



# Büyük Veri Optimizasyonu için Kaynak-Bağlantılı Harmoni Arama Algoritmasının Performans Analizi

## Performance Analysis of Source-Linked Harmony Search Algorithm for Big Data Optimization

Serhat Celil İLERİ  
Ondokuz Mayıs Üniversitesi  
Bilgisayar Mühendisliği Bölümü  
Samsun, Türkiye  
celil.ileri@bil.omu.edu.tr  
ORCID: 0000-0002-0259-0791

Selçuk ASLAN  
Erciyes Üniversitesi  
Uçak Mühendisliği Bölümü  
Kayseri, Türkiye  
selcukaslan@erciyes.edu.tr  
ORCID: 0000-0002-9145-239X

Sercan DEMİRCİ  
Ondokuz Mayıs Üniversitesi  
Bilgisayar Mühendisliği Bölümü  
Samsun, Türkiye  
sercan.demirci@omu.edu.tr  
ORCID: 0000-0001-6739-7653

### Öz

Bu çalışmada, Harmoni Arama algoritmasının (Harmony Search algorithm, HSA) mevcut veriden faydalanarak başlangıç çözümlerini üretme yaklaşımı ile güçlendirilmiş varyantı olan Kaynak-Bağlantılı Harmoni Arama algoritmasının (Source-Linked HSA, slinkHSA) performansı elektroensefalografi (EEG) sinyallerinde gürültü minimizasyonu gerektiren büyük veri optimizasyonu üzerinden incelenmiştir. slinkHSA ile elde edilen sonuçlar diğer meta-sezgisel teknikler tarafından bulunan sonuçlar üzerinden kıyaslanmıştır. Karşılaştırmalar, başlangıç harmonilerini EEG sinyalleri kullanılarak üretmenin çözümlerinin kalitesini önemli ölçüde katkıda bulunduğunu ve algoritmanın yakınsama hızını artırdığını göstermiştir.

**Anahtar sözcükler:** Harmoni Arama Algoritması, Kaynak-Bağlantılı HSA, Büyük Veri Optimizasyonu, EEG

### Abstract

In this study, the performance of Source-Linked Harmony Search algorithm (slinkHSA) that is a new variant of the Harmony Search algorithm (HSA) powered with the usage of existing data for generating initial solutions was investigated over a big data optimization problem requiring minimization

of measurement noise for electroencephalography (EEG) signals. The results obtained by the mentioned HSA variant were also compared to the results of other meta-heuristic techniques. Comparative studies showed that generating initial harmonies by guiding the existing EEG signals significantly contributes to the qualities of the solutions and increases the convergence speed of the algorithm.

**Keywords:** Harmony Search Algorithm, Source-linked Harmony Search Algorithm, Big Data Optimisation, EEG

### 1. Giriş

Haberleşme ve iletişim teknolojilerinde yaşanan hızlı gelişmeler, veri depolama maliyetlerinin düşmesi ve kapasite artışını da beraberinde getirmiştir. Günlük hayatın vazgeçilmezi haline gelen akıllı telefon, tablet gibi cihazların yanı sıra günlük olarak kullandığımız birçok eşya da üzerinde barındırdığı duyargalar yardımıyla akıllı hale gelerek insan yaşamını büyük oranda değiştirmiştir. Yaşanan bu gelişmelerin sonucu olarak birçok kurum ve kuruluşta çalışma düzeni de aşamalı olarak büyük oranda değişime uğramıştır. Bahsedilen teknolojik gelişmelerle birlikte, organizasyonların birçoğu müşterilerinin verilerini kolaylıkla elde etme ve ilerleyen zamanlarda analiz etmek üzere veriyi depolama imkânlarını elde etmiştir [1]. Yakın zamana kadar veri tabanları belirli ölçütlere uyan sınırlı sayıda günümüze oranla çok daha az sayıda veriyi depolamak için kullanılmıştır.

Günümüzde ise çok çeşitli kaynaklardan sürekli ve yüksek miktarda veri toplanmakta, işlenmekte ve buradan elde edilen bilgi ile insandan bağımsız karar verme mekanizmalarını da içerisinde barındıran pek çok akıllı sistem geliştirilmektedir. Bu denli hızla büyüyen veriler hem depolama alanı hem de veri işleme hızlarının sınırlarını zorlayacak seviyelere kolaylıkla ulaşabilmektedir. Bu nedenle geleneksel veri tabanlarında önceden belirlenmiş tablolar ve sütunlar ile saklanan veri kümeleri ile az önce bahsettiğimiz hızla büyüyen veri kümelerini birbirinden ayırtmak için geçtiğimiz yıllarda "büyük veri" adı verilen bir kavram ortaya çıkmıştır.

