

Öznelik seçimi problemleri için ikili beyaz köpekbalığı optimizasyon algoritması

Binary white shark optimization algorithm for feature selection problems

Funda KUTLU ONAY*, 

Amasya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 05200, Amasya

• Geliş tarihi / Received: 15.09.2022

• Kabul tarihi / Accepted: 21.01.2023

Öz

Öznelik seçimi, makine öğrenmesi problemlerinde kabul edilebilir bir sınıflandırma doğruluğunu hedeflerken, aynı zamanda büyük ölçekli bir veri kümesinden gereksiz, alakasız ve gürültülü öznelikleri elimine etme işlemidir. Aslında öznelik seçimi de bir optimizasyon problemi olarak nitelendirilebilir. Literatürde metasezgisel optimizasyon algoritmalarının, optimum öznelik alt kümelerinin bulunmasında başarılı performansla sahip olduğu çalışmalar mevcuttur. Bu çalışmada da beyaz köpek balığı optimizasyon algoritması (BKO), S, V ve U-şekilli transfer fonksiyonları ile ikili forma dönüştürülerek öznelik seçimi için kullanılmıştır. Önerilen yöntemler UCI veri deposundaki sekiz farklı veri kümesi üzerinde uygulanmış ve sınıflandırma doğrulukları, uygunluk değerleri ve seçilen öznelik sayıları yönünden incelenmiştir. Sınıflandırıcı olarak k-en yakın komşuluk sınıflandırıcısı kullanılmıştır. Daha sonra farklı metasezgisel algoritmalarla karşılaştırılarak Freidman sıralama testi uygulanmıştır. Deneysel sonuçlar önerilen metotların, öznelik seçiminde başarılı olduğunu ve sınıflandırma başarısını artırdığını göstermektedir. Özellikle V ve U-şekilli versiyonların daha kararlı ve yüksek doğrulukla sonuçlar ürettiği yorumu yapılabilir.

Anahtar kelimeler: Beyaz köpek balığı optimizasyonu algoritması, Metasezgisel algoritmalar, Öznelik seçimi, Sınıflandırma, Transfer fonksiyonu

Abstract

In this study, the geometric process (GP) model is considered in order to calculate the debugging and testing costs of a software product. Under the assumption of the GP model, the debugging and testing costs of the software product are obtained depending on the first and second moment functions of the GP. It is observed that the values of the first and second moment functions of the process must be known in order to calculate the debugging and testing costs. At the same time, the calculation of moment functions also depends on both the distribution of the first interarrival time of the GP and the estimates of the model and distribution parameters. In this study, the proposed debugging and testing costs are calculated for the data set containing 136 failure times of a real-time command and control system. For this dataset, it has been shown in previous studies that the GP with gamma distribution can be proposed as a model. Under gamma distribution assumption, the maximum likelihood estimates of the model parameters are obtained. Using the estimates of the model parameters, the first and second moment functions of the GP are calculated with the help of the numerical methods proposed for these functions. Finally, the debugging and testing costs are obtained for the data set.

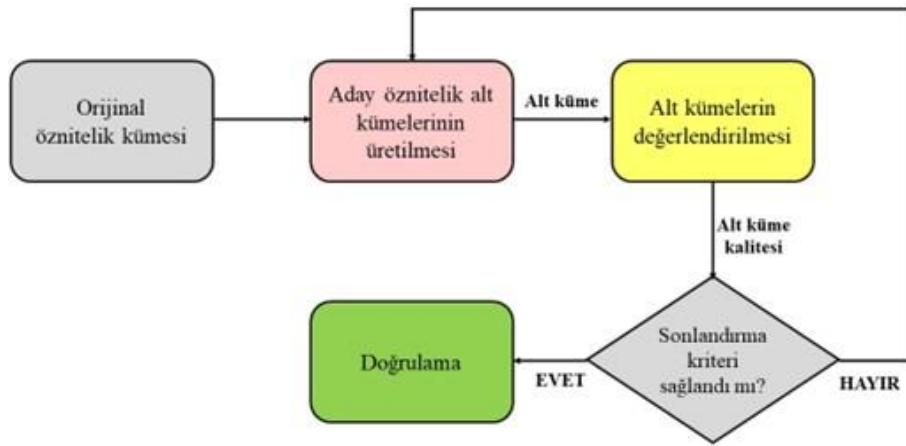
Keywords: White shark optimization algorithm, Metaheuristic algorithms, Feature selection, Classification Transfer function

* Funda KUTLU ONAY; funda.kutlu@amasya.edu.tr

1. Giriş

1.1. Introduction

Son zamanlarda, teknolojinin gelişmesinin bir sonucu olarak birçok alanda gerçekleştirilen gerçek dünya uygulamalarında düzenli olarak büyük miktarda bir veri üretimi ve bu verilerin depolanması durumu söz konusudur. Bu verilerden birçoğu gereksiz veya problemle ilgisi olmayan veriler olabilir. Bu nedenle, anlamlı bilgiyi ortaya çıkarmak ve analiz etmek gerekir. Hesaplama karmaşıklığını azaltabilmek ve etkin bir sınıflandırma performansı yakalayabilmek için bu ayıklama işleminin yürütülmesi kaçınılmaz olabilir. Öznitelik seçimi (ÖS), verilerden gereksiz ve alakasız özelliklerin ortadan kaldırıldığı bir ön işleme fazıdır. Sınıflandırma problemleri için ÖS, özellikle büyük veri kümeleri içeren problemler için, zorlu ve hesaplama açısından pahalı bir süreç olabilir (Dash & Liu, 1997). ÖS sürecinin temel amacı, öğrenme algoritmasının öğrenme hızı veya sınıflandırma doğruluğu gibi önemli parametrelerinin iyileştirilmesi için en iyi öznitelik alt kümesini belirlemektir. Şekil 1'de ÖS sürecinin akış diyagramı ifade edilmiştir. Buna göre, ilk aşamada yeni bir öznitelik alt kümesi oluşturmak için yöntemler uygulanır. Daha sonra, üretilen çözümlerin etkinliğini değerlendirmek üzere sınıflandırıcı ve uygunluk fonksiyonu kullanılır. Bu işlem sonlandırma kriteri sağlanana dek devam eder. Kriter sağlanmadığı durumda, yeniden aday öznitelik alt kümeleri oluşturulur.



Şekil 1. Öznitelik seçim sürecinin akış diyagramı

Figure 1. Flow chart of the feature selection process

ÖS, aslında özniteliklerin optimum kombinasyonunu bulmaya yönelik bir arama sürecidir. Örneğin N özellikli bir öznitelik kümesinin, 2^N-1 adet alt kümesi mevcuttur ve bunu değerlendirmek zor bir problemdir (Siedlecki & Sklansky, 1993). Bu durumda en iyi alt kümeyi bulmak için arama algoritması kullanılamaz. Öznitelik seçme algoritmaları filtre modeli, sarmal modeli ve gömülü model olmak üzere sınıfta ayrılmaktadır (Saeys, Inza, & Larranaga, 2007). Filtre modeli algoritmalar, bazı istatistiksel varsayımlar ile her bir özniteliğin sonuca katkısını hesaplayarak öznitelikleri değerlendirir ve belli bir eşik değerinin üzerindeki öznitelikleri hesaba katarak öznitelik seçimi gerçekleştirir. ReliefF (Robnik-Šikonja & Kononenko, 2003), Fskor (Duda, Hart, & Stork, 2001), bilgi kazancı (Quinlan, 1986) filtre modeli öznitelik çıkarım algoritmalarına örnek olarak gösterilebilir. Sarma modeli öznitelik seçim algoritmaları, sınıflandırma doğruluk oranını arttırmayı hedef alarak öznitelikleri seçmeyi hedefler. Dolayısıyla öznitelik sayısı arttıkça algoritmanın çalışma süresi de artmaktadır. Bundan dolayı sarmal yöntemler, filtre tabanlı yöntemlere göre daha yavaş çalışırlar. Genetik algoritma (Jh, 1975), sıralı geri eleme ve sıralı ileri eleme (Kittler, 1978) sarmal modeli öznitelik seçim yöntemlerine örnek gösterilebilir. Son olarak, gömülü modeli algoritmalar, sınıflandırma yapılırken öznitelik seçimi yapan modellerdir. Dolayısıyla gömülü yöntemler de filtre yöntemlerinden daha yavaş çalışmaktadırlar. Karar ağaçları (Grabczewski & Jankowski, 2005) ve destek vektör makineleri (Pal & Foody, 2010) gömülü yöntemlerin en sık kullanılan örneklerindedir. Metasezgisel algoritmalar (MA) da bu kısımda devreye girerler. MA, global optimizasyon problemlerine en uygun çözümleri bulmak için geliştirilen ve son zamanlarda birçok araştırma alanında popüler olarak kullanılan algoritmalar ve umut verici sonuçlar üretmektedirler (Abdel-Basset, Abdel-Fatah, & Sangaiyah, 2018). Bunlardan başlıcaları olarak parçacık sürü optimizasyonu (Poli, Kennedy, & Blackwell, 2007), evrimsel algoritmalar (Bäck & Schwefel, 1993), yapay arı kolonisi (Karaboga, Gorkemli, Ozturk, & Karaboga, 2014), gri kurt optimizasyonu (Long & Xu, 2016), Harris'in şahinleri (Fan, Chen, & Xia, 2020), balçık kalıbı algoritması (S. Li, Chen, Wang, Heidari, & Mirjalili, 2020), sinüs-kosinüs algoritması (Mirjalili,

2016) ve aritmetik optimizasyon (Abualigah, Diabat, Mirjalili, Abd Elaziz, & Gandomi, 2021) örnek verilebilir.

ÖS süreci için algoritmanın genel yapısını bozmaksızın yani MA ile aynı süreci uygulayarak ikilileştirme işlemi yapılması gerekir. İkilileştirme işlemi ile konum değerleri 0 veya 1 değerlerini alırlar ve bu işlem transfer fonksiyonu denilen matematiksel ifadelerle gerçekleştirilir. Bir girdi transfer fonksiyonuna geçirildikten sonra rastgele bir sayı ile karşılaştırılır ve karşılaştırma sonucunda göre 0 veya 1 değeri verilir. Literatürde ikilileştirme işlemi uygulanıp öznitelik seçiminde kullanılan çok sayıda metasezgisel algoritma vardır. Bunlardan en yenileri olarak ikili balina optimizasyon algoritması (Hussien, Hassanien, Houssein, Amin, & Azar, 2020), ikili sinüs kosinüs algoritması (Taghian & Nadimi-Shahraki, 2019), ikili atom arama optimizasyonu (Too & Rahim Abdullah, 2020), ikili karadelik optimizasyonu (Awadallah, Hammouri, Al-Betar, Braik, & Abd Elaziz, 2022), ikili salp sürüsü optimizasyonu (Dehghanivd., 2021), ikili kelebek optimizasyonu (Arora & Anand, 2019), ikili güve alevi optimizasyonu (Nadimi-Shahraki, Banaie-Dezfouli, Zamani, Taghian, & Mirjalili, 2021), ikili çekirge optimizasyonu (Hichem, Elkamel, Rafik, Mesaaoud, & Ouahiba, 2019), ikili Harris'in şahinleri optimizasyonu (Thaher, Heidari, Mafarja, Dong, & Mirjalili, 2020) ve İGKO (Luo & Zhao, 2019) örnek verilebilir.

