



DOĞA İLHAMLI OPTİMİZASYON KULLANARAK ÖZELLİK SEÇİMİ VE MAKİNE ÖĞRENMESİ İLE MEME KANSERİ TEŞHİSİ

Onur SEVLİ*

Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü,
Burdur, Türkiye

Anahtar Kelimeler

*Meme Kanseri Teşhisi,
Doğa İlhamlı Optimizasyon,
Özellik Seçimi,
Makine Öğrenmesi.*

Öz

Meme kanseri kadınlarda en yaygın görülen kanser türü ve en çok ölüme sebep olan hastalıktır. Erken teşhis ve tedavi edilirse iyileşme ve hayatta kalma olasılığı yükselmektedir. Makine öğrenmesi, medikal alanda farklı uygulamalarıyla kanser türlerinin erken teşhisinde uzmanlar için başarılı bir karar destek sistemi rolü üstlenmektedir. Bir hastalığın tanısı için toplanan çok sayıda özelliğin tümünün birlikte kullanımını analiz süreci ve başarısını olumsuz etkileyebilmektedir. Toplanan verilerden en etkin özniteliklerin seçilmesi ve bunları kullanarak tahminlemeler yapılması teşhis başarısını artırabilmektedir. Bu çalışmada meme kanseri konusunda literatürde yaygın kullanılan UCI WDBC veri seti üzerinde KNN, RO ve DVM algoritmaları ile öznitelik seçimi olmadan ve öznitelik seçimi uygulanarak sınıflandırmalar gerçekleştirilmiştir. Öznitelik seçimi konusunda başarılı sonuçlar sağlayan doğa ilhamlı algoritmalarından Guguk Kuşu Arama (GKA), Parçacık Sürü Optimizasyonu (PSO), Balina Optimizasyon (BO) algoritması ve Kızıl Geyik (KG) algoritması kullanılarak ve bu algoritmaların 25, 50 ve 75 parçacıklı sürüleri ile 30 öznitelikli orijinal WDBC veri seti üzerinde öznitelik seçimleri gerçekleştirilmiştir. En yüksek doğruluk 75 parçacıklı GKA ile seçilen 16 adet öznitelik ile RO sınıflandırıcı kullanılarak %99.12 olarak elde edilmiştir. Öznitelik seçimi yapılarak gerçekleştirilen sınıflandırmalardaki doğruluklarının, seçim yapılmadan elde edilen sonuçlardan yüksek olduğu görülmüştür. Eldeki bulgular literatürdeki çalışmalar ile karşılaştırılmış ve daha yüksek başarı sağladığı gözlenmiştir.

BREAST CANCER DIAGNOSIS WITH FEATURE SELECTION USING NATURE-INSPIRED OPTIMIZATION AND MACHINE LEARNING

Keywords

*Breast Cancer Diagnosis,
Nature-Inspired Optimization,
Feature Selection,
Machine Learning.*

Abstract

Breast cancer is the most common type of cancer in women and is the leading cause of death. If diagnosed and treated early, the rate of recovery and survival increases. Machine learning, with different applications in the medical field, plays a successful decision support role for experts in the early diagnosis of cancer types. Using all of the many features collected to diagnose a disease can negatively affect the analysis process and success. Selecting the most effective features from the collected data and making predictions using them can increase the success of the diagnosis. In this study, classifications were carried out on the UCI WDBC dataset, which is widely used in the literature on breast cancer, with KNN, RF, and SVM, without feature selection and by applying feature selection. Feature selection operations were performed on the original WDBC dataset containing 30 features, using nature-inspired algorithms Cuckoo Search (CS), Particle Swarm Optimization (PSO), Whale Optimization (WOA), and Red Deer Algorithm (RDA) consisting of 25, 50, and 75 particles. The highest accuracy was obtained with the RO classifier as 99.12%, using 16 features selected with 75-particle CS. It was observed that the accuracy of the classifications made by feature selection was higher than the results obtained without feature selection. The findings were compared with the current studies in the literature and it was observed that it provided higher success.

* İlgili yazar / Corresponding author: onursevli@mehmetakif.edu.tr, +90-248-213-4130

Alıntı / Cite

Sevli, O. (2022). Doğa İlhamlı Optimizasyon Kullanarak Özellik Seçimi ve Makine Öğrenmesi ile Meme Kanseri Teşhisi, Mühendislik Bilimleri ve Tasarım Dergisi, 10(2), 442-452.

Yazar Kimliği / Author ID (ORCID Number)

O. Sevli, 0000-0002-8933-8395

Makale Süreci / Article Process

Başvuru Tarihi / Submission Date	14.11.2021
Revizyon Tarihi / Revision Date	16.01.2022
Kabul Tarihi / Accepted Date	24.01.2022
Yayın Tarihi / Published Date	30.06.2022

1. Giriş (Introduction)

Kanser, hücrelerin büyümesini düzenlemekten sorumlu genlerdeki mutasyonlar veya anormal değişikliklerin sonucu olarak ortaya çıkar. Normal şartlarda vücuttaki hücreler düzenli hücre bölünmeleri ile yenilenirler. Yaşlanan hücreler ölürken yerlerini yeni hücreler alır. Ancak oluşan anormallikler sebebiyle değişen hücre kontrolsüz ve düzensiz olarak bölünmeye devam ederek tümör oluşturma potansiyeli kazanır. Tümörler iyi huylu (benign) veya kötü huylu (malignant) olabilir. İyi huylu tümörler yavaş büyürler, yakınındaki dokuları istila etmez ve vücudun diğer bölgelerine yayılmazlar. Bu nedenle iyi huylu tümörler kanserli olarak kabul edilmezler. Kötü huylu tümörler ise kanserlidir ve kontrol altına alınmazlarsa hızla çoğalarak vücudun diğer bölgelerine yayılabilirler.

Meme tümörleri, meme dokusunda özellikle süt kanalları ve bezlerindeki hücrelerin olağandışı bölünmesi sonucu büyüyen ve farklılaşan kitleler şeklinde kendini gösterir (Jerez-Aragonés vd., 2003). Kitlenin pürüzsüz yapıda olması ve sınırlarının belirli olması iyi huylu olduğunu, düzensiz sınırlara sahip olması ve pürüzlü olması ise kötü huylu olduğunu, yani kanser riski taşıdığını gösterir (Papageorgiou vd., 2015). Meme kanseri kadınlarda en yaygın görülen kanser türü ve en yüksek oranda ölüme sebep olan hastalıktır (Shrivastavat vd., 2013). Son dönemdeki istatistikler meme kanserinin %11.7'lik görülme oranıyla akciğer kanserini geride bıraktığını ve kanser kaynaklı ölümlerin genelinde %6.9'luk bir paya sahip olduğunu göstermektedir (Sung vd., 2021). Diğer tüm kanser türlerinde olduğu gibi meme kanserinin de erken dönemde teşhis ve tedavi edilmesi, iyileşme ve hayatta kalma olasılığını yükseltir. Ancak düzenli ve periyodik kontrollerin yapılmaması halinde çoğu hastaya ileri safhalarda tanı konulabilmektedir. Meme kanseri tanısı uzmanlık gerektirmekte fakat uzmanlar tarafından manuel olarak gerçekleştirilen testler bazen yanlış sonuçlar verebilmektedir (Kumar & Bhargava, 2019). Bu durum uzmanların sağlıklı kararlar vermelerine yardımcı olacak otomatik karar destek sistemlerini gündeme getirmektedir (Sinha vd., 2020).