Bir veri yığınının büyük veri olarak ele alınabilmesi için "hacim" (volume), "hız" (velocity), "çeşitlilik" (variety), "doğruluk" (veracity) ve "değer" (value) olmak üzere İngilizce karşılıklarının her birisi v harfi ile başlaması nedeni ile 5V olarak isimlendirilen özellikleri taşıması beklenmektedir. Hacim kavramı verinin çok yüksek sayılarda olmasını ve giderek daha da büyümesini ifade eder, bu da veri depolama için gerekli alanla eşdeğerdir. Hız kavramı verinin ilgili kaynaklardan toplanma hızını, çeşitlilik kavramı ise verinin farklı kaynaklardan toplanıyor olmasını ifade eder. Doğruluk kavramı da toplanılan verilerin ve veri kaynağının kalitesini, kullanılabilirliğini ve doğruluğunu ifade etmek için kullanılır. Değer kavramı ise toplanması, depolanması ve işlenmesi için geleneksel yöntemlerin dışına çıkılarak yürütülen sürecin sonucunda elde edilen verinin değerini ifade etmektedir [2]. Büyük veri ile ilişkili problemler sağlık, lojistik, bankacılık, hava trafiği, otomotiv endüstrisi, petrol ve gaz tüketimi, telekomünikasyon, turizm, nakliye, finans, sosyal medya ve çevrimiçi servisler, eğitim ve ticaret gibi çeşitli alanlarda kullanılmaktadır [3,4]. Geçtiğimiz yıllarda Abbass ve arkadaşları tarafından ortaya atılan yeni bir büyük veri optimizasyon problemi araştırmacıların ilgisini çekmiş ve birçok meta-sezgisel algoritma kullanarak mevcut yöntemlerle çözüm üretilmeye çalışılmıştır [5-7]. Abbass ve arkadaşları tarafından ortaya konulan problem Elektroensefalografi (EEG) sinyallerini iki parçaya ayırmaktadır. Bu parçalardan ilki beyin aktivitelerini elde etmek için gerekli bilgiyi sağlamak üzere seçilen kaynak sinyale benzeyen bir sinyali ifade eder. İkinci parça ise tarama sırasında sinyalde meydana gelen gürültülerle eşleştirilir.

Zhang ve arkadaşları tarafından yapılan çaprazlama ve mutasyon operatörlerini yeniden modelleyerek yeni bir popülasyon bazlı teknik sundukları çalışma sinyal ayrıştırma tabanlı büyük veri problemlerini meta-sezgisel olarak çözen ilk çalışmalardandır [8]. Ayrıca bu çalışmalarını gradyan tabanlı yerel arama yöntemi ve sinyal ayrıştırma kullanarak iyileştirmişler, MOME/D olarak adlandırılan çok amaçlı ayrıştırmalı memetik algoritmayı sunmuşlardır [9]. Geliştirdikleri bu yöntem EEG sinyalleri temelli büyük veri optimizasyon problemlerinde Genetik Algoritma (GA) ve Diferansiyel Gelişim (DE) algoritmalarından daha başarılı sonuçlar ortaya koymuştur. Elsayed ve Sarker tarafından geliştirilen otomatik diferansiyel gelişim sisteminde (ADEF) birçok diferansiyel gelişim varyantı birlikte koşturularak belirli bir iterasyon sonucunda tüm popülasyonlardaki en iyi bireylerin seçilerek en uygun diferansiyel gelişim varyantı ile

işleme devam edilmektedir [10-11]. Elsayed ve Sarker tarafından geliştirilen bu yöntem altı farklı büyük veri probleminde çeşitli diferansiyel gelişim varyantlarından önemli ölçüde daha iyi sonuçlar elde etmiştir. El Majdouli ve arkadaşları tarafından Havai Fişek (FW) algoritması kullanılarak EEG sinyalleri temelli büyük veri problemleri çözülmeye çalışılmış, probleme aday çözümler üretilmesi sırasında parametrelerin alabileceği alt ve üst sınırlar arasında rastgele değerler seçmek yerine doğrudan EEG sinyalleri kullanılarak aday çözüm değerleri üretilmiştir. Kullanılan bu teknik ile FW algoritmasının geleneksel yöntemlerle koşturulmasına kıyasla çok daha hızlı şekilde iyi sonuçlar üretmeye başladığı görülmüştür [12]. Loukdache ve arkadaşları tarafından EEG sinyalleri bazlı büyük veri problemlerinin çözümü için bir Klonal Seçim algoritması önerilmiş, algoritmalarında El Majdouli ve arkadaşları tarafından tanıtılan popülasyon oluşturma tekniğini kullanmışlardır [13].