Bu çalışmada ÖS için beyaz köpek balığı algoritmasının (BKO) ikilileştirilmesi işlemi gerçekleştirilmiştir. BKO, köpek balıklarının avlanma ve avı izleme davranışlarından esinlenerek öne sürülen bir algoritmadır (Braik, Hammouri, Atwan, Al-Betar, & Awadallah, 2022). BKO'da köpek balıkları buldukları durumlara göre farklı hareket davranışları sergilerler. Köpek balıklarının konumları bu davranışlara göre güncellenecek ve böylece optimum çözüme ulaşılabilecektir. İkili beyaz köpekbalığı meta sezgisel algoritmasının bazı avantajlarından aşağıda bahsedilmiştir.

- Sömürü ve keşif arasındaki dengeyi kontrol altında alır. Dolayısıyla küresel en küçük değere yakınsamada, yerel en küçük değerde takılma ihtimalini azaltır.
- Yüksek yakınsama hızıyla zorlu optimizasyon problemlerinde küresel çözüme yakınsayabilmektedir.
- Matematiksel modeli ile yüksek boyutlu, çeşitli mühendislik optimizasyon problemlerini ele almayı uygun hale getirir.

İBKO ise BKO'nun ikilileştirilmiş versiyonunu temsil eder. Burada ikili değerlere dönüştürülmesi S, U ve V-şekilli transfer fonksiyonları ile elde edilmiştir ve bu versiyonlara sırasıyla SBKO, VBKO ve UBKO isimleri verilmiştir. İBKO alt kümeler üzerinden sınıflandırma işlemi yapılarak optimum öznitelik kümesini seçer. Dolayısıyla, sarmal bir öznitelik seçim metodudur.

Çalışmanın geri kalanı şu şekilde organize edilmiştir: Bölüm 2'de BKO'nun matematiksel alt yapısından ve algoritmanın işleyiş mekanizmasından bahsedilecektir. Bölüm 3'te BKO'nun ikilileştirilmesinde kullanılan transfer fonksiyonları açıklanacaktır. Bölüm 4'te elde edilen deneysel sonuçlara, bu sonuçların diğer algoritmalarla karşılaştırılmasına ve sıralama analizine yer verilecektir. Bölüm 5'te elde edilen sonuçlardan, önerilen yöntemin katkısından ve gelecekte yapılması planlanan çalışmalardan söz edilecektir.

2. Beyaz Köpek Balığı Optimizasyon Algoritması (BKO)

2. White Shark Optimization Algorithm (WSO)

Bu kısımda, köpek balıklarının avlanma ve avı izleme davranışlarının temel alan BKO'nun matematiksel modeli üzerinde durulacaktır. Büyük beyaz köpek balıkları, her ne kadar okyanusta avın yerini belirleme yeteneğine sahip olsalar da okyanusun derinlerindeki besin kaynaklarına ulaşabilmek için kapsamlı araştırmalar yapmak zorundadırlar. Bu amaçla köpek balıklarının üç temel hareketi esas alınmıştır. Bunlar:

- (1) Avın hareketi nedeniyle oluşan dalgalara dayalı ava doğru hareket: Burada köpek balığı, işitme ve koklama duyuları ile dalgalı hareketle avına ulaşmaya çalışır.
- (2) Okyanusun derinlerinde rastgele arama hareketi: Bunun için köpek balıkları ava doğru hareket ederek optimal avına yakın durur.
- (3) Optimal ava yakın olan bir avı bulma hareketi: Bu durumda büyük beyaz köpek balığı optimum avına çok yakın olan en iyi beyaz köpek balığına doğru hareket eder.

Bu davranışlara dayanarak, avın uygun şekilde bulunamaması durumunda en iyi çözüme ulaşmak için tüm beyaz köpekbalıklarının konumları güncellenecektir. Matematiksel modelin uygulanışı alt başlıklar altında detaylandırılmıştır.

2.1. BKO'nun başlatılması

2.1. Initialization of WSO

Bütün popülasyon tabanlı algoritmalarda olduğu gibi, BKO'da da ilk çözümler havuzu rastgele olarak oluşturulur. Problemin boyutu d ve köpek balığı popülasyonu ise N olsun. d -boyutlu bir arama uzayında N tane köpek balığı için aday çözümler matrisi ω , denklem 1'deki gibi tanımlanır.

$$\omega = \begin{bmatrix} \omega(1,1) & \cdots & \omega(1,d) \\ \vdots & \ddots & \vdots \\ \omega(N,1) & \cdots & \omega(N,d) \end{bmatrix}_{N \times d} \quad (1)$$

Burada ω , tüm köpek balıklarının pozisyonunu; d , ilgili problem için karar değişkenlerinin boyutunu ve $\omega(i,d)$, i . köpek balığının d . boyuttaki pozisyonunu ifade eder. Başlangıç popülasyonu denklem 2'deki gibi rastgele başlatma ile oluşturulur.

$$\omega(i,j) = low_j + rand \times (up_j - low_j) \quad (2)$$

Burada $\omega(i,j)$, i . köpek balığının j . boyuttaki başlangıç konumunu; low_j ve up_j sırasıyla arama uzayının j . boyutta alt ve üst sınır değerlerini ve $rand$, $[0,1]$ aralığında üretilen rastgele bir sayıyı göstermektedir. Beyaz köpek balığının her yeni konumu için uygunluk fonksiyonu ile çözüm kalitesi değerlendirilir. Eğer yeni konum mevcut konumdan daha iyiyse, mevcut konum yenisiyle güncellenir. Aksi halde mevcut değerini korur.

2.2. Hareket stratejileri

2.2. Movement strategies

Ava doğru hareket: Beyaz köpekbalığı, avının hareketi esnasında algıladığı dalgaların tereddütüne dayanarak dalgalı bir hareketle avına doğru hareket eder. Bu durum denklem 3'teki gibi modellenmiştir.

$$\vartheta_{t+1}^i = \mu \left[\vartheta_t^i + p_1 (\omega_{bestpos_t} - \omega_t^i) \times r_1 + p_2 (\omega_{best}^{\vartheta_t^i} - \omega_t^i) \times r_2 \right] \quad (3)$$

Burada, i , beyaz köpek balıklarının indeksini (1'den N 'ye kadar); ϑ_{t+1}^i , i . köpek balığının t . adımdaki yeni hız vektörünü; ϑ_t^i , t . adımdaki hızı, $\omega_{bestpos_t}$, t . adımda herhangi bir köpek balığının o ana kadarki en iyi konum vektörünü, ω_t^i , i . köpek balığının t . adımdaki konum vektörünü; $\omega_{best}^{\vartheta_t^i}$, sürü tarafından bilinen en iyi konum vektörünü; ϑ^i , denklem 4'te açıkça verildiği üzere en iyi konuma ulaşan beyaz köpekbalıklarının i . indeks vektörünü temsil eder. Ayrıca, r_1 ve r_2 , $[0,1]$ aralığında elde edilen iki rastgele sayı; p_1 ve p_2 , ω_t^i 'nin üzerinde $\omega_{bestpos_t}$ ve $\omega_{best}^{\vartheta_t^i}$ 'nin etkisini kontrol eden parametrelerdir ve sırasıyla denklem 5-6'daki gibi ifade edilirler. Son olarak μ , BKO'da daralma etkisini kontrol eden parametre olarak denklem 7'deki şekilde önerilmiştir.

$$\vartheta = [N \times rand(1, N)] + 1 \quad (4)$$

$$p_1 = p_{max} + (p_{max} - p_{min}) \times e^{-\left(\frac{4T}{t}\right)^2} \quad (5)$$

$$p_2 = p_{min} + (p_{max} - p_{min}) \times e^{-\left(\frac{4T}{t}\right)^2} \quad (6)$$

Denklem 4'teki $rand(1, N)$, $[0,1]$ aralığında normal dağılıma sahip rastgele sayı vektörüdür. Denklem 5-6'da kullanılan t ve T sırasıyla mevcut iterasyonu ve maksimum iterasyon sayısını gösterir. p_{min} ve p_{max} ırasıyla beyaz köpek balıklarının en iyi hareketi için gerekli en küçük ve en büyük hızları temsil eder.

$$\mu = \frac{2}{|2-\tau-\sqrt{\tau^2-4\tau}|} \quad (7)$$

Denklem 7'deki τ hızlanma katsayısıdır ve değeri 4.125 olarak belirlenmiştir.