Meme kanseri tanısında konsültasyon raporları, histopatolojik görüntüler, kan testi sonuçları, mamografik görüntüler gibi çok sayıda veri elde edilir. Bu veriler doğru kullanıldığında bir destek sistemi rolü üstlenecek güçlü tahmin modelleri oluşturulabilir. Yapay zekanın bir alt dalı olan makine öğrenmesi, verilerin analiz edilmesi ve en iyi tanılama için veriler arasındaki ilişkilerin ortaya çıkarılmasını sağlar (Bayrak vd., 2019). Makine öğrenmesi alanındaki sınıflandırma teknikleri, meme kanseri gibi kanser türlerinin nispeten erken aşamada, kısa sürede ve etkin şekilde tespitinde de kullanılabilir.

Makine öğrenmesi teknikleri ile meme kanseri tespiti konusunda Wisconsin Breast Cancer Diagnostic (WDBC) veri seti literatürdeki çalışmalarda yaygın olarak kullanılmaktadır. Bu veri seti meme dokusundaki kitleleri tespit etmek için ince iğne aspirasyon (fine-needle aspiration / FNA) biyopsisi tekniği ile elde edilen görüntülerin sayısallaştırılmış özelliklerinden oluşmaktadır. Genel olarak bu veri seti üzerinde, farklı yöntemler kullanılarak gerçekleştirilen son dönemdeki çalışmalar şu şekilde özetlenebilir:

Khorshid vd. (2021) Matlab kullanarak, 5 farklı makine öğrenmesi tekniği ile meme tümörü tespitine yönelik bir çalışma gerçekleştirmişlerdir. Destek Vektör Makinesi (DVM), Lojistik Regresyon (LR), K-En yakın Komşu (KNN), Ağırlıklandırılmış KNN(W-KNN) ve Gaussian Naïve Bayes (Gaussian NB) algoritmaları kullanarak gerçekleştirdikleri sınıflandırmalarda en yüksek doğruluğu %96.7 olarak W-KNN ile elde etmişlerdir. Ahuja vd. (2021) ise veri seti üzerinde PCA ile boyut indirgeme yapıp daha sonra aykırı değerlerin temizlenmesinin ardından gerçekleştirdikleri sınıflandırmalarda W-KNN yöntemiyle %98.9 doğruluğa ulaşmışlardır. Vijayakumar vd. (2021) derin sinir ağları (DNN) ile farklı aktivasyon fonksiyonları ve optimize ediciler kullanarak sınıflandırma çalışmaları gerçekleştirmişlerdir. ReLU aktivasyon fonksiyonu ve Adagrad optimize edici kullanarak gerçekleştirdikleri sınıflandırma çalışmasında %97.94 doğruluk değerine ulaşmışlardır. Feroz vd. (2021) farklı makine öğrenmesi algoritmaları kullanarak gerçekleştirdikleri karşılaştırmalı sınıflandırmalarda KNN ve Rastgele Orman (RO) yöntemleri ile %97.14 doğruluk değerine ulaşmışlardır. Muhammad Amin ve Inna (2021) ise Lojistik Regresyon (LR) yöntemi ile %93.9 doğruluk elde etmişlerdir. Jabbar (2021) Bayes ağı ve Radial Basis Function (RBF) birleşimi ile oluşturduğu kolektif öğrenme (ensemble learning) algoritmasını kullanarak %97 doğruluk değerine ulaşmıştır.

Assegie (2021) KNN algoritmasının hiperparametrelerinin varsayılan değerlerini kullanarak ve ayrıca hiperparametre optimizasyonu yaparak gerçekleştirdiği iki ayrı sınıflandırmada sırayla %90.10 ve %94.35 doğruluk elde etmiştir. Gupta ve Kaushik (2021) KNN, LR, RO, DVM ve sinir ağları (NN) algoritmalarının birleşimi ile oluşturdukları kolektif öğrenme tekniği ile gerçekleştirdikleri sınıflandırmada %97.9 doğruluğa ulaşmışlardır. Ghosh vd. (2021) bir derin öğrenme tekniği olan Uzun Kısa Süreli Bellek (LSTM) modeli kullanarak gerçekleştirdikleri sınıflandırmada %99.1 doğruluk elde etmişlerdir. Lahoura vd. (2021) KNN, Naïve Bayes (NB) ve DVM algoritmaları kullanarak sırası ile %90.64, %84.80, %92.98 sınıflandırma doğruluğuna ulaşmışlardır.

Pravesjit vd. (2021) veri seti üzerinde Parçacık Sürü Optimizasyon (PSO) algoritması ve NB sınıflandırıcı kullanarak gerçekleştirdikleri sınıflandırmada %97.27 doğruluk değerine ulaşmışlardır. Jijitha ve Amudha (2021) ise Genetik Algoritmalar (GA) ile öznitelik seçimi yaparak ve öznitelik seçimi olmaksızın sınıflandırmalar gerçekleştirmişlerdir. Öznitelik seçimi olmadan LR ile %96.5, KNN ile %79.02 doğruluk elde ederken, öznitelik seçimi uygulandığı durumda doğruluk değerleri %98.24 ve %95.1'e yükselmiştir. Devi vd. (2021) veri setinde öznitelik seçimi uygulayarak elde ettikleri 16 elemanlı alt set üzerinde LR ve NN'den oluşan kolektif sınıflandırma algoritması kullanarak %98.5 doğruluk elde etmişlerdir.

Parekh ve Dahiya (2021) DVM ve NN kullanarak gerçekleştirdikleri karşılaştırmalı sınıflandırma çalışmasında %96.48 ve %96.04 doğruluk elde etmişlerdir. Naji vd. (2021), veri setinden seçtikleri özellikler üzerinde DVM, RO, LR, Karar Ağacı (KA) ve KNN kullanarak gerçekleştirdikleri karşılaştırmalı sınıflandırma çalışmasında %93.2 ile %97.2 arasında değişen doğruluk değerleri elde etmişlerdir. %97.2 oranla en yüksek sonucu DVM algoritması sağlamıştır. Magboo ve Magboo (2021) veriler üzerinde ölçekleme, PCA ve yeniden örnekleme yöntemleri ile birlikte LR, NB, KNN ve DVM algoritmaları kullanarak sınıflandırmalar gerçekleştirmişlerdir. En yüksek doğruluğu %80 olarak LR sınıflandırıcı ile birlikte PCA kullandıkları durumda elde etmişlerdir. Al-Azzam, ve Shatnawi (2021), meme kanseri tespitinde danışmanlı ve yarı danışmanlı makine öğrenmesi tekniklerinin performanslarını karşılaştıran bir çalışma gerçekleştirmişlerdir. Danışmanlı öğrenmede KNN ve yarı danışmanlı öğrenmede LR modeli %98 ile en yüksek doğruluğu sağlamıştır. Pawar vd. (2021) veri seti üzerinde özelliklerin ölçeklendirilmesinin ardından DVM, KNN, KA, RO, AdaBoost ve XGBoost olmak üzere altı farklı sınıflandırıcı ile karşılaştırmalı bir çalışma gerçekleştirmişlerdir. En yüksek doğruluk değeri %98.24 olarak XGBoost sınıflandırıcı ile elde edilmiştir. Al-Yaseen vd. (2021) veri setini %75 eğitim ve %25 test seti olarak ayırmışlar, eğitim seti üzerinde KMeans kullanarak gerçekleştirdikleri kümeleme ile özellikleri indirgeyerek eğitim süresini kısaltmayı hedeflemişlerdir. İndirgenen veriler üzerinde DVM kullanarak gerçekleştirdikleri sınıflandırmada % 98.067 doğruluk elde etmişlerdir. Ara vd. (2021) öznitelik seçiminin ardından LR, KNN, KA, NB, RO ve DVM algoritmaları kullanarak gerçekleştirdikleri sınıflandırmalarda en yüksek doğruluğu %96.5 olarak DVM ile elde etmişlerdir.