Meselhi ve arkadaşları grafik işleme birimi üzerinde diferansiyel gelişim bazlı bir algoritma kullanarak EEG sinyallerinden gürültüleri temizlemeye çalışmışlardır [14]. Wang ve arkadaşları Ateşböceği algoritması (FA) ile EEG sinyalleri bazlı büyük veri optimizasyon problemlerini çözmeye çalışmışlardır [15]. Yi ve arkadaşları tarafından genetik algoritma kullanılarak geliştirilen NSGA-III olarak adlandırılan yöntem ile EEG sinyali tabanlı büyük veri problemleri çözülmeye çalışılmıştır [16]. Aslan tarafından yapılan çalışmada Yapay Arı Kolonisi (ABC) algoritmasının EEG sinyali tabanlı büyük veri optimizasyon problemine özgü varyantı geliştirilmiştir. slinkABC olarak adlandırılan yeni ABC varyantı ile yapılan denemelerde yeni önerilen yöntemin diğer ABC varyantlarından ve karşılaştırılan diğer algoritmalarından büyük oranda daha iyi sonuç verdiği görülmüştür [17].

Bu çalışmada ise Harmoni Arama Algoritmasının EEG sinyalleri temelli büyük veri probleminin çözümü ve algoritmanın harmoni bestelenmesi ve güncellenmesi sırasında diğer algoritmalarda uygulandığında başarıyı büyük oranda artıran bir yöntem olan kaynak sinyalden faydalanarak çözümler üretecek şekilde güncellenmiş yeni bir sürümü olan kaynak bağlantılı HSA (slinkHSA) tanıtılacaktır. Çalışmanın ilerleyen kısımlarında Bölüm-2'de EEG sinyali bazlı büyük veri problemleri anlatılacak, Bölüm-3'te Harmoni Arama Algoritması açıklanacak, Bölüm-4'te yeni önerilen slinkHSA'da yapılan geliştirmelerden bahsedilecektir. Elde edilen deneysel sonuçlar ve diğer algoritmalarla karşılaştırmalar Bölüm-5'te yapılacak, Bölüm-6'da sonuçlar tartışılarak ileride yapılabilecek çalışmalardan bahsedilecektir.

## 2. EEG Sinyali Temelli Büyük Veri Optimizasyon Problemleri

EEG sinyalleri, beynin değişik uyarılara karşı verdiği tepkilerin zamansal olarak belirli bir sıklıkta kaydedilmesiyle oluşur. Klinik çalışmalarda önemli bir yer tutan bu sinyaller, ölçüm sırasında dış etkenler de dahil olmak üzere birçok gürültü kaynağından gelen gürültü sinyalini de içerisinde barındırmaktadır [18]. Abbass ve arkadaşları tarafından önerilen bir büyük veri optimizasyonu probleminde EEG

sinyallerindeki gürültülerin temizlenmesi bir optimizasyon problemi olarak ele alınmıştır [5-7]. Saniyelik olarak ikili formattaki 20 KB (Kilobayt) veri ve 0,5 Mb (Megabit) boyutundaki veri alınmakta ve işlenerek optimize edilmek istenilmektedir. Ölçümler sırasında bir saniyelik aralıkta alınan veri uzunluğu 256 olan bağımsız zaman serileri şeklinde gruplandırılmaktadır. Dört, on iki ve on dokuz örneğe sahip olan veriler sırasıyla D4, D12 ve D19 olarak adlandırılmaktadır. Ayrıca kaynak zaman serilerine ek gürültü eklenerek değiştirilmiş sinyaller de yine sırayla D4 sinyali için D4N, D12 sinyali için D12N ve D19 sinyali için D19N olarak adlandırılmaktadır. D4 ve D4N örnekleri altı farklı sinyal kaynağından 256 Hz sıklıkta alınan örneklerden oluşmaktadır. Altı farklı sinyal kaynağından dört tanesi EEG ile doğrudan ilişkili sinyaller olup diğer iki sinyal kaynağı ise elektromiyografi (EMG) için kullanılan sinyallerdir.

Ölçümler sırasında dış etkenler veya çeşitli nedenlerle sinyallerde meydana gelen bozulma ve değişimleri EEG sinyallerinde canlandırabilmek için EMG'den elde edilen beşinci ve altıncı sinyaller her bir saniyelik ölçümlerin son 250 ve 500 milisaniyelik kısmında aktif hale getirilmektedir. D12, D12N, D19 ve D19N örnekleri için Abbass ve arkadaşları 25 farklı sinyal kaynağından faydalanmıştır. D12 ve D12N örnekleri için 25 sinyal kaynağından EEG ile ilişkili olan ilk 12 tanesi kullanılmış, D19 ve D19N örnekleri için EEG ile ilişkili olan ilk 19 sinyal kullanılmıştır. Toplamda 25 sinyal kaynağından geriye kalan 6 tanesi ise EMG için kullanılan sinyaller olup bu sinyallerden 20, 22 ve 24 numaralı sinyaller her bir saniyelik ölçümlerin son 250 milisaniyelik kısmında aktifleştirilmiş, 21, 23 ve 25 numaralı sinyaller ise her bir saniyelik ölçümlerin 500 milisaniyelik kısmında aktifleştirilmiştir. Bu şekilde sinyallerde oluşturulan gürültülerin birbirleriyle senkron olmamaları sağlanmıştır.

