *Bioengineering*, 10(6), 639. <https://doi.org/10.3390/bioengineering10060639>
- Kaleli, S. S. (2023). Bist-30 Şirketlerinin Pandemi Öncesi-Sonrası Satış Verilerinin Genetik Algoritma ile Analizi ve Optimum Portföy Oluşturma. *MANAS Sosyal Araştırmalar Dergisi*, 12(2), 557-565. <https://doi.org/10.33206/mjss.1215054>
- Karaboğa, D. (2020). Yapay zeka optimizasyon algoritmaları (7.Baskı). Nobel Akademi Yayıncılık.
- Katoch, S., Chauhan, S. S., & Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia tools and applications*, 80, 8091-8126. <https://doi.org/10.1007/s11042-020-10139-6>
- Keklik, G., & Özcan, B. D. (2023). Genetik Algoritmaların İşleyişi ve Genetik Algoritma Uygulamalarında Kullanılan Operatörler. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 6(1), 1052-1066. <https://doi.org/10.47495/okufbed.1161413>
- Khan, J. R., Chowdhury, S., Islam, H., & Raheem, E. (2019). Machine learning algorithms to predict the childhood anemia in Bangladesh. *Journal of Data Science*, 17(1), 195-218. 10.6339/JDS.201901\_17(1).0009
- Lambora, A., Gupta, K., & Chopra, K. (2019, February). Genetic algorithm-A literature review. In 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 380-384). IEEE, 10.1109/COMITCon.2019.8862255
- Mirjalili, S., Song Dong, J., Sadiq, A. S., & Faris, H. (2020). Genetic algorithm: Theory, literature review, and application in image reconstruction. *Nature-inspired optimizers: Theories, literature reviews and applications*, 69-85. [https://doi.org/10.1007/978-3-030-12127-3\\_5](https://doi.org/10.1007/978-3-030-12127-3_5)
- Narayan, Y. (2021). Comparative analysis of SVM and Naive Bayes classifier for the SEMG signal classification. *Materials Today: Proceedings*, 37, 3241-3245. <https://doi.org/10.1016/j.matpr.2020.09.093>
- Öklü, M., & Canbay, P. (2023). Makine Öğrenmesi Yöntemleri ile Şehirlerin Hava Kalitesi Tahmini. *International Journal of Advances in Engineering and Pure Sciences*, 35(1), 39-53. <https://doi.org/10.7240/jeps.1175507>
- Pertiwi, D. A. A., Ahmad, K., Salahudin, S. N., Annegrat, A. M., & Muslim, M. A. (2024). Using genetic algorithm feature selection to optimize XGBoost performance in Australian credit. *Journal of Soft Computing Exploration*, 5(1), 92-98. <https://doi.org/10.52465/josce.v5i1.302>
- Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883. 10.4249/scholarpedia.1883
- Saputra, D. C. E., Sunat, K., & Ratnaningsih, T. (2023, February). A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia. In *Healthcare* (Vol. 11, No. 5, p. 697). MDPI. <https://doi.org/10.3390/healthcare11050697>
- Seymen, V. (2014). Demir eksikliği anemisi hastalığının tespitinde kullanılan sınıflandırma algoritmalarının karşılaştırılması (Master's thesis, Sakarya Üniversitesi (Turkey)).
- Shukla, A., Pandey, H. M., & Mehrotra, D. (2015, February). Comparative review of selection techniques in genetic algorithm. In 2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE) (pp. 515-519). IEEE. 10.1109/ABLAZE.2015.7154916
- Taşcı, E., & Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. *Akademik Bilişim*, 1(1), 4-18.
- Türkkahraman, Ş. M., & Karabulut, K. (2023). Sosyal Ağ Varlığında Takım Oluşturma Problemine Hibrit Bir Genetik Algoritma Önerisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 25(73), 181-192. <https://doi.org/10.21205/deufmd.2023257315>
- Vohra, R., Dudyala, A. K., Pahareeya, J., & Hussain, A. (2022). Decision rules generation using decision tree classifier and their optimization for anemia classification. In *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021* (pp. 721-737). Singapore: Springer Nature Singapore. [https://doi.org/10.1007/978-981-16-6723-7\\_53](https://doi.org/10.1007/978-981-16-6723-7_53)

- Wan, H. (2024). Applying the genetic algorithm to optimization problems. *WIT Transactions on Information and Communication Technologies*, 2. 10.2495/AIENG930332
- Weile, D. S., & Michielssen, E. (1997). Genetic algorithm optimization applied to electromagnetics: A review. *IEEE Transactions on Antennas and Propagation*, 45(3), 343-353. 10.1109/8.558650
- World Health Organization (WHO) (2008). Worldwide prevalence of anaemia 1993-2005: WHO global database on anaemia.
- World Health Organization (WHO). (2014). Global nutrition targets 2025: Stunting policy brief (No. WHO/NMH/NHD/14.3).
- World Health Organization (WHO). (2024). Guideline on haemoglobin cutoffs to define anaemia in individuals and populations. World Health Organization.
- Yağmur, N., Temurtaş, H., & Dağ, İ. (2023). Anemi Hastalığının Yapay Sinir Ağları Yöntemleri Kullanılarak Sınıflandırılması. *Journal of Scientific Reports-B*, (008), 20-34.
- Yıldız, T. K., Yurtay, N., & Öneç, B. (2021). Classifying anemia types using artificial learning methods. *Engineering Science and Technology, an International Journal*, 24(1), 50-70. <https://doi.org/10.1016/j.jestch.2020.12.003>
- Zhang, S., & Li, J. (2021). KNN classification with one-step computation. *IEEE Transactions on Knowledge and Data Engineering*, 35(3), 2711-2723. 10.1109/TKDE.2021.3119140