Bu çalışmada WDBC veri seti üzerinde farklı makine öğrenmesi teknikleri kullanılarak meme kanseri durumunun tespitine yönelik sınıflandırmalar gerçekleştirilmiştir. Orijinal veri seti üzerinde gerçekleştirilen sınıflandırmalar yanında sınıflandırma öncesinde 4 farklı Doğa ilhamlı Optimizasyon Algoritması (DOA) kullanılarak öznitelik seçim işlemi uygulanmıştır. Her bir sınıflandırıcının performansı öznitelik seçimi olmaksızın ve öznitelik seçimi uygulanarak raporlanmıştır. Elde edilen bulgular aynı veri setini kullanan ve son dönemde yapılan literatürdeki güncel çalışmalar ile karşılaştırılmıştır.

Devam eden bölümlerde kullanılan veri setinin özellikleri, kullanılan sınıflandırma ve optimizasyon yöntemleri, araştırma bulguları ve karşılaştırma ile sonuç bölümleri yer almaktadır.

2. Materyal ve Yöntem (Material and Method)

2.1. Veri Seti (Dataset)

Çalışmada kullanılan Wisconsin Breast Cancer Diagnostic (WDBC) veri seti, University of California Irvine (UCI) veri deposundan elde edilmiştir. Veriler Wisconsin Üniversite Hastanesi'nde Dr. Wolberg vd. tarafından toplanmış ve kamuya açık olarak paylaşımına sunulmuştur (UCI, 2021).

Veri setindeki öznitelikler, meme kitlelerinin FNA biyopsisi ile elde edilen görüntülerin sayısallaştırılması ile elde edilmiştir. Bu öznitelikler görüntüdeki hücre çekirdeğinin morfolojik özellikleridir. 30 adet tanımlayıcı özellik ve bir adet teşhis parametresi olmak üzere toplam 31 adet öznitelik bulunmaktadır. Her bir hücre çekirdeği için 10 ana özelliğin ortalama, standart hata ve en kötü (en büyük) değeri ile toplam 30 öznitelik tanımlanmıştır. 10 ana özellik ve elde edilmiş biçimleri Tablo 1'de verilmiştir. Bu 10 adet özelliğin ortalama (mean), standart hata (error) ve en kötü (worst) olmak üzere üçer farklı değeri ile toplam 30 adet öznitelik oluşturulmuştur. Bu ölçümlere kitlenin iyi huylu (Benign) ya da kötü huylu (Malignant) olma durumunu belirten Diagnosis isimli teşhis özniteliği eklenerek veri seti tanımlanmıştır. Veri seti içerisinde toplam 569 adet örnek yer almaktadır. Bunlardan 357 adedi iyi huylu, 212 adedi ise kötü huylu yani kanserlidir.

Tablo 1. Verinin ana özellikleri (Main features of the data)

ÖZELLİK ADI	AÇIKLAMA	ELDE EDİLİŞ ŞEKLİ
Radius	Yarıçap	Sınırın kütle merkezinden sınır noktalarının her birine doğru olan radyal doğru parçalarının uzunluğunun ortalaması şeklinde hesaplanmıştır.
Texture	Doku	Piksellerindeki gri yoğunluklarının varyansı alınarak elde edilmiştir.
Perimeter	Çevre	Ardışık sınır noktaları arasındaki mesafelerin toplamı ile hesaplanmıştır.
Area	Alan	Sınırın iç kısmındaki pikseller sayılarak elde edilmiştir.
Smoothness	Düzgünlük, pürüzsüzlük	Her bir radyal çizginin uzunluğu ile onu çevreleyen iki radyal çizginin ortalama uzunluğu arasındaki fark ölçülerek elde edilmiştir.
Compactness	Kompaktlık	Çevrenin karesinin alana bölümü ile hesaplanmıştır
Concavity	İçbükeylik	Hücre çekirdeğinin sınırındaki girintilerin boyutu ölçülerek elde edilmiştir.
Concave points	İçbükey noktalar	Sınırın içbükey bölgelerinde yer alan sınır noktalarının sayısıdır
Symmetry	Simetri	Hücre çekirdeğinin ana eksenine dik çizgi çiftleri arasındaki orantılı uzunluk farkı bulunarak ölçülmüştür.
Fractal dimension	Fraktal boyut	Çekirdeğin çevresi giderek büyüyen cetveller ile ölçülmüştür. Değerlerin log-log ölçeğinde çizilmesi ve düşey eğimin hesaplanması ile fraktal boyuta bir yaklaşık değer negatif olarak elde edilir.

2.2. Sınıflandırmada Kullanılan Makine Öğrenmesi Teknikleri (Machine Learning Techniques Used for Classification)

Bu çalışmada sınıflandırma için literatürde de yaygın olarak kullanılan 3 farklı makine öğrenmesi algoritması kullanılmıştır. Bunlar; K-En yakın Komşu (K-Nearest Neighbor / KNN), Rastgele Orman (RO) ve Destek Vektör Makinesi (DVM)' dir.

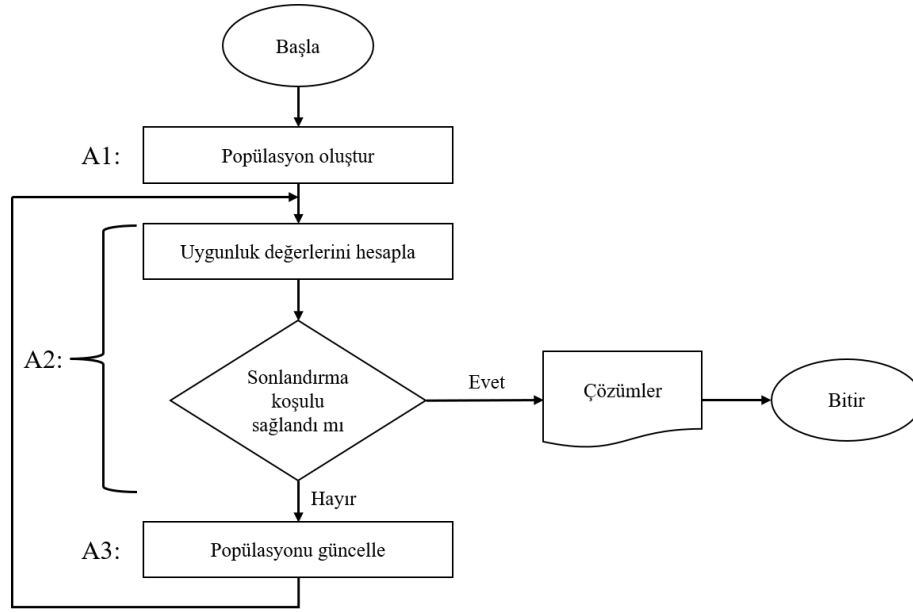
KNN algoritması, literatüre Fix ve Hodges (1952) tarafından kazandırılan, sınıflandırma ve regresyon problemleri için yaygın olarak kullanılan bir yöntemidir. Sınıfı belirlenmek istenen nokta K adet en yakın komşunun oyuyla en yaygın olan sınıfa atanır. KNN mesafeye dayalı ve en temel makine öğrenmesi algoritmalarından biridir. KNN ile gerçekleştirilen sınıflandırmada çıktı, bir sınıfın üyeliğidir.