Optimum ava doğru hareket: Beyaz köpek balıkları, genellikle avın hareketinden kaynaklanan dalgaları duyduklarında veya avın kokusunu aldıklarında ava doğru hareket ederler. Bazı durumlarda ise av ya yiyecek aramak için ya da beyaz bir köpekbalığı ona doğru hareket ettiğinde bulunduğu yeri terk eder. Balık sürüsünün besin kaynağı arama davranışı gibi, beyaz köpekbalığı av aramak için rastgele yerlerde gezinir. Bu bağlamda, söz konusu hareketi tanımlamak için denklem 8'deki konum güncelleme stratejisi kullanılır.

$$\omega_{t+1}^i = \begin{cases} \omega_t^i \cdot \neg \oplus \omega_0 + up \times av + low \times bv ; rand < mv \\ \omega_t^i + \frac{\vartheta_t^i}{f_{wavy}} ; rand \geq mv \end{cases} \quad (8)$$

Burada ω_{t+1}^i , i . beyaz köpek balığının $(t+1)$. adımıdaki yeni konum vektörünü; \neg , değil işlemi; av ve bv , denklem 9 ve 10'daki gibi tanımlanmış 1b ikili vektörleri; low ve up , sırasıyla arama uzayının alt ve üst sınırlarını; ω_0 , denklem 11'de tanımlanan lojik vektörü; f_{wavy} , denklem 12'de gösterildiği gibi tanımlanan beyaz köpek balığının dalgalı hareketinin frekansını; $rand$, $[0,1]$ aralığında üretilen rastgele bir sayıyı ve mv , iterasyon sayısına bağlı olarak köpek balığı avına yaklaştıkça artan hareket kuvvetini temsil eder ve denklem 13'teki gibi tanımlanır.

$$av = \text{sgn}(\omega_t^i - up) > 0 \quad (9)$$

$$bv = \text{sgn}(\omega_t^i - low) < 0 \quad (10)$$

$$\omega_0 = \oplus (av, bv) \quad (11)$$

Denklem 11'de \oplus , XOR işlemine karşılık gelmektedir. Denklem 9 ve 10, arama uzayının tüm muhtemel alanlarına ulaşmayı destekler.

$$f_{wavy} = f_{min} + \frac{f_{max} - f_{min}}{f_{max} + f_{min}} \quad (12)$$

Denklem 12'de f_{min} ve f_{max} , sırasıyla, dalgalı hareketin minimum ve maksimum frekanslarını ifade eder.

$$mv = \frac{1}{(c_1 + e^{(T/2-t)/c_2})} \quad (13)$$

Burada, c_1 ve c_2 , keşif ve sömürü davranışlarını yönetmek için kullanılan iki pozitif sabittir.

En iyi beyaz köpek balığına doğru hareket: Büyük beyaz köpek balıkları, avına en yakın olana karşı konumlarını koruyabilen bir davranış sergilerler. Bu davranış denklem 14'teki gibi hesaplanır.

$$(\omega_{t+1}^i)' = \omega_{bestpos_t} + rnd_1 \times \overrightarrow{Dist_\omega} \times \text{sgn}(rnd_2 - 0.5), rnd_3 < ss \quad (14)$$

Burada, $(\omega_{t+1}^i)'$, i . beyaz köpek balığının avın konumuna göre güncellenmiş konumunu; $\text{sgn}(rnd_2 - 0.5)$ aramanın yönünü değiştirmek için kullanılan ve -1 veya 1 üreten işaret fonksiyonunu; $\overrightarrow{Dist_\omega}$, köpek balığı ile av arasındaki denklem 15'teki gibi ifade edilen uzaklığı ve ss , beyaz köpek balıklarının koku ve görme duyularının gücünü ifade eden parametreyi denklem 16'daki gibi temsil etmektedir.

$$\overrightarrow{Dist_\omega} = |rand \times (\omega_{bestpos_t} - \omega_t^i)| \quad (15)$$

$$ss = |1 - e^{(-cs \times t/T)}| \quad (16)$$

Burada cs , keşif ve sömürü safhalarını kontrol etmede kullanılan pozitif bir sabit olup, bu çalışma için değeri 5×10^{-4} olarak belirlenmiştir.

2.3. Balık sürüsü davranışı

2.3. Fish school behavior

Beyaz köpek balıklarının sürü davranışı denklem 17 ile tanımlanabilir. Bu denkleme göre ilk iki optimum çözüm korunur, diğer köpek balıklarının konumu da bu iki optimum konuma göre güncellenir. Büyük beyaz köpek balıklarının son konumu, optimum avına çok yakın olan arama alanı içinde bir yerde olacaktır. Balık sürüsü davranışı ve beyaz köpek balıklarının en iyi beyaz köpek balığına doğru hareketi, BKO'nun bütüncül davranışını belirler ve bu, daha iyi keşif ve kullanma özelliklerinin kapsamını genişletir.

$$\omega_{t+1}^i = \omega_t^i + (\omega_{t+1}^i)' / (2 \times rand) \quad (17)$$

2.4. Keşif ve sömürü fazı

2.4. Exploration and exploitation phase

BKO'nun matematiksel modelini oluşturan denklemlerden bahsedildi. Buna göre bu denklemler içerisinde kullanılan ve keşif ve sömürü yeteneğini artıran parametreler şunlardır:

μ : BKO'nun yerel optimuma takılmasını önleyerek keşif fazını kontrol eder.

p_1, p_2 : Köpek balıklarının yerel ve global arama esansındaki hızlarını ayarlamak için kullanılır. BKO'nun iterasyonları boyunca bu değerler güncellenir. BKO'nun başarısı üzerinde oldukça etkisi vardır.

mv : İlk iterasyonlarda av ve köpek balığı arasındaki mesafe oldukça fazladır. mv 'ye göre beyaz köpek balıklarının konumu güncellenir. Denklem içerisinde kullanılan c_1 ve c_2 parametreleri sırasıyla sömürü ve keşif kapasitesini artırmayı hedefler.

$(\omega_{t+1}^i)' = \omega_{bestpos_t} + rnd_1 \times \overrightarrow{Dist_\omega} \times sgn(rnd_2 - 0.5)$: Denklem 14'teki $sgn(rnd_2 - 0.5)$, keşfin yönünü kontrol eder. rnd_2 sayesinde, negatif ve pozitiflerin eşit olasılığı olan tutarlı bir dağılım sağlar. rnd_1 ise arama uzayında çözümlerin rastgele davranmasına yardımcı olur

3. İkili Beyaz Köpek Balığı Optimizasyon Algoritması (İBKO)

3. Binary White Shark Optimization Algorithm (BWSO)

Metasezgisel algoritmalarda arama uzayının, popülasyonun her bir bireyi için kayan nokta konum vektörüne sahip sürekli bir uzay olduğu kabul edilir. BKO'nun temel versiyonu da sürekli optimizasyon problemlerini çözüm getirmek için önerilmiştir. Bu problemler temel kalite testi fonksiyonları olabileceği gibi gerçek dünya problemleri de olabilir (Dhiman & Kumar, 2019; Houssein, Saad, Hashim, Shaban, & Hassaballah, 2020; Y. Li, Zhu, & Liu, 2020). BKO'nun ikili versiyonu olan İBKO'da gerçekleştirilen, transfer fonksiyonuna göre sürekli verilerinin ikili değerlere dönüştürülmesidir. Bu çalışmada ise S, V ve U şekilli transfer fonksiyonları ile BKO'nun ikili versiyonları elde edilmiş ve sırasıyla SBKO, VBKO ve UBKO olarak isimlendirilmiştir.

S-şekilli transfer fonksiyonun matematiksel ifadesi denklem 18'de ve grafiği Şekil 2a'da verilmiştir.

$$S(x_i^{dim}(t+1)) = 1 / (1 + e^{-x_i^{dim}(t)}) \quad (18)$$

Burada $S(x_i^{dim}(t+1))$, i . bireyin ikili konum değerinin dim . boyuttaki değişme olasılığını verir. Daha sonra denklem 19'daki gibi eşik değeri ile karşılaştırılarak ikili değer belirlenir. Burada $rand$, normal dağılıma sahip $[0,1]$ aralığında rastgele bir değer değeridir.

$$x_i^{dim}(t+1) = \begin{cases} 1, & rand \leq S(x_i^{dim}(t+1)) \\ 0, & dd \end{cases} \quad (19)$$

V-şekilli transfer fonksiyonun matematiksel ifadesi denklem 20'de ve grafiği Şekil 2b'de verilmiştir.

$$V(x_i^{dim}(t+1)) = \left| \frac{2}{\pi} \arctan \left(\frac{\pi}{2} x_i^{dim}(t) \right) \right| \quad (20)$$

Burada $V(x_i^{dim}(t+1))$, i . bireyin ikili konum değerinin dim . boyuttaki değişme olasılığını verir. İkili değerlerin belirlenmesi denklem 21 ile gerçekleştirilir.

$$x_i^{dim}(t+1) = \begin{cases} 1, & rand \leq V(x_i^{dim}(t+1)) \\ 0, & dd \end{cases} \quad (21)$$

Son olarak U-şekilli transfer fonksiyonunun matematiksel ifadesi denklem 22'de ve grafiği Şekil 2c'de verilmiştir. Burada, $U(x_i^{dim}(t+1))$, i . bireyin ikili konum değerinin dim . boyuttaki değişme olasılığını verir. α ve β kontrol parametreleri olmak üzere; α , transfer fonksiyonunun eğimini ve β , kullanılan transfer fonksiyonunun havuz genişliğini temsil eder.

$$U(x_i^{dim}(t+1)) = \alpha |x_i^{dim}(t)|^\beta \quad (22)$$

İkili değerler ise denklem 23 ile üretilir.

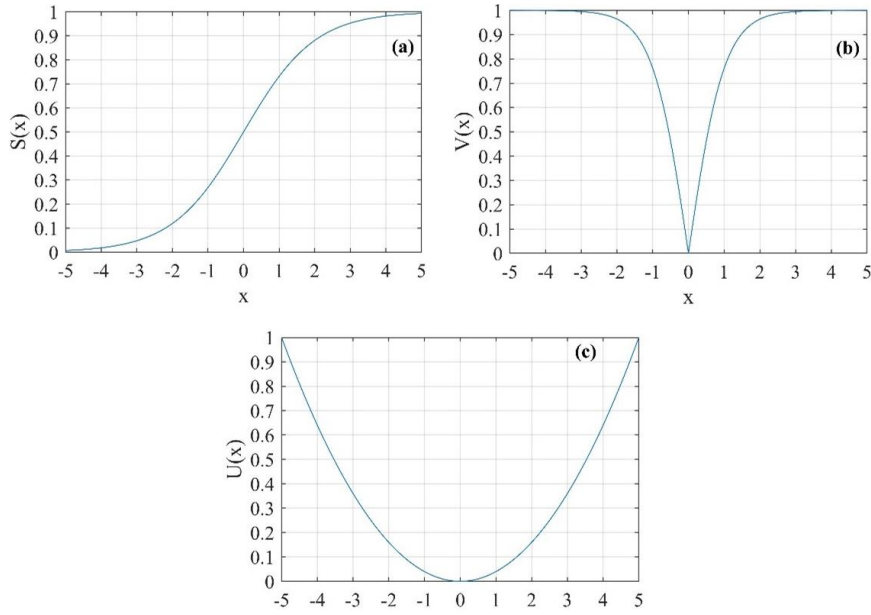
$$x_i^{dim}(t+1) = \begin{cases} 1, & rand \leq U(x_i^{dim}(t+1)) \\ 0, & dd \end{cases} \quad (23)$$