RO algoritması, eğitim esnasında çok sayıda karar ağacı oluşturularak, her bir ağacın ürettiği sonuçların modu veya ortalamasını alıp çıktı sınıfın belirlenmesini sağlayan kolektif bir öğrenme algoritmasıdır. Ho (1995) tarafından oluşturulan yöntem dayanan RO, daha sonra Breiman (2001) tarafından geliştirilerek literatüre kazandırılmıştır. RO, geleneksel karar ağaçlarında yaygın problemlerden biri olan aşırı uydurma (overfitting) sorununa hem veri seti, hem de öznitelikleri çok sayıda parçaya bölüp birden çok ağaç üzerinde işleyerek çözüm getirir.

DVM algoritması, Vapnik vd. (1997) tarafından literatüre kazandırılan, istatistiksel öğrenme teorisine dayalı bir yöntemdir. DVM, her biri ayrı kategorilere ait olarak işaretlenmiş eğitim setinden öğrenerek, yeni örnekleri bu sınıflardan birine olasılıklı olmayacak şekilde atayan bir model oluşturur. Veri örneklerinin yer aldığı düzlemde, sınıfları birbirinden ayırmak için, iki sınıfın üyelerinden en uzak mesafede olacak şekilde bir karar sınırının çizilmesi sağlanır. DVM'nin, aşırı uydurma problemi karşısındaki hassasiyetinin düşük olması ve yüksek doğruluk sağlaması kullanım yaygınlığını arttırmaktadır.

2.3. Öznitelik Seçiminde Kullanılan Optimizasyon Algoritmaları (Optimization Algorithms Used for Feature Selection)

Doğa ilhamlı optimizasyon algoritmaları (DOA), bitki ve hayvan davranışlarından, sürü zekası, fiziksel, biyolojik, kimyasal sistemler dahil olmak üzere doğadan ilham alan bir grup algoritma olarak tanımlanır (Fister Jr vd., 2013). Bu algoritmalar global optimize edicilerdir ve optimal çözümü bulmak için dengeli global ve yerel arama bileşenlerine sahiptirler (Dhal vd., 2019). Parçacık temelli olarak çalışan DOA'lar, çok sayıda parçacık ile ideal sonucu arama ve hesaplama süreçlerini yerine getirirler. Temel mantıkları benzer olan DOA'ların genel işleyiş biçimini gösterir akış diyagramı Şekil 1'de verilmiştir.



Şekil 1. DOA genel işleyiş diyagramı (General mechanism diagram of nature inspired optimization algorithms)

A1 olarak işaretlenmiş adımda optimizasyon algoritmasının başlangıç popülasyonu mümkün olduğu kadar geniş çözüm alanını kapsayacak şekilde rastgele yöntemlerle oluşturulur. Popülasyon büyüklüğü gereksinimlere bağlı olarak ve elde edilen tecrübeler doğrultusunda seçilir. DOA'ların çoğu iteratif yöntemler kullanır ve maksimum iterasyon sayısı, hassasiyet eşiği gibi sonlandırma parametreleri de A1 adımında belirlenir. A2 adımında her bir parçacığın uygunluk değerleri bir uygunluk fonksiyonuna bağlı olarak hesaplanır. Uygunluk fonksiyonu, her bir çözümün performansını yansıtan benzersiz bir göstergedir ve genellikle maksimum veya minimum değere sahip olan hedef fonksiyona uygun şekilde tasarlanmıştır. Genel olarak her bireyin kendi yerel optimal çözümü vardır ve tüm popülasyonun da global bir optimumu vardır. A2 adımında, her iterasyonda popülasyonun uygunluk değerleri hesaplanır ve global en iyi çözüm, sonlandırma koşullarını sağlıyorsa çıkış olarak verilir. Aksi halde iç etkileşimlerle popülasyonu iyileştirmek için gerekli işlemlerin uygulandığı A3 adımına geçilir. Güncellenen popülasyon sonraki iterasyonu gerçekleştirmek üzere A2 adımına geri döner.

Gerçek dünya problemlerinde mevcut olan büyük verilerin çok sayıda özellik ve/veya örneğe sahip olması analiz etme sürecini zorlaştırmakta ve başarısı düşük sonuçların üretilmesine neden olmaktadır. Öznitelik seçimi (feature selection), optimal çözümleri elde etmek için bir veri kümesi oluşturan özelliklerin tümünden en uygun özellikleri belirleme ve seçme işlemidir. Öznitelik seçiminin iki temel amacı sınıflandırma doğruluğunu en üst düzeye çıkarmak ve öznitelik sayısını en aza indirmektir (Arora & Anand, 2019). Bu iki amaç çoğu durumda birbirleriyle çatışsa da ikisi arasında bir denge oluşturup mümkün olduğunca az ve etkin özelliklerle en yüksek doğruluğa ulaşmak hedeflenir.

DOA'lar veri setlerinden öznitelik alt seti seçimi konusunda son dönemlerde yaygın olarak kullanılmaktadır. Bu çalışmada öznitelik seçimi için Guguk Kuşu Arama (GKA) algoritması (Cuckoo search), Parçacık Sürü Optimizasyonu (PSO) algoritması (Particle swarm optimization), Balina Optimizasyon (BO) algoritması (Whale optimization) ve Kızıl Geyik (KG) algoritması (Red deer algorithm) olmak üzere 4 farklı DOA kullanılmıştır.

GKA algoritması, bazı guguk kuşu türlerinin yumurtalarını diğer ev sahibi türlerin yuvalarına bırakarak asalak biçimde üreme davranışlarından esinlenerek Yang ve Deb (2009) tarafından formülize edilen bir optimizasyon algoritmasıdır. Ev sahibi kuş yumurtaların kendine ait olmadığını keşfederse bu yumurtaları yuvadan atabilir ya da yuvasını terk ederek başka bir yerde yeni yuva kurabilir. Tapera gibi bazı guguk kuşu türleri asalak üreme için evrimleşerek yumurtalarını seçilen ev sahibi türün yumurtalarına benzer renk ve desende yapma konusunda ustalaşmışlardır. GKA algoritmasında, içerisinde yumurtaların yer aldığı her bir yuva, çözüm uzayında bir parçacığı temsil eder. Parçacıklar çözüm uzayına Lévy Uçuşları adlı rastgele hareket desenine bağlı olarak konumlanır. Her bir parçacığın bulunduğu konumda uygunluk değeri hesaplanır. Uygunluk fonksiyonunda iyi değerler üreten yuvalar/yumurtalar bir sonraki nesle aktarılır. Yuvalarını başka kuşlara kaptıran kuşlar (parçacıklar) ise yeni yuva arayışına girerek optimizasyon sürecini devam ettirirler.

PSO algoritması, balık ve kuş gibi sürüler halinde yaşayan canlıların yiyecek bulma gibi temel ihtiyaçlarını giderirken bireysel olarak gerçekleştirdikleri davranışların sürüdeki diğer bireyleri etkileyerek, sürünün amacına daha kolay ulaşmasını sağladığının gözlenmesinden hareketle Kennedy ve Eberhart (1995) tarafından geliştirilen

bir optimizasyon algoritmasıdır. PSO'da çözüme ulaşmaya çalışan her bir bireye parçacık, parçacıklardan oluşan popülasyona ise sürü adı verilir. Her parçacığın çözüme ne kadar yaklaştığı uygunluk fonksiyonu ile hesaplanır. Parçacığın hareketi ve yeni durumu, parçacığın kendi en iyi çözümü ve tüm popülasyondaki global en iyi çözüm temel alınarak, bir değişim hızı fonksiyonu ile belirlenir. Ardından uygunluk fonksiyonu kullanılarak çözüme ne kadar yaklaşıldığı tekrar kontrol edilir. Bu döngü belirlenen iterasyon şartları sağlanana kadar devam eder.

BO algoritması, Mirjalili ve Lewis (2016) tarafından geliştirilen, kambur balinaların avlanma davranışlarından esinlenen bir optimizasyon algoritmasıdır. Kambur balinalar "kabarık ağı avlanma manevrası" adı verilen bir taktik kullanırlar. Yüze yakın küçük balıkları avlamayı tercih eden kambur balinalar, yiyecek arama esnasında spiraller çizerek ve belirgin kabarcıklar çıkararak hareket ederler. BO algoritmasında avını arayan her bir balina bir parçacığı, çok sayıda balinadan oluşan popülasyon ise sürüyü oluşturur. BO'da popülasyon çok boyutlu bir uzayda arama yapar. Bir balina ile avı arasındaki mesafe maliyet değerine karşılık gelir. Bu algoritmada balinaların konumu zamana bağlı bir değerdir. Bireyin zamana bağlı konumu üç operasyonel süreçle ölçümlenir: Keşif aşaması, avın çevrelediği halkanın küçültülmesi ve kabarcık ağı saldırısı. Kambur balina avın konumunu belirleyip onu çevreler. Arama uzayındaki optimal çözümün yeri önceden bilinmediğinden BO mevcut en iyi adayın hedef av olduğunu ve optimuma en yakın olduğunu varsayar. Algoritma en iyi arama ajanını belirlemeye çalışırken diğer ajanlar da konumlarını en iyiye göre günceller. Konumların belirlenmesi esnasında kambur balinaların hareketlerini taklit eden spiral bir matematiksel fonksiyon kullanılır. Optimum çözüme ulaşana kadar güncelleme süreci devam eder.

KG algoritması İskoç kızıl geyiğinin (*Cervus Elaphus Scoticus*) üreme davranışından esinlenerek Fathollahi-Fard vd. (2020) tarafından geliştirilen bir optimizasyon algoritmasıdır. Popülasyon erkek ve dişi geyiklerden oluşur. Üreme sezonunda erkek geyikler sık sık kükrerler. Yüksek sesle kükreyen erkek dişiler tarafından daha çok tercih edilir. Kükrere oranı hem üreme başarısı hem de savaşma yeteneği ile pozitif ilişkilidir. Kükrere oranı yüksek erkekler hem diğer erkekler üzerinde baskı oluşturma hem de dişiler tarafından daha yüksek farkedilirliğe sahip olmalarından dolayı üreme başarıları artar. Güçlü kükreyen erkekler grup içerisinde komutan rolüne yükselir. Her komutan kendine ait bölgelerde kendi çiftleşme alanlarını oluşturur. Yakın çiftleşme alanlarındaki komutan geyikler birbirleri ile savaşır ve galip olan komutan diğerinin çiftleşme alanına girer. KG algoritması rastgele bir ilk popülasyonla başlar. Her bir bireyin çözümü yakınlığı bir uygunluk fonksiyonu ile belirlenir. En iyi çözümleri temsil eden geyiklerin kükrere oranları yüksektir. Yüksek kükrere oranına sahip belirli sayıda geyik komutan olarak kendi çiftleşme alanlarını oluşturur. Her komutan kendi çiftleşme alanında belirli bir yüzdelik dilimde dişi ile çiftleşir. Aynı zamanda komutanlar savaşarak galip geldikleri diğer komutanların çiftleşme alanında da belirli bir yüzdelik oranda dişi geyik ile çiftleşir. Aynı zamanda erkek geyikler kendilerine en yakın dişi ile de çiftleşir. Bu şekilde oluşturulan yeni popülasyondan uygunluk fonksiyonuna göre yeni nesiller seçilir. Optimum çözüme ulaşana kadar süreç devam eder.

3. Deneysel Sonuçlar (Experimental Results)

Bu çalışmada 30 adet giriş özneliği ve 1 adet hedef teşhis parametresinden oluşan WDBC veri seti üzerinde sınıflandırmalar gerçekleştirilmiştir. İlk aşamada veri setindeki tüm öznelikler kullanılarak 3 farklı makine öğrenmesi algoritmasının sınıflandırma başarıları test edilmiştir. Ardından veri setinin orijinal öznelikleri içerisinde dört farklı DOA kullanarak öznelik alt setleri seçilmiş ve bu setler üzerinde aynı sınıflandırma algoritmaları kullanılarak işlemler tekrarlanmıştır. Orijinal veri setinde yer alan 30 adet giriş özneliği ve veri seti içindeki sıra numaraları Tablo 2'de yer almaktadır.

Tablo 2. WDBC veri setinin öznelikleri (Features of the WDBC dataset)

ÖZNETELİK NO	ÖZNETELİK ADI	ÖZNETELİK NO	ÖZNETELİK ADI
1	mean radius	16	compactness error
2	mean texture	17	concavity error
3	mean perimeter	18	concave points error
4	mean area	19	symmetry error
5	mean smoothness	20	fractal dimension error
6	mean compactness	21	worst radius
7	mean concavity	22	worst texture
8	mean concave points	23	worst perimeter
9	mean symmetry	24	worst area
10	mean fractal dimension	25	worst smoothness
11	radius error	26	worst compactness
12	texture error	27	worst concavity
13	perimeter error	28	worst concave points
14	area error	29	worst symmetry
15	smoothness error	30	worst fractal dimension

Sınıflandırma işlemlerinin başarısını raporlamak için karmaşıklık matrislerinden elde edilen değerler ile hesaplanan metrikler kullanılmıştır. Bir karmaşıklık matrisinin genel yapısı Şekil 2’de verilmiştir.

		Gerçek sınıf	
		Meme kanseri	Kanser değil
Tahmin edilen sınıf	Meme kanseri	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Kanser değil	Yanlış Negatif (YN)	Doğru Negatif (DN)

Şekil 2. Karmaşıklık matrisi (Confusion matrix)

Sınıflandırıcı tarafından meme kanseri olduğu belirlenen örneğin gerçek sınıfı da meme kanseri ise bu durum Doğru Pozitif (DP), gerçek sınıfı kanser değil ise Yanlış Pozitif (YP) olarak adlandırılır. Benzer şekilde sınıflandırıcının kanserli olmadığını belirlediği örneğin gerçek sınıfı da kanserli değil ise bu durum Doğru Negatif (DN), gerçek sınıfı meme kanseri ise Yanlış Negatif (YN) olarak adlandırılır. Bu değerler kullanılarak hesaplanan performans metrikleri ve formülleri Tablo 3’te verilmiştir.