Buna göre BKO'nun ikilileştirilmiş versiyonuna ilişkin algoritma aşağıdaki gibidir.

```

1: Problem parametreleri başlatılır.
2: BKO'nun parametreleri başlatılır.
3: Rastgele başlangıç konumları üretilir.
4: Başlangıç popülasyonun hızları başlatılır ve konumları değerlendirilir.
5: while(t<T) do
6:    $\vartheta$ ,  $p_1, p_2, \mu, av, bv, \omega_0, f_{wavy}, mv, ss$  güncellenir.
7:   for i=1:N do
8:      $\vartheta_{t+1}^i = \mu [\vartheta_t^i + p_1(\omega_{bestpos_t} - \omega_t^i) \times r_1 + p_2(\omega_{best}^{\vartheta_t} - \omega_t^i) \times r_2]$ 
9:   end for
10:  for i=1:N do
11:    if rand < mv then
12:       $\omega_t^i \times \neg \oplus \omega_0 + up \times av + low \times bv$ 
13:    else
14:       $\omega_t^i + \vartheta_t^i / f_{wavy}$ 
15:    end if
16:  end for
17:  for i=1:N do
18:    if rand < ss then
19:       $\overrightarrow{Dist}_\omega = |[(rand \times (\omega)_{bestpos_t} - \omega_t^i)]|$ 
20:      if i==1 then
21:         $(\omega_{t+1}^i) = \omega_{bestpos_t} + rnd_1 \times \overrightarrow{Dist}_\omega \times sgn(rnd_2 - 0.5)$ 
22:      else
23:         $(\omega_{t+1}^i)' = \omega_{bestpos_t} + rnd_1 \times \overrightarrow{Dist}_\omega \times sgn(rnd_2 - 0.5)$ 
24:         $\omega_{t+1}^i = \omega_t^i + (\omega_{t+1}^i)' / (2 \times rand)$ 
25:      end if
26:    end if
27:  end for
28: Değerlendirilir ve pozisyonlar güncellenir.
29: Transfer fonksiyonu (TF) seçilir (S, V veya U).
30:   if rand ≤  $(\omega_{t+1}^i)$  then
31:      $\omega_{t+1}^i = 1$ 
32:   else
33:      $\omega_{t+1}^i = 0$ 
34:   end if
35: end while
36: Optimal sonuçları döndürülür.

```



Şekil 2. Transfer fonksiyonları (a) S-şekilli (b) V-şekilli (c) U-şekilli.
Figure 2. Transfer functions (a) S-shaped (b) V-shaped (c) U-shaped.

4. Deneysel Sonuçlar

4. Experimental Results

Deneysel çalışmalar MATLAB R2016a versiyonunda, AMD Ryzen 5 3500X 6-çekirdek işlemci ve 16 GB RAM belleğe sahip bilgisayarda gerçekleştirilmiştir. Önerilen yöntemler için UCI veri deposundan alınan 8 farklı kalite testi seti kullanılmıştır. Bu verilere ait öznelik sayısı, örnek sayısı, sınıf sayısı, çalışma alanı ve sınıf dağılımı bilgileri Tablo 1’de açıklanmıştır.

Tablo 1. Kullanılan veri setleri ve özellikleri

Table 1. Datasets used and their properties

Veri seti	Öznelik sayısı	Örnek sayısı	Sınıf sayısı	Çalışma alanı	Sınıf dağılımı
Sonar	60	208	2	Biyoloji	%46, %54
Ionosphere	33	351	2	Elektromanyetik	%35, %65
Breast cancer	9	699	2	Biyoloji	%65, %35
Clean1	168	476	2	Yapay	%43, %57
Spect	22	267	2	Biyoloji	%20, %80
Waveform	40	5000	3	Fizik	%34, %33, %33
Satellite	36	6435	6	Fiziksel	%23, %11, %21, %10, %12, %23
Spambase	57	4061	2	Bilgisayar	%69, %31

Her bir veri seti için önerilen üç yaklaşım 30 tekrarla çalıştırılmıştır. İstatistiksel sonuçlar olarak ortalama uygunluk değeri (OUD), ortalama sınıflandırma doğruluğu (OSD), ortalama seçilen örnek sayısı (OSÖS) ve bunların standart sapma değerleri (SS) hesaplanmıştır. Maksimum iterasyon sayısı (epok) 100 olarak belirlenmiş ve sınıflandırıcı olarak k-en yakın komşuluk (k-EYK) sınıflandırıcı 5 komşuluk değeri için uygulanmıştır. Sınıflandırma verileri, %80 eğitim-%20 test verisi olacak şekilde rastgele olarak oluşturulmuştur. Önerilen SBKO, VBKO ve UBKO yöntemleri; İAAO, İPSO (Khanesar, Teshnehlab, & Shoorehdeli, 2007), İHŞO, İGKO, İGAO algoritmaları ile karşılaştırılmıştır. Her bir algoritmanın başlangıç ve temel parametreleri Tablo 2’de verilmiştir.

Tablo 2. Algoritmaların parametre ayarları
Table 2 Parameter settings of the algorithms

Algoritma	Parametre	Değer
İBKO	f_{\min}	0.07
	f_{\max}	0.75
	τ	4.125
	a_0	6.25
	a_1	100
	a_2	0.0005
İAAO	α	50
	β	0.2
İPSO	c_1, c_2	2
	w_{\min}	0.4
İHŞO	w_{\max}	0.9
	β	1.5
İGKO	a	[2 0]

Adil bir karşılaştırma yapabilmek adına her algoritma rastgele olarak başlatılmış ve popülasyon sayısı, maksimum iterasyon ve koşum sayısı algoritmalar arasında eşit tutulmuştur. Uygunluk fonksiyonu parametresi $\alpha=0.99$ olarak alınmıştır. Önerilen algoritmaların başarısını ispatlamak ve elde edilen sonuçların tutarlılığını göstermek için Friedman sıralama testi uygulanmış ve yöntemler kendi aralarında derecelendirilmiştir.

4.1. Önerilen yöntemlerin sayısal sonuçları

4.1. Numerical results of the proposed methods

İlk olarak önerilen S, V ve U-şekilli versiyonlar farklı popülasyon sayıları için ($N=[10, 20, 30, 50]$) uygulanmış ve gelecek işlemlerde kullanılacak olan optimum popülasyon sayısı belirlenmiştir. Popülasyon değerlerine göre S, V ve U-şekilli transfer fonksiyonlarına göre OSD sonuçları sırasıyla Tablo 3-5'te verilmiştir. Tablo 3'te SBKO'ya göre en yüksek OSD doğruluğu tüm veri setleri için en yüksek değerine $N=50$ olduğu durumda ulaşmıştır. Tablo 4'te VBKO için, Breast cancer ve Waveform veri setlerinde ideal popülasyon sayısı $N=30$ iken, diğer veri setleri için yine $N=50$ olarak tespit edilmiştir. Son olarak Tablo 5'te ise UBKO OSD sonuçlarına göre Waveform veri seti için $N=10$, Satellite veri seti için $N=30$ ve geri kalan veri setleri için $N=50$ olarak gözlemlenmiştir.

Tablo 3. SBKO'nun farklı popülasyon boyutları için OSD sonuçları
Table 3. ACA results of the SWSO for different population values

Veri seti	10		20		30		50	
	OSD	$\pm SS$	OSD	$\pm SS$	OSD	$\pm SS$	OSD	$\pm SS$
Sonar	0.8376	0.0215	0.8390	0.0285	0.8372	0.0143	0.8540	0.0146
Ionosphere	0.9059	0.0092	0.9085	0.0060	0.9097	0.0066	0.9140	0.0071
Breast cancer	0.9843	0.0015	0.9842	0.0014	0.9844	0.0012	0.9845	0.0013
Clean1	0.9123	0.0098	0.9181	0.0097	0.9236	0.0104	0.9251	0.0089
Spect	0.8784	0.0121	0.8845	0.0094	0.8890	0.0109	0.8980	0.0118
Waveform	0.8569	0.0054	0.8622	0.0057	0.8532	0.0082	0.8695	0.0037
Satellite	0.9246	0.0034	0.9247	0.0032	0.9252	0.0026	0.9252	0.0017
Spambase	0.9530	0.0038	0.9423	0.0055	0.9436	0.0042	0.9439	0.0052

Tablo 4. VBKO'nun farklı popülasyon boyutları için OSD sonuçları
Table 4. ACA results of the VWSO for different population values

Veri seti	10		20		30		50	
	OSD	$\pm SS$	OSD	$\pm SS$	OSD	$\pm SS$	OSD	$\pm SS$
Sonar	0.8562	0.0210	0.8685	0.0196	0.8807	0.0143	0.8890	0.0227
Ionosphere	0.9355	0.0054	0.9358	0.0061	0.9371	0.0074	0.9477	0.0063
Breast cancer	0.9843	0.0013	0.9838	0.0012	0.9849	0.0012	0.9843	0.0014
Clean1	0.9229	0.0226	0.9468	0.0318	0.9523	0.0313	0.9754	0.0245
Spect	0.8834	0.0187	0.8924	0.0122	0.8969	0.0116	0.9030	0.0097

Tablo 4. Devamı
Table 4. continuing

Waveform	0.8567	0.0100	0.8618	0.0045	0.8635	0.0066	0.8625	0.0056
Satellite	0.9212	0.0030	0.9228	0.0028	0.9220	0.0035	0.9240	0.0020
Spambase	0.9393	0.0045	0.9399	0.0034	0.9412	0.0040	0.9437	0.0029

Tablo 5. UBKO'nun farklı popülasyon boyutları için OSD sonuçları
Table 5. ACA results of the UWSO for different population values

Veri seti	10		20		30		50	
	OSD	±SS	OSD	±SS	OSD	±SS	OSD	±SS
Sonar	0.9456	0.0180	0.9598	0.0159	0.9703	0.0154	0.9754	0.0103
Ionosphere	0.9783	0.0137	0.9840	0.0116	0.9911	0.0057	0.9938	0.0070
Breast cancer	0.9825	0.0026	0.9838	0.0026	0.9844	0.0009	0.9843	0.0011
Clean1	0.9712	0.0200	0.9853	0.0103	0.9862	0.0089	0.9919	0.0083
Spect	0.9262	0.0102	0.9307	0.0114	0.9352	0.0083	0.9421	0.0092
Waveform	0.8592	0.0076	0.8516	0.0036	0.8528	0.0030	0.8551	0.0058
Satellite	0.9202	0.0037	0.9231	0.0027	0.9247	0.0031	0.9245	0.0018
Spambase	0.9361	0.0119	0.9404	0.0130	0.9408	0.0133	0.9421	0.0143

Ancak genel olarak popülasyon boyutlarına göre OSD değerlerine bakıldığında sonuçlar birbirine oldukça yakındır. Örneğin Tablo 4'e göre VBKO için Breast cancer veri setinde, N=30'un N=50'ye göre doğruluk oranı yalnızca %0.04 daha iyidir. Durum Waveform veri seti için de farklı değildir. Tablo 5'te ise Waveform veri seti N=10 değerinde 0.8592 sınıflandırma doğruluğu elde edilmişken, N=50 için bu değer 0.8551'dir ve yine Tablo 4'tekine benzer bir oran söz konusudur. Bu oran çok küçük olduğundan, bundan sonraki karşılaştırmaların tümü N=50 için gerçekleştirilmiştir.