Tablo 3. Model performans metrikleri (Model performance metrics)

METRİK	MATEMATİKSEL İFADESİ
Doğruluk	$(DP + DN) / (DP + YP + YN + DN)$
Kesinlik	$DP / (DP + YP)$
Duyarlılık	$DP / (DP + YN)$
F1 Skoru	$2 * kesinlik * duyarlılık / (kesinlik + duyarlılık)$

İlk etapta WDBC veri setinde yer alan ve Tablo 2’de belirtilen 30 özneliğin tamamı kullanılarak KNN, RO ve DVM yöntemleri ile sınıflandırmalar gerçekleştirilmiştir. Kullanılan sınıflandırma yöntemlerinin parametreleri üzerinde herhangi bir değişiklik yapılmamış olup varsayılan değerleriyle kullanılmıştır. 5 kat çapraz doğrulama kullanılarak gerçekleştirilen sınıflandırma işlemlerinin sonuçları Tablo 3’te belirtilen metrikler ile raporlanmıştır. Öznelik seçimi uygulanmadan gerçekleştirilen bu sınıflandırmalar sonucunda elde edilen bulgular Tablo 4’te verilmiştir.

Tablo 4. Öznelik seçimi yapılmadan elde edilen sonuçlar (Results without feature selection)

ÖZNETELİK SAYISI	SINIFLANDIRICI	DOĞRULUK	KESİNLİK	DUYARLILIK	F1 SKORU
30 (tamamı)	KNN	0,929825	0,910256	0,986111	0,946667
	RO	0,938596	0,933333	0,972222	0,952381
	DVM	0,921053	0,909091	0,972222	0,939597

İkinci aşamada WDBC veri setinde yer alan öznelikler üzerinde GKA, PSO, BO, KG olmak 4 farklı DOA ile, farklı sayıda parçacıklar kullanılarak öznelik alt seti seçimleri gerçekleştirilmiştir. Her bir DOA için 25, 50 ve 75 adet parçacık kullanılarak, her biri 100’er iterasyon çalıştırılmıştır. Seçilen öznelikler üzerinde KNN, RO ve DVM yöntemleri ile sınıflandırmalar gerçekleştirilmiştir. Her bir durumda seçilen öznelikler ve bunlar üzerinde gerçekleştirilen sınıflandırmalara ait bulgular Tablo 5’te yer almaktadır.

Tablo 5’teki sonuçlar incelendiğinde her bir DOA için parçacık sayısındaki artışın elde edilen sınıflandırma başarılarında da artış sağladığı görülmektedir. Ayrıca parçacık sayısındaki artış genel olarak daha çok özneliğin seçilmesini sağlamıştır. En yüksek doğruluk değeri GKA algoritması ile 75 parçacık kullanılarak seçilen 16 adet öznelik üzerinde RO sınıflandırıcı ile %99.12 olarak elde edilmiştir. Kesinlik ve duyarlılık metrikleri için %100 değer elde edilen farklı durumlar bulunmakla birlikte bu iki metriğin dengesini ifade eden F1 Skorunun en yüksek değeri yine aynı sınıflandırma işleminde %99.30 olarak elde edilmiştir. Farklı sınıflandırmalarda seçilen en az öznelik sayısı 12 olup, bu sayıda öznelikle elde edilen en yüksek doğruluk %97.36 olarak RO algoritması ile 25 parçacıklı PSO ve 50 parçacıklı KG kullanılan durumlarda elde edilmiştir. En az parçacıkla, en az sayıda öznelik seçilerek en iyi doğruluğu PSO ile birlikte RO kullanılan durum sağlamıştır. Gerçekleştirilen sınıflandırmalar sonucunda, kullanılan DOA ve sınıflandırıcı bazında elde edilen en yüksek doğruluk değerleri özet olarak Tablo 6’da verilmiştir.

Tablo 5. Öznitelik seçimi ile elde edilen sonuçlar (Results with feature selection)

DOA	PARÇACIK SAYISI	SEÇİLEN ÖZİNTELİKLER	ÖZİNTELİK SAYISI	SINIFLANDIRICI	DOĞRULUK	KESİNLİK	DUYARLILIK	F1 SKORU
GKA	25	1-5-7-8-10-21-22-23-29	9	KNN	0,93859649	0,97101449	0,93055556	0,95035461
				RO	0,96491228	0,94736842	1	0,97297297
				DVM	0,96521739	0,97260274	0,97260274	0,97260274
	50	1-3-6-11-13-19-20-21-22-23-26-27-28-29-30	15	KNN	0,97368421	0,98591549	0,97222222	0,97902098
				RO	0,98245614	1	0,97222222	0,98591549
				DVM	0,97368421	0,97260274	0,98611111	0,97931034
	75	2-3-5-6-8-9-10-11-12-17-18-19-20-22-23-26	16	KNN	0,98245614	0,98611111	0,98611111	0,98611111
				RO	0,99122807	1	0,98611111	0,99300699
				DVM	0,97368421	0,97260274	0,98611111	0,97931034
PSO	25	2-8-10-12-13-17-18-20-23-25-28-29	12	KNN	0,95614035	0,95890411	0,97222222	0,96551724
				RO	0,97368421	0,97260274	0,98611111	0,97931034
				DVM	0,92982456	0,93243243	0,95833333	0,94520548
	50	1-2-3-5-7-8-9-11-12-22-23-26-27-28-29-30	16	KNN	0,96491228	0,97222222	0,97222222	0,97222222
				RO	0,96491228	0,95945946	0,98611111	0,97260274
				DVM	0,95614035	0,93506494	1	0,96644295
	75	1-2-3-4-5-6-9-10-14-17-18-19-20-24-27-29-30	17	KNN	0,95614035	0,95890411	0,97222222	0,96551724
				RO	0,98245614	0,97297297	1	0,98630137
				DVM	0,95614035	0,94666667	0,98611111	0,96598639
BO	25	1-5-6-10-11-12-13-17-21-25-27-29-30	13	KNN	0,93859649	0,93333333	0,97222222	0,95238095
				RO	0,95614035	0,93506494	1	0,96644295
				DVM	0,92982456	0,9	1	0,94736842
	50	1-4-5-6-9-12-13-15-16-18-20-22-24-26-28-29-30	17	KNN	0,95614035	0,94666667	0,98611111	0,96598639
				RO	0,97368421	0,97260274	0,98611111	0,97931034
				DVM	0,93859649	0,92207792	0,98611111	0,95302013
	75	3-5-7-9-10-12-13-14-15-17-18-19-21-22-23-28-29-30	18	KNN	0,94736842	0,94594595	0,97222222	0,95890411
				RO	0,96491228	0,97222222	0,97222222	0,97222222
				DVM	0,92982456	0,91025641	0,98611111	0,94666667
KG	25	1-2-6-7-11-13-15-16-17-18-20-21-22-23-25-29-30	17	KNN	0,95614035	0,95890411	0,97222222	0,96551724
				RO	0,96491228	0,94736842	1	0,97297297
				DVM	0,94736842	0,94594595	0,97222222	0,95890411
	50	5-8-9-12-15-17-19-20-21-22-26-27	12	KNN	0,96491228	0,95945946	0,98611111	0,97260274
				RO	0,97368421	0,98591549	0,97222222	0,97902098
				DVM	0,94736842	0,94594595	0,97222222	0,95890411
	75	2-3-6-8-11-12-16-18-21-22-23-28-29	13	KNN	0,95614035	0,93506494	1	0,96644295
				RO	0,97368421	0,96	1	0,97959184
				DVM	0,93859649	0,91139241	1	0,95364238

Tablo 6. Elde edilen en iyi doğruluk değerleri (The best accuracy values obtained)

DOA	KNN	RO	DVM	Ortalama
-	0,929825	0,938596	0,921053	0,929825
GKA	0,982456	0,991228	0,973684	0,982456
PSO	0,964912	0,982456	0,95614	0,967836
BO	0,95614	0,973684	0,95614	0,961988
KG	0,964912	0,973684	0,947368	0,961988
Ortalama	0,959649	0,97193	0,950877	

Tablo 6'daki değerler doğrultusunda öz nitelik seçimi olmadan ve farklı DOA'lar ile öz nitelik seçimi yapılarak elde edilen sonuçlar Şekil 2'de grafik olarak da gösterilmiştir.