Buna göre N=50 için OSD değerleri Tablo 6'da özetlenmiştir. Burada, Breast cancer, Waveform ve Satellite veri setleri için SBKO ile diğer versiyonlarına göre daha yüksek sınıflandırma doğruluğu elde edildiği sonucuna varılmaktadır. Geri kalan veri setleri için ise UBKO'nun diğer versiyonlarından daha yüksek OSD'ye sahip olduğu görülmektedir.

Tablo 6. Önerilen yöntemlerin OSD değerlerine göre karşılaştırılması
Table 6. Comparison of the proposed methods according to their ACA values

Veri seti	SBKO		VBKO		UBKO	
	OSD	±SS	OSD	±SS	OSD	±SS
Sonar	0.8540	0.0146	0.8813	0.0269	0.9365	0.0554
Ionosphere	0.9140	0.0071	0.9378	0.0154	0.9686	0.0370
Breast cancer	0.9845	0.0013	0.9843	0.0014	0.9843	0.0011
Clean1	0.9251	0.0089	0.9754	0.0245	0.9920	0.0129
Spect	0.9005	0.0094	0.9030	0.0097	0.9297	0.0223
Waveform	0.8695	0.0037	0.8628	0.0059	0.8634	0.0061
Satellite	0.9252	0.0017	0.9240	0.0020	0.9244	0.0019
Spambase	0.9421	0.0042	0.9438	0.0030	0.9451	0.0045

Metasezgisel algoritmaların başarıları uygunluk değerlerine göre yorumlanır. Uygunluk değeri ise aslında bir hata değeridir ve başarılı bir MA'nın düşük hata değeri üretmesi beklenir. Uygunluk değeri, denklem 24'e göre hesaplanır.

$$Fit = \alpha \times err + \beta \frac{|Self|}{|TotF|} \quad (24)$$

Burada; α ve β , sınıflandırma kalitesini ve öznelik indirgemelerini etkileyen iki parametredir. β , $1-\alpha$ değerine eşittir. err , sınıflandırıcıdan elde edilen hata oranıdır. Bu çalışma için k-EYK sınıflandırıcısından elde edilen hata değerine karşılık gelmektedir. $|Self|$, seçilen öznelik sayısını ifade ederken, $|TotF|$ ise toplam öznelik sayısını temsil eder. Buna göre önerilen yöntemler için OUD sonuçları Tablo 7'de rapor edilmiştir. Burada, denklem 24'ten de yorumlanacağı üzere Tablo 6'yı destekleyen uygunluk değerleri üretilmiştir.

Tablo 7. Önerilen yöntemlerin ortalama uygunluk değerlerine (OUD) göre karşılaştırılması
Table 7. Comparison of the proposed methods according to their average fitness values (AFV)

Veri seti	SBKO		VBKO		UBKO	
	OUD	±SS	OUD	±SS	OUD	±SS
Sonar	0.1519	0.0149	0.1242	0.0270	0.0687	0.0556
Ionosphere	0.0917	0.0072	0.0667	0.0161	0.0357	0.0379
Breast cancer	0.0216	0.0000	0.0216	0.0000	0.0216	0.0000
Clean1	0.0810	0.0090	0.0296	0.0250	0.0132	0.0130
Spect	0.1053	0.0096	0.1030	0.0096	0.0763	0.0226
Waveform	0.1376	0.0036	0.1446	0.0055	0.1446	0.0058
Satellite	0.0816	0.0012	0.0827	0.0016	0.0822	0.0019
Spambase	0.0636	0.0040	0.0624	0.0031	0.0616	0.0037

Bir başka karşılaştırma metriği ise seçilen öznitelik sayısıdır. Her ne kadar sonuçlar, önerilen üç versiyon için de sınıflandırma doğruluğu ve uygunluk değerleri açısından tatmin edici olsa da bu değerlerin en az öznitelik ile elde edilmesi de önerilen yöntemin başarısına katkıda bulunacaktır. Tablo 8’de önerilen yöntemlerin OSÖS değerleri verilmiştir.

Tablo 8. Önerilen yöntemlerin OSÖS değerlerine göre karşılaştırılması
Table 8. Comparison of the proposed methods according to the ANSF values

Veri seti	SBKO		VBKO		UBKO	
	OSÖS	±SS	OSÖS	±SS	OSÖS	±SS
Sonar	26.1364	9.1459	25.6364	8.7098	27.0455	7.0405
Ionosphere	16.1250	2.9534	12.9583	6.2169	13.2917	4.8318
Breast cancer	5.2759	1.1618	5.1034	1.2348	5.1379	1.0255
Clean1	88.4615	8.1610	79.6875	10.9467	84.6250	14.3475
Spect	10.2632	2.7252	11.0000	2.1264	11.3333	2.5988
Waveform	22.6667	7.1063	23.7778	5.0936	26.2222	5.2863
Satellite	21.6154	2.7549	21.0556	3.1895	21.0000	3.0870
Spambase	26.1364	9.1459	25.6364	8.7098	27.0455	7.0405

Burada veri setine göre farklı versiyonlarda başarı gözlenmiştir. SBKO ile ortalama olarak Spect veri setinde 22’den 10.2632’e ve Waveform veri setinde 40’tan 22.6667 düzeyinde indirgeme gerçekleşmiştir. VBKO ortalama olarak, Sonar veri setinde 60’tan 25.6364, Ionosphere veri setinde 34’ten 12.9583, Breast cancer veri setinde 9’dan 5.1034, Clean1 veri setinde 166’dan 79.6875 ve Spambase veri setinde 57’den 25.6364 öznitelige indirgeme sağlamıştır. UBKO ise Satellite veri seti için 35 öznitelikten 21 öznitelige indirgeme sağlamış ve bu sayıyı 21.0556 ile VBKO takip etmektedir.

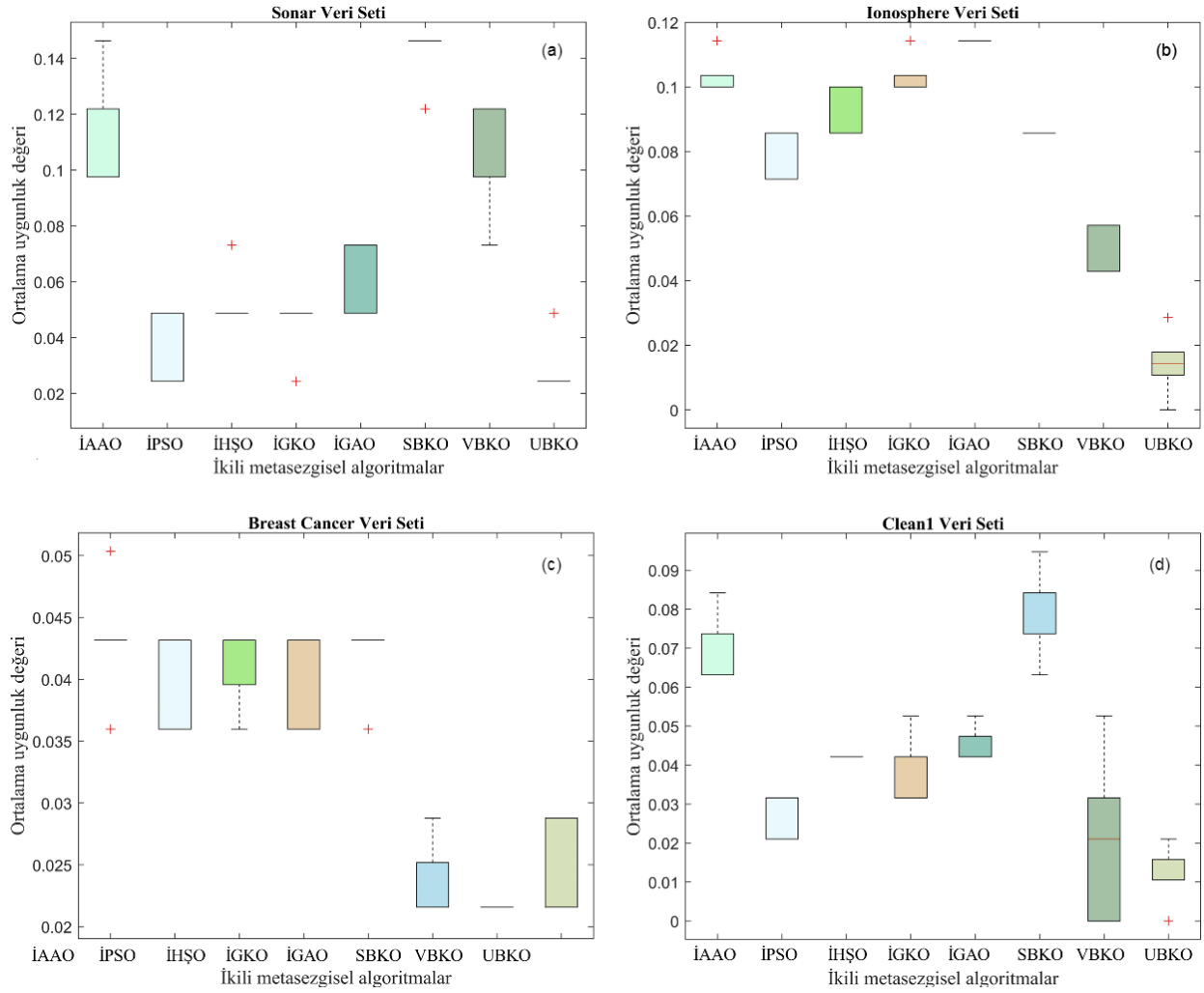
4.2. Önerilen yöntemlerin diğer ikili yöntemlerle karşılaştırılması

4.2. Comparison of the proposed methods with other binary methods

Bu kısımda SBKO, VBKO ve UBKO önerilen yöntemlerinin diğer ikili metasezgisel algoritmalarla karşılaştırılması yapılmıştır. Bu karşılaştırmada kullanılacak olan algoritmalar, literatürde popüler ve bilinen algoritmalar olan ikili atom arama optimizasyonu (İAAO), ikili parçacık sürü optimizasyonu (İPSO), ikili gri kurt (İGKO), ikili Harris’in şahinleri optimizasyonu (İHŞO) ve ikili güve alevi optimizasyonu (İGAO)’dur. Tüm veri setlerinin değerlendirilmesinde k-EYK sınıflandırıcı kullanılmış ve k değeri 5 olarak belirlenmiştir.

Tüm algoritmalar, tüm veri setleri üzerinde daha önceki bölümlerde bahsedilen metriklerle değerlendirilmiştir. Şekil 3-4’te, farklı veri setleri için 8 farklı öznitelik seçim algoritmasının karşılaştırılması grafiksel olarak görselleştirilmiştir. Karşılaştırma metriği olarak OUD kullanılmıştır. Buna göre Şekil 3a’da Sonar veri seti için, UBKO’nun en iyi uygunluk değerine sahip olduğu ve bu değeri İHŞO ve İGKO’nun takip ettiği söylenebilir. Şekil 3b’de Ionosphere için yine UBKO belirgin bir farkla en iyi uygunluk değerine sahiptir. Breast cancer veri seti için Şekil 3c’de önerilen üç versiyonun da başarılı olduğu yorumu yapılabilir. Şekil 3d’ye göre ise Clean1 veri seti için VBKO daha başarılı olmuştur.

Şekil 4a’da Spect veri setinde UBKO’nun belirgin bir uygunluk değerine sahip olduğu açıktır. Şekil 4b, 4c ve 4d için önerilen yöntemlerin birbirine yakın uygunluk değerleri ürettiği ve karşılaştırılan diğer yöntemlerden daha başarılı oldukları söylenebilir.



Şekil 3. Önerilen yöntemlerin farklı ikili metotlarla OUD'ye göre karşılaştırma sonuçları. (a) Sonar veri seti (b) Ionosphere veri seti (c) Breast cancer veri seti (d) Clean1 veri seti.

Figure 3. Comparison results of proposed methods with different binary methods according to ACA. (a) Sonar dataset (b) Ionosphere dataset (c) Breast cancer dataset (d) Clean1 dataset.

Şu ana kadarki ortalama sınıflandırma doğruluğu ve ortama uygunluk değeri sonuçlarına bakıldığında, önerilen yöntemler içerisinde UBKO'nun genel olarak başarılı sonuçlar ürettiği yorumu yapılabilir. Ancak bu yorumun net bir şekilde yapılabilmesi için verilerin sınıf dağılımlarına göre de incelenmesi gerekmektedir. Çünkü dengeli olmayan veriler için tek başına sınıflandırma doğruluğu ve uygunluk değeri yeterli olmayabilir. Bu nedenle mikro ve makro F1 skor değerlerinin de hesaplanması gerekir.

F1 skoru, sınıf başına hesaplanan bir metriktir. Birden fazla sınıf içeren bir veri kümesi için genel F1 skoru, mikro ve makro toplama yöntemleri ile elde edilir. F1 skoru denklem 25 ile, denklem 26 ve 27'yi kullanarak hesaplanır.

$$F1 \text{ Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (25)$$

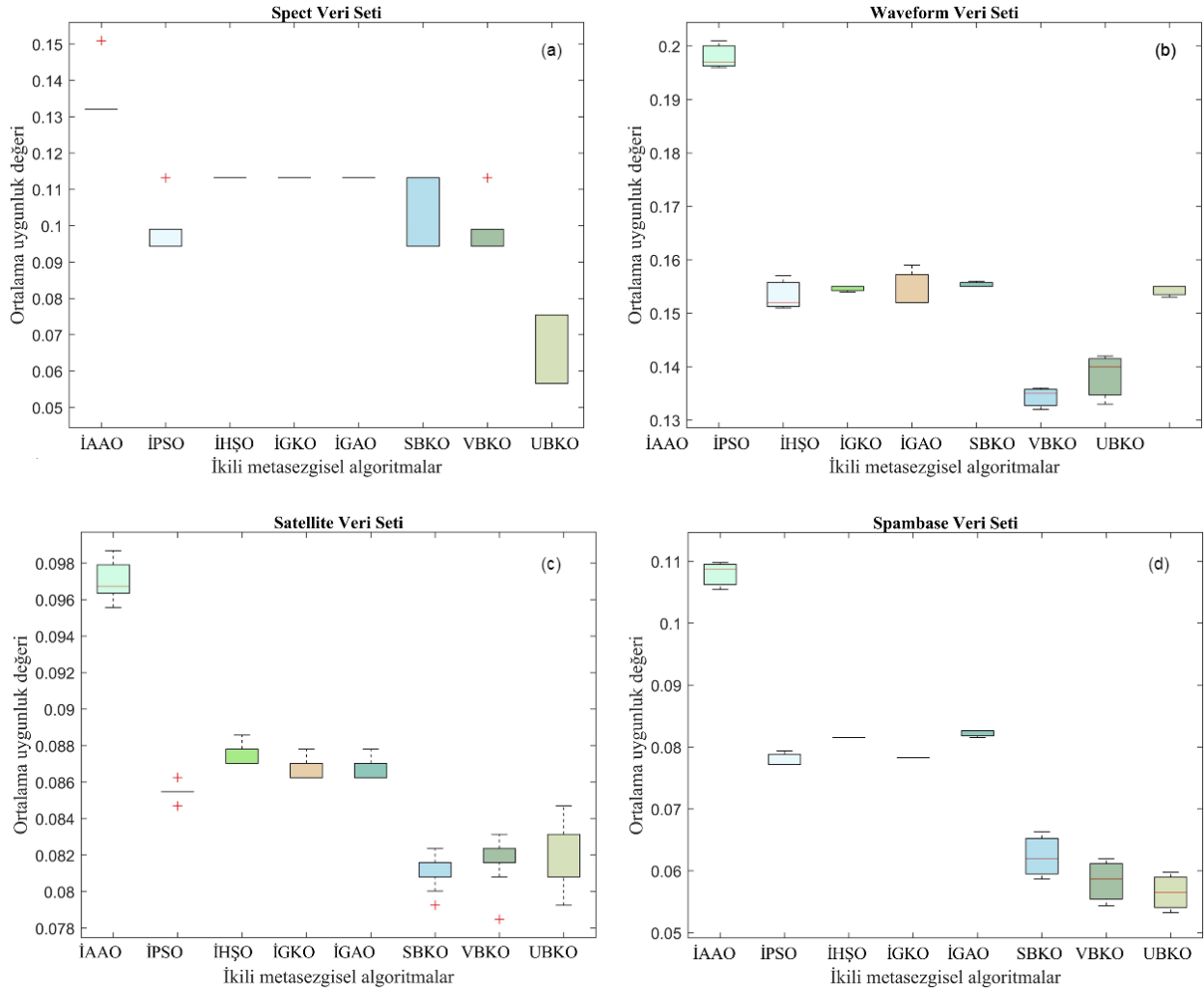
$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (26)$$

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (27)$$

Doğru Pozitif (DP): Pozitif olan verilerin pozitif olarak bulunma oranı

Yanlış Pozitif (YP): Negatif olanların pozitif olarak bulunma oranı

Yanlış Negatif (YN): Pozitif olanların negatif olarak bulunma oranı



Şekil 4. Önerilen yöntemlerin farklı ikili metotlarla OUD'ye göre karşılaştırma sonuçları. (a) Spect veri seti (b) Waveform veri seti (c) Satellite veri seti (d) Spambase veri seti.

Figure 4. Comparison results of proposed methods with different binary methods according to ACA. (a) Spect dataset (b) Waveform dataset (c) Satellite dataset (d) Spambase dataset.

Mikro F1 skoru normal F1 formülüne karşılık gelir ancak her sınıf için ayrı ayrı olmak yerine, toplam DP, YP ve YN değerlerine bağlıdır ve denklem 28'e göre hesaplanır.

$$\text{Mikro F1 Skoru} = \frac{DP}{DP + 0.5 \times (YP + YN)} \quad (28)$$

Makro F1 skoru, sınıf başına hesaplanan F1 puanlarının ağırlıksız ortalamasıdır ve denklem 29'daki gibi hesaplanır.

$$\text{Makro F1 Skoru} = \frac{F1 \text{ skorları toplamı}}{\text{sınıf sayısı}} \quad (29)$$

Tablo 9'da UBKO'nun F1-skoruna göre diğer metasezgisel algoritmalarla karşılaştırılmasına yer verilmiştir. Buna göre, Spect veri setinin mikro ve makro F1-skor değerlerinin arasındaki farkın belirgin ölçüde olduğu ve bu da veri setinin dengeli olmayan veri seti olduğu yorumunu ortaya çıkarır. Mikro ve makro-F1 skorlarına göre, Sonar, Clean1 ve Spambase veri setleri hariç, diğer tüm veri setlerinde en yüksek değerler elde edilmiştir. Sonar ve Spambase için en iyi F1 skorlarını İGKO sağlarken, Clean1 için İAAO sağlamıştır. Ancak, mikro ve makro F1 skorlarının hesaplanması, daha çok Spect veri seti için anlamlıdır ve Spect veri seti için UBKO hem mikro hem de makro F1 skoru açısından diğer yöntemlerden daha yüksek değere sahiptir.