Şekil 2. Özellik seçimi ile ve özellik seçimi olmadan elde edilen en iyi doğruluk değerleri (Best accuracy values obtained with and without feature selection)

Sonuçlar incelendiğinde öz nitelik seçimi yapılan tüm durumlarda elde edilen başarının öz nitelik seçimi yapılmadan gerçekleştirilen sınıflandırmalardan daha yüksek olduğu görülmektedir. Tüm sınıflandırıcılar için en yüksek doğruluk değerlerinin GKA algoritması kullanılarak öz nitelik seçimi yapılan durumlarda elde edildiği görülmektedir. GKA kullanılarak elde edilen sınıflandırma doğruluklarının ortalaması %98.24 iken, ikinci sırada %96.78 ortalama ile PSO, ardından %96.19 ile BO ve KG gelmektedir. Kullanılan 3 sınıflandırıcı içerisinde en yüksek başarı RO ile elde edilmiştir. RO kullanılarak elde edilen doğruluk değerlerinin ortalaması %97.19 iken, bunu %95.96 ile KNN ve %95.08 ile DVM takip etmektedir.

Bu çalışmada elde edilen bulguların, aynı veri setini kullanılan literatürdeki güncel çalışmaların bulguları ile karşılaştırılması Tablo 7'de verilmiştir.

Tablo 7. Bulguların literatürdeki çalışmalar ile karşılaştırılması (Comparison of the findings with the literature)

	REFERANS	YÖNTEM	EN İYİ DOĞRULUK (%)
1	(Khorshid vd., 2021)	W-KNN	96.7
2	(Ahuja vd., 2021)	W-KNN + PCA	98.9
3	(Vijayakumar vd., 2021)	ReLU DNN + Adagrad	97.94
4	(Feroz vd., 2021)	KNN ve RO	97.14
5	(Muhammad Amin & Inna, 2021)	LR	93.9
6	(Jabbar, 2021)	Bayes + RBF	97
7	(Assegie, 2021)	Optimized KNN	94.35
8	(Gupta & Kaushik, 2021)	Ensemble learning	97.9
9	(Ghosh vd., 2021)	LSTM	99.10
10	(Lahoura vd., 2021)	DVM	92.98
11	(Pravesjit vd., 2021)	NB+PSO	97.27
12	(Jijitha & Amudha, 2021)	LR+GA	98.24
13	(Devi vd., 2021)	LR+NN	98.5
14	(Parekh vd., 2021)	DVM	96.48
15	(Naji vd., 2021)	DVM	97.2
16	(Magboo & Magboo, 2021)	LR+PCA	80
17	(Al-Azzam & Shatnawi, 2021)	KNN ve LR	98
18	(Pawar vd., 2021)	XGBoost	98.24
19	(Al-Yaseen vd., 2021)	KMeans + DVM	98.067
20	(Ara vd., 2021)	DVM	96.5
	Bu çalışma	RO+GKA	99.122

Tablo 7'de yer alan çalışmalar ile karşılaştırıldığında bu çalışmada GKA algoritması ve RO sınıflandırıcı kullanılarak elde edilen %99.12'lik doğruluk oranının diğer çalışmalardan daha yüksek olduğu görülmektedir.

DOA kullanan diğer iki çalışmada Pravesjit vd. (2021) PSO ve NB ile %97.27, Jijitha ve Amudha (2021) GA ve LR kullanarak %98.24 ile bu çalışmadan daha düşük bir doğruluk elde etmişlerdir. Devi vd. (2021) sınıflandırma öncesinden gerçekleştirdikleri öznitelik seçiminde bu çalışmaya benzer şekilde 16 elemanlı bir öznitelik alt seti elde etmişler, bunun üzerinde LR ve NN'den oluşan kolektif bir yöntemle %98.5 oranla bu çalışmadan daha düşük bir doğruluğa ulaşmışlardır. Bu çalışma ile benzer RO sınıflandırıcı kullanan Feroz vd. (2021) tarafından yürütülen diğer çalışmada elde edilen doğruluk ise %97.14'tür. Bu çalışmaya en yakın doğruluk ise %99.1 olarak Ghosh vd. (2021) tarafından daha kompleks bir derin öğrenme yöntemi ile elde edilmiştir.

4. Sonuç (Conclusion)

Bu çalışmada meme kanseri tespitinde literatürde yaygın kullanılan WDBC veri seti üzerinde öznitelik seçimi ve farklı sınıflandırıcıların performansları üzerine test ölçümleri gerçekleştirilmiştir. Öznitelik seçimi için farklı doğa ilhamlı optimizasyon algoritmaları kullanılarak başarıları karşılaştırılmıştır. 30 öznitelikten oluşan orijinal veri seti üzerinde KNN, RO ve DVM algoritmaları ile gerçekleştirilen sınıflandırmalarda en yüksek doğruluk %93.85 olarak RO sınıflandırıcı ile elde edilmiştir. Veri setindeki tüm özniteliklerin kullanılması analiz sürecini zorlaştırmakta ve başarısı düşük sonuçların üretilmesine neden olabilmektedir. Bunu test etmek amacıyla orijinal veri seti üzerinde GKA, PSO, BO ve KG olmak üzere 4 ayrı optimizasyon algoritması ile öznitelik seçim işlemleri gerçekleştirilmiştir. Ardından seçilen öznitelik alt setleri üzerinde aynı sınıflandırıcılarla gerçekleştirilen sınıflandırma işlemlerinde %10'a varan başarı artışı olduğu gözlenmiştir. Optimizasyon algoritmalarının parçacık sayıları değiştirilerek performans değişimleri izlenmiş, genel anlamda parçacık sayısının artması ile seçilen özniteliklerdeki değişimin başarıyı arttırdığı görülmüştür. Öznitelik seçimi ile gerçekleştirilen sınıflandırmalarda en başarılı sonuç 75 parçacıklı GKA algoritması ve RO sınıflandırıcı kullanılan durumda %99.12 olarak elde edilmiştir. Elde edilen bulgular literatürde aynı veri setini kullanan güncel çalışmaların bulguları ile karşılaştırılmış ve genelinden daha yüksek sonuç elde edildiği görülmüştür. Geçekleştirilen bu çalışma meme kanseri tespiti konusunda uzman bir karar destek sisteminin oluşturulmasında başarı ile kullanılabilecek niteliktedir.

Çıkar Çatışması (Conflict of Interest)

Yazar tarafından herhangi bir çıkar çatışması beyan edilmemiştir. No conflict of interest was declared by the author.