Tablo 9. Mikro ve makro F1 skorlarının karşılaştırılması**Table 9.** Comparison of micro and macro F1 scores

Veri seti	UBKO		İAAO		İPSO		İHŞO		İGKO		İGAO	
	Mikro F1	Makro F1	Mikro F1	Makro F1	Mikro F1	Makro F1	Mikro F1	Makro F1	Mikro F1	Makro F1	Mikro F1	Makro F1
Sonar	0.8390	0.8365	0.8146	0.8131	0.8293	0.8240	0.7805	0.7793	0.9114	0.8913	0.7415	0.7232
Ionosphere	0.8914	0.8750	0.8029	0.7568	0.8857	0.8674	0.8771	0.8308	0.8143	0.7918	0.8229	0.7701
Breast cancer	0.9856	0.9831	0.9453	0.9372	0.9604	0.9588	0.9496	0.9407	0.9496	0.9464	0.9410	0.9294
Clean1	0.9095	0.9092	0.8358	0.8320	0.9158	0.9139	0.8358	0.8343	0.9145	0.8992	0.8526	0.8526
Spect	0.8340	0.7691	0.7547	0.6330	0.7925	0.7307	0.7925	0.7102	0.7170	0.6757	0.7585	0.6803
Waveform	0.8360	0.8360	0.7724	0.7719	0.7945	0.7943	0.7670	0.7663	0.8280	0.8282	0.7078	0.7059
Satellite	0.9223	0.9057	0.8931	0.8725	0.9037	0.8851	0.9040	0.8896	0.9153	0.9029	0.8886	0.8676
Spambase	0.9091	0.8921	0.7996	0.7889	0.9022	0.8983	0.7976	0.7895	0.9342	0.9309	0.8715	0.8621

4.3. Freidman sıralama testi değerlendirmesi

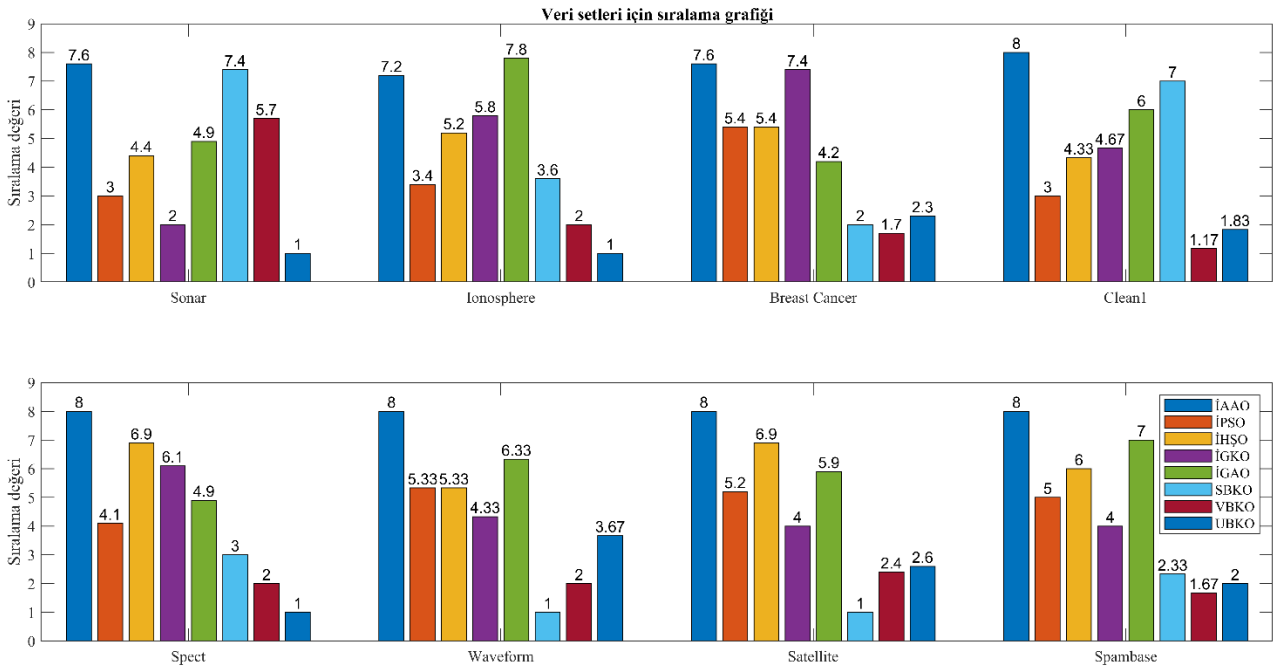
4.3. Evaluation of the Freidman's rank test

Önerilen yöntemlerin önemini kanıtlamak için istatistiksel testlerin yapılması gerekli görülebilir. Bu bağlamda, Friedman sıralama testi kullanılarak, önerilen yöntemlerin karşılaştırılan algoritmalar arasındaki başarıları derecelendirilmiştir. Friedman sıralama testi, ortalama değerler üzerinden sonuçların birbirlerinden anlamlı bir şekilde ayrılıp ayrılmadığını test eden istatistiksel bir yöntemdir (Friedman, 1940).

Tablo 9'da OSD ve SS değerlerine göre Friedman sıralama test sonuçları verilmiştir. Burada SBKO, VBKO ve UBKO, Ionosphere, Breast cancer, Spect, Waveform, Satellite ve Spambase veri setlerinde ilk üçe yerleşmeyi başarmıştır. Clean1 veri setinde ise VBKO ve UBKO versiyonları yüksek sınıflandırma doğruluğu ile sırasıyla 1. ve 2. Sıradadır. Son olarak Sonar veri seti için önerilen yöntemler arasında UBKO'nun başarıları en yüksektir. Bunu İGKO ve İPSO takip etmektedir. Tablo 10'daki gibi hesaplanan sıralama değerleri Şekil 5'teki görselleştirilmiştir. Buradan da önerilen yöntemlerin analiz edilen veri setlerinin büyük çoğunluğunda karşılaştırılan diğer MA'lardan başarılı olduğu yorumu açıkça yapılabilir.

Tablo 10. OSD ve SS değerlerine göre sıralama analizi sonuçları**Table 10.** Rank analysis results according to ACA and SD values

Veri seti		İAAO	İPSO	İHŞO	İGKO	İGAO	SBKO	VBKO	UBKO
Sonar	OSD	0.8690	0.9566	0.9506	0.9578	0.9382	0.8768	0.9117	0.9768
	±SS	0.0136	0.0005	0.0105	0.0003	0.0099	0.0091	0.0120	0.0020
	Sıralama	7.60	3.00	4.40	2.00	4.90	7.40	5.70	1.00
Ionosphere	OSD	0.8998	0.9266	0.9094	0.9028	0.8914	0.9203	0.9533	0.9859
	±SS	0.0061	0.0071	0.0067	0.0056	0.0003	0.0002	0.0059	0.0100
	Sıralama	7.20	3.40	5.20	5.80	7.80	3.60	2.00	1.00
Breast cancer	OSD	0.9605	0.9677	0.9677	0.9617	0.9695	0.9860	0.9862	0.9857
	±SS	0.0027	0.0000	0.0000	0.0000	0.0010	0.0006	0.0005	0.0006
	Sıralama	7.60	5.40	5.40	7.40	4.20	2.00	1.70	2.30
Clean1	OSD	0.9276	0.9734	0.9631	0.9580	0.9539	0.9323	0.9981	0.9963
	±SS	0.0052	0.0002	0.0003	0.0057	0.0004	0.0003	0.0033	0.0032
	Sıralama	8.00	3.00	4.33	4.67	6.00	7.00	1.17	1.83
Spect	OSD	0.8696	0.8968	0.8918	0.8924	0.8932	0.9118	0.9128	0.9495
	±SS	0.0079	0.0073	0.0007	0.0005	0.0002	0.0006	0.0007	0.0006
	Sıralama	8.00	4.10	6.90	6.10	4.90	3.00	2.00	1.00
Waveform	OSD	0.80981	0.85395	0.85355	0.85463	0.85305	0.86811	0.86357	0.85843
	±SS	0.00223	0.00381	0.00075	0.00371	0.00089	0.00303	0.00105	0.00454
	Sıralama	8.00	5.33	5.33	4.33	6.33	1.00	2.00	3.67
Satellite	OSD	0.9083	0.9201	0.9192	0.9213	0.9196	0.9261	0.9243	0.9239
	±SS	0.0009	0.0005	0.0008	0.0004	0.0008	0.0019	0.0011	0.0017
	Sıralama	8.0	5.2	6.9	4.0	5.9	1.0	2.4	2.6
Spambase	OSD	0.8971	0.9287	0.9263	0.9300	0.9251	0.9431	0.9463	0.9440
	±SS	0.0014	0.0013	0.0005	0.0004	0.0007	0.0045	0.0047	0.0037
	Sıralama	8.00	5.00	6.00	4.00	7.00	2.33	1.67	2.00



Şekil 5. Önerilen ve karşılaştırılan ikili algoritmaların veri setlerine göre sıralama değerleri
Figure 5. Rank values of proposed and compared binary algorithms according to data sets

4.4. Çalışma zamanı analizi

4.4. Run time analysis

Örnek olarak Ionosphere veri seti için 1000 iterasyon ve 50 popülasyonda önerilen yöntem ile diğer ikili metasezgisel algoritmalar, çalışma süresi açısından karşılaştırılmış ve Tablo 11 elde edilmiştir. Buna göre, önerilen yöntemlerden UBKO'nun genel olarak diğer ikili yöntemlerden daha hızlı çalıştığı söylenebilir ve bu süre 11.947 s olarak ölçülmüştür. Aynı zamanda sınıflandırma doğruluğu olarak da UBKO'nun genellikle yüksek sonuçlar ürettiği de Tablo 6'da belirtilmiştir. Bununla birlikte, UBKO'nun SBKO ve VBKO'dan da daha kısa sürede sonuç ürettiği açıktır. Çünkü, S- ve V-şekilli transfer fonksiyonlarında, U-şekilli transfer fonksiyonuna göre hesaplaması uzun süren fonksiyonlar kullanılmaktadır. S-şekilli transfer fonksiyonu doğal logaritmik (e^x) ifadeler içerirken, V-şekilli transfer fonksiyonu ise trigonometrik ifadeler barındırır.

Tablo 11. Ionosphere veri seti için ortalama çalışma süreleri
Table 11. Average run times for the Ionosphere dataset

Yöntem	Ortalama çalışma süresi (s)
SBKO	12.420
VBKO	12.781
UBKO	11.947
İAAO	12.500
İPSO	12.751
İHŞO	19.221
İGKO	13.381
İGAO	12.026

5. Tartışma ve sonuçlar

5. Discussion and conclusions

Çalışmada metasezgisel bir algoritma olan beyaz köpek balığı algoritmasının, makine öğrenmesinin önemli bir adımı olan öznelik seçimi sürecinde kullanılmak üzere ikili versiyonları geliştirilmiştir. Bu işlem transfer fonksiyonları denilen matematiksel ifadelerle gerçekleştirilir. Bu çalışma için S, V ve U-şekilli transfer fonksiyonları uygulanmıştır. Geliştirilen versiyonlar, UCI veri deposundan alınan Sonar, Ionosphere, Breast cancer, Clean1, Spect, Waveform, Satellite ve Spambase veri setlerinin öznelik seçiminde test edilmiştir. Değerlendirme metriği olarak ortalama sınıflandırma doğruluğu, ortalama uygunluk değeri,

ortalama seçilen öznitelik sayısı ve bunların standart sapma değerleri hesaplanmıştır. Ayrıca, verilerin sınıf oranını da dikkate alan bir metrik olan F1 skoru da tüm veri setlerine uygulanmıştır. Elde edilen sonuçlar, literatürdeki beş farklı MA ile karşılaştırılmış ve Freidman sıralama testi ile sonuçlar doğrulanmıştır.