Kaynaklar (References)

- Ahuja, A., Al-Zogbi, L., & Krieger, A. (2021). Application of Noise-Reduction Techniques to Machine Learning Algorithms for Breast Cancer Tumor Identification. *Computers in Biology and Medicine*, 104576.
- Al-Azzam, N., & Shatnawi, I. (2021). Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer. *Annals of Medicine and Surgery*, 62, 53-64. <https://doi.org/10.1016/j.amsu.2020.12.043>
- Al-Yaseen, W., Jehad, A., Abed, C. I., & Idrees, A. K. (2021). The use of modified K-Means algorithm to enhance the performance of support vector machine in classifying breast cancer. *Int. J. Intell. Eng. Syst*, 14(2).
- Ara, S., Das, A., & Dey, A. (2021). Malignant and Benign Breast Cancer Classification using Machine Learning Algorithms. 2021 International Conference on Artificial Intelligence (ICAI), 97-101.
- Arora, S., & Anand, P. (2019). Binary butterfly optimization approaches for feature selection. *Expert Systems with Applications*, 116, 147-160. <https://doi.org/10.1016/j.eswa.2018.08.051>
- Assegie, T. A. (2021). An optimized K-Nearest Neighbor based breast cancer detection. *Journal of Robotics and Control (JRC)*, 2(3), 115-118.
- Bayrak, E. A., Kırıcı, P., & Ensari, T. (2019). Comparison of machine learning methods for breast cancer diagnosis. 2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT), 1-3.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Devi, M. V., Sravani, M., Ramya, K., Bindulakshmisai, N., Parameshwari, V., & others. (2021). Breast Cancer Diagnosis Using Adaptive Voting Ensemble Machine Learning Algorithm. *UGC Care Group I Listed Journal*, 11(1), 495-501.
- Dhal, K. G., Ray, S., Das, A., & Das, S. (2019). A Survey on Nature-Inspired Optimization Algorithms and Their Application in Image Enhancement Domain. *Archives of Computational Methods in Engineering*, 26(5), 1607-1638. <https://doi.org/10.1007/s11831-018-9289-9>
- Fathollahi-Fard, A. M., Hajiaghahi-Keshteli, M., & Tavakkoli-Moghaddam, R. (2020). Red deer algorithm (RDA): A new nature-inspired meta-heuristic. *Soft Computing*, 24(19), 14637-14665.
- Feroz, N., Ahad, M. A., & Doja, F. (2021). Machine Learning Techniques for Improved Breast Cancer Detection and Prognosis—A Comparative Analysis. *Çinde Applications of Artificial Intelligence and Machine Learning* (ss. 441-455). Springer.
- Fister Jr, I., Yang, X.-S., Fister, I., Brest, J., & Fister, D. (2013). A brief review of nature-inspired algorithms for optimization. *arXiv preprint arXiv:1307.4186*.
- Fix, E., & Hodges Jr, J. L. (1952). Discriminatory analysis-nonparametric discrimination: Small sample performance. *California Univ Berkeley*.

- Ghosh, P., Azam, S., Hasib, K. M., Karim, A., Jonkman, M., & Anwar, A. (2021). A Performance Based Study on Deep Learning Algorithms in the Effective Prediction of Breast Cancer. 2021 International Joint Conference on Neural Networks (IJCNN), 1-8.
- Gupta, N., & Kaushik, B. N. (2021). Prognosis and Prediction of Breast Cancer Using Machine Learning and Ensemble-Based Training Model. *The Computer Journal*.
- Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd international conference on document analysis and recognition*, 1, 278-282.
- Jabbar, M. A. (2021). Breast Cancer Data Classification Using Ensemble Machine Learning. *Engineering and Applied Science Research*, 48(1), 65-72.
- Jerez-Aragonés, J. M., Gómez-Ruiz, J. A., Ramos-Jiménez, G., Muñoz-Pérez, J., & Alba-Conejo, E. (2003). A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial intelligence in medicine*, 27(1), 45-63.
- Jijitha, S., & Amudha, T. (2021). Breast cancer prognosis using machine learning techniques and genetic algorithm: Experiment on six different datasets. *İçinde Evolutionary Computing and Mobile Sustainable Networks (ss. 703-711)*. Springer.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-international conference on neural networks*, 4, 1942-1948.
- Khorshid, S. F., Abdulazeez, A. M., & Sallow, A. B. (2021). A comparative analysis and predicting for breast cancer detection based on data mining models. *Asian Journal of Research in Computer Science*, 45-59.
- Kumar, S., & Bhargava, C. (2019). A review on Breast Cancer Analysis using Machine Learning. *International Journal of Emerging Technologies and Innovative Research*, 6, 326-329.
- Lahoura, V., Singh, H., Aggarwal, A., Sharma, B., Mohammed, M. A., Damaševičius, R., Kadry, S., & Cengiz, K. (2021). Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. *Diagnostics*, 11(2), 241.
- Magboo, V. P. C., & Magboo, M. S. A. (2021). Machine Learning Classifiers on Breast Cancer Recurrences. *Procedia Computer Science*, 192, 2742-2752. <https://doi.org/10.1016/j.procs.2021.09.044>
- Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, 95, 51-67.
- Muhammad Amin, B., & Inna, E. (2021). Breast Cancer Prediction Model Using Machine Learning. *Journal of Data Science*, 2021(02).
- Naji, M. A., Filali, S. E., Aarika, K., Benlahmar, E. H., Abdelouahid, R. A., & Debauche, O. (2021). Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis. *Procedia Computer Science*, 191, 487-492. <https://doi.org/10.1016/j.procs.2021.07.062>
- Papageorgiou, E. I., Subramanian, J., Karmegam, A., & Papandrianos, N. (2015). A risk management model for familial breast cancer: A new application using Fuzzy Cognitive Map method. *Computer methods and programs in biomedicine*, 122(2), 123-135.
- Parekh, D. H., Dahiya, V., & others. (2021). Predicting breast cancer using machine learning classifiers and enhancing the output by combining the predictions to generate optimal F1-score. *Biomedical and Biotechnology Research Journal (BBRJ)*, 5(3), 331.
- Pawar, S., Bagal, P., Shukla, P., & Dawkhar, A. (2021). Detection of Breast Cancer using Machine Learning Classifier. 2021 Asian Conference on Innovation in Technology (ASIANCON), 1-5. <https://doi.org/10.1109/ASIANCON51346.2021.9544767>
- Pravesjit, S., Longpradit, P., Kantawong, K., Pengchata, R., & Oul, N. (2021). A Hybrid PSO with Rao Algorithm for Classification of Wisconsin Breast Cancer Dataset. 2021 2nd International Conference on Big Data Analytics and Practices (IBDAP), 68-71. <https://doi.org/10.1109/IBDAP52511.2021.9552152>
- Shrivastavat, S. S., Sant, A., & Aharwal, R. P. (2013). An overview on data mining approach on breast cancer data. *International Journal of Advanced Computer Research*, 3(4), 256.
- Sinha, N. K., Khulal, M., Gurung, M., & Lal, A. (2020). Developing a web based system for breast cancer prediction using xgboost classifier. *International Journal of Engineering Research Technology (IJERT)*, 9.
- Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3), 209-249.
- UCI. (2021). Breast Cancer Wisconsin (Diagnostic) Data Set. Erişim tarihi 04 Ekim 2021, [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- Vapnik, V., Golowich, S. E., Smola, A., & others. (1997). Support vector method for function approximation, regression estimation, and signal processing. *Advances in neural information processing systems*, 281-287.
- Vijayakumar, K., Kadam, V. J., & Sharma, S. K. (2021). Breast cancer diagnosis using multiple activation deep neural network. *Concurrent Engineering*, 29(3), 275-284.
- Yang, X.-S., & Deb, S. (2009). Cuckoo Search via Lévy flights. 2009 World Congress on Nature Biologically Inspired Computing (NaBIC), 210-214. <https://doi.org/10.1109/NABIC.2009.5393690>