Buna göre önerilen SBKO, VBKO ve UBKO yöntemlerinin genel olarak sekiz veri setinde de doğruluk ve uygunluk değeri açısından başarıya ulaştığı söylenebilir. Ancak bunlar arasında en tutarlı ve tüm veri setleri için başarı sağlayan UBKO olmuştur. Öznitelik indirgemesi açısından ise VBKO diğer versiyonlarına göre daha az öznitelikle etkin bir doğruluğa ulaşmıştır. Bununla birlikte, dengeli olmayan veri setleri için uygunluk değeri ve sınıflandırma doğruluğu metrikleri tek başına yeterli olmayacağından, sınıf oranına bağlı hesaplanan bir metrik olan mikro ve makro F1 skorları da incelenmiştir. Burada da genel olarak UBKO'nun en yüksek F1 skorlarına sahip olduğu görülmüştür. Özellikle dengeli olmayan Spect veri setinde, yine en yüksek mikro ve makro F1 skoru sonucu UBKO ile elde edilmiştir.

Önerilen yöntem, çalışma süreleri açısından da diğer metasezgisellerle karşılaştırılmıştır. Buna göre, UBKO'nun çalışma süresinin SBKO ve VBKO ile birlikte, diğer metasezgisel algoritmalarından daha kısa olduğu görülmüştür. Bunun nedeni, U-şekilli transfer fonksiyonun hesaplama süresi kısa olan fonksiyonlardan oluşmasıdır. Dezavantaj olarak ise, önerilen yöntem sarmal bir öznitelik seçim yöntemi olduğundan, bilgi kazanımı, ReliefF gibi geleneksel filtre tabanlı yöntemlere göre daha yavaş çalışmaktadır.

Tüm sonuçlar göz önüne alındığında, gelecek çalışmalar için S, V ve U şekilli transfer fonksiyonlarının farklı MA'larda ve bunların geliştirilmiş ve hibrit versiyonlarında da kullanılabilirliği mümkün görülmektedir. Aynı zamanda yüksek örnek ve öznitelik sayısına sahip veri setlerinde (büyük veri kümelerinde) de uygulanabilir hale getirilmesi gelecek çalışma planları arasındadır. Ayrıca, yeni modifiye transfer fonksiyonları önerilip, MA'ların ikilileştirilmesinde kullanılabilir ve öznitelik seçimi için yeni ikili yöntemler literatüre kazandırılabilir.

Yazar katkısı

Author contribution

Funda KUTLU ONAY: literatür taraması, metodoloji, deneysel çalışmalar, ilgili tabloların/şekillerin hazırlanması ve yorumlanması, bulgular ve sonuç kısmının şekillendirilmesi, makalenin düzenlenmesi ve revizyon süreçlerinin yürütülmesi.

Etik beyanı

Declaration of ethical code

Bu makalenin yazarları, bu çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve / veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı

Conflicts of interest

Yazarlar herhangi bir çıkar çatışması olmadığını beyan eder.

Kaynaklar

References

- Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Metaheuristic algorithms: A comprehensive review. *Computational intelligence for multimedia big data on the cloud with engineering applications*, 185-231. doi:https://doi.org/10.1016/B978-0-12-813314-9.00010-4
- Abualigah, L., Diabat, A., Mirjalili, S., Abd Elaziz, M., & Gandomi, A. H. (2021). The arithmetic optimization algorithm. *Computer Methods in Applied Mechanics*, 376, 113609. doi:https://doi.org/10.1016/j.cma.2020.113609
- Arora, S., & Anand, P. (2019). Binary butterfly optimization approaches for feature selection. *Expert Systems with Applications*, 116, 147-160. doi:https://doi.org/10.1016/j.eswa.2018.08.051

- Awadallah, M. A., Hammouri, A. I., Al-Betar, M. A., Braik, M. S., & Abd Elaziz, M. (2022). Binary Horse herd optimization algorithm with crossover operators for feature selection. *Computers in Biology Medicine*, *141*, 105152. doi:<https://doi.org/10.1016/j.compbiomed.2021.105152>
- Bäck, T., & Schwefel, H.-P. (1993). An overview of evolutionary algorithms for parameter optimization. *Evolutionary computation*, *1*(1), 1-23. doi:<https://doi.org/10.1162/evco.1993.1.1.1>
- Braik, M., Hammouri, A., Atwan, J., Al-Betar, M. A., & Awadallah, M. A. J. K.-B. S. (2022). White Shark Optimizer: A novel bio-inspired meta-heuristic algorithm for global optimization problems. *243*, 108457. doi:<https://doi.org/10.1016/j.knosys.2022.108457>
- Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, *1*(1-4), 131-156. doi:[https://doi.org/10.1016/S1088-467X\(97\)00008-5](https://doi.org/10.1016/S1088-467X(97)00008-5)
- Dehghani, M., Montazeri, Z., Dehghani, A., Malik, O. P., Morales-Menendez, R., Dhiman, G., Nouri, N., Ehsanifar, A., Guerrero, J. M., & Ramirez-Mendoza, R. A. (2021). Binary spring search algorithm for solving various optimization problems. *Applied Sciences*, *11*(3), 1286. doi:<https://doi.org/10.3390/app11031286>
- Dhiman, G., & Kumar, V. (2019). Seagull optimization algorithm: Theory and its applications for large-scale industrial engineering problems. *Knowledge-based systems*, *165*, 169-196. doi:<https://doi.org/10.1016/j.knosys.2018.11.024>
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification*, Hoboken. In: NJ: Wiley.
- Fan, Q., Chen, Z., & Xia, Z. (2020). A novel quasi-reflected Harris hawks optimization algorithm for global optimization problems. *Soft Computing*, *24*(19), 14825-14843. doi:<https://doi.org/10.1007/s00500-020-04834-7>
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, *11*(1), 86-92. doi:<https://doi.org/10.1214/aoms/1177731944>
- Grabczewski, K., & Jankowski, N. (2005). *Feature selection with decision tree criterion*. Fifth International Conference on Hybrid Intelligent Systems (HIS'05) (ss. 6 pp.).
- Hichem, H., Elkamel, M., Rafik, M., Mesaaoud, M. T., & Ouahiba, C. (2019). A new binary grasshopper optimization algorithm for feature selection problem. *Journal of King Saud University-Computer Information Sciences*. doi:<https://doi.org/10.1016/j.jksuci.2019.11.007>
- Houssein, E. H., Saad, M. R., Hashim, F. A., Shaban, H., & Hassaballah, M. (2020). Lévy flight distribution: A new metaheuristic algorithm for solving engineering optimization problems. *Engineering Applications of Artificial Intelligence*, *94*, 103731. doi:<https://doi.org/10.1016/j.engappai.2020.103731>
- Hussien, A. G., Hassanien, A. E., Houssein, E. H., Amin, M., & Azar, A. T. (2020). New binary whale optimization algorithm for discrete optimization problems. *Engineering Optimization*, *52*(6), 945-959. doi:<https://doi.org/10.1080/0305215X.2019.1624740>
- Jh, H. (1975). Adaptation in natural and artificial systems. *Ann Arbor*.
- Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2014). A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, *42*(1), 21-57. doi:<https://doi.org/10.1007/s10462-012-9328-0>
- Khanesar, M. A., Teshnehlab, M., & Shoorehdeli, M. A. (2007). *A novel binary particle swarm optimization*. 2007 Mediterranean conference on control & automation (ss. 1-6).
- Kittler, J. (1978). Feature set search algorithms. *Pattern recognition signal processing*, 41-60.
- Li, S., Chen, H., Wang, M., Heidari, A. A., & Mirjalili, S. (2020). Slime mould algorithm: A new method for stochastic optimization. *Future Generation Computer Systems*, *111*, 300-323. doi:<https://doi.org/10.1016/j.future.2020.03.055>
- Li, Y., Zhu, X., & Liu, J. (2020). An improved moth-flame optimization algorithm for engineering problems. *Symmetry*, *12*(8), 1234. doi:<https://doi.org/10.3390/sym12081234>

- Long, W., & Xu, S. (2016). A novel grey wolf optimizer for global optimization problems. 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (ss. 1266-1270).
- Luo, K., & Zhao, Q. (2019). A binary grey wolf optimizer for the multidimensional knapsack problem. *Applied Soft Computing*, 83, 105645. doi:<https://doi.org/10.1016/j.asoc.2019.105645>
- Mirjalili, S. (2016). SCA: a sine cosine algorithm for solving optimization problems. *Knowledge-based systems*, 96, 120-133. doi:<https://doi.org/10.1016/j.knosys.2015.12.022>
- Nadimi-Shahraki, M. H., Banaie-Dezfouli, M., Zamani, H., Taghian, S., & Mirjalili, S. (2021). B-MFO: a binary moth-flame optimization for feature selection from medical datasets. *Computers in Biology*, 10(11), 136.
- Pal, M., & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience Remote Sensing* 48(5), 2297-2307. doi:<https://doi.org/10.1109/TGRS.2009.2039484>
- Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization. *Swarm intelligence*, 1(1), 33-57. doi:<https://doi.org/10.1007/s11721-007-0002-0>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- Robnik-Šikonja, M., & Kononenko, I. J. M. I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. 53(1), 23-69.
- Saeys, Y., Inza, I., & Larranaga, P. J. b. (2007). A review of feature selection techniques in bioinformatics. 23(19), 2507-2517. doi:<https://doi.org/10.1093/bioinformatics/btm344>
- Siedlecki, W., & Sklansky, J. (1993). On automatic feature selection. In *Handbook of pattern recognition and computer vision* (ss. 63-87): World Scientific.
- Taghian, S., & Nadimi-Shahraki, M. H. (2019). Binary sine cosine algorithms for feature selection from medical data. *Advanced Computing: An International Journal (ACIJ)*, 10. doi:<https://doi.org/10.5121/acij.2019.10501>
- Thaher, T., Heidari, A. A., Mafarja, M., Dong, J. S., & Mirjalili, S. (2020). Binary Harris Hawks optimizer for high-dimensional, low sample size feature selection. In *Evolutionary machine learning techniques* (ss. 251-272): Springer.
- Too, J., & Rahim Abdullah, A. (2020). Binary atom search optimisation approaches for feature selection. *Connection Science*, 32(4), 406-430. doi:<https://doi.org/10.1080/09540091.2020.1741515>