EEG sinyallerini temel alarak önerilen bu büyük veri probleminin temel amacı bir sinyal örneğini iki parçaya bölmektir. Oluşturulmak istenen parçalardan ilki kaynak sinyalin gürültülerden arındırılmış halini ifade edecek şekilde kaynak sinyale mümkün olduğunca yakın olmalı, ikinci parça ise sinyalin barındırdığı beyindeki elektriksel aktiviteler haricindeki gürültüyü ifade edilecek şekilde belirlenmelidir.  $N$  zaman serisi sayısını ve  $M$  serilerin uzunluğunu ifade etmek üzere  $S$  kaynak sinyali ifade eden  $N \times M$  boyutlu bir matris ile gösterilsin.  $A$  kare dönüşüm matrisi  $N \times N$  boyutlu bir matris ve  $X$  matrisi Eşitlik (1)'deki denkliği sağlayacak  $N \times M$  boyutlu bir matristir.

$$X = A \times S \quad (1)$$

Bu üç matris kullanılarak  $S$  matrisi  $S1$  ve  $S2$  olarak adlandırılan  $N \times M$  boyutlu iki matrise bölünmek istenilmektedir.  $S1$  matrisi gürültülerden temizlenmiş sinyali,  $S2$  matrisi ise sinyaldeki gürültüyü temsil etmektedir.  $S1$  matrisinin değerleri beynin ürettiği elektriksel aktivitelerden kaynaklı sinyalleri korumak için mümkün olduğunca  $S$  matrisine yakın değerler olarak seçilmelidir.  $S1$  ve  $S2$  sinyalleri birleştirildiğinde Eşitlik (2)'deki gibi  $S$  matrisine eşit olmalıdır.

$$S = S1 + S2 \quad (2)$$

Buna ek olarak  $S1$  ve  $S2$  matrisleri  $A$  matrisi ile dönüştürüldükten sonra toplanıldığında  $X$  matrisi Eşitlik (3)'te gösterildiği gibi elde edilmelidir.

$$X = (A \times S1) + (A \times S2) \quad (3)$$

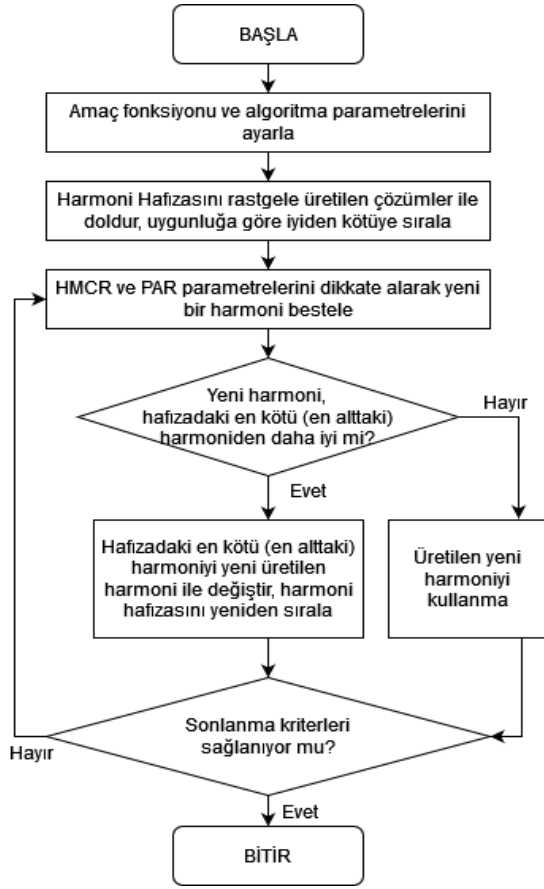
$S1$  ve  $S2$  matrislerini elde etmenin doğrudan bir yolu olmayıp istatistiksel yöntemler kullanılarak en uygun değerlerin bulunması gerekmektedir. Abbass ve arkadaşlarının çalışmalarında *Pearson* korelasyon katsayıları kullanılarak bu matrislerin değerlerinin elde edilmesi için gerekli hesaplamalar açıklanmıştır. Sonuç olarak elde edilen büyük veri optimizasyon probleminde  $S1$  matrisinin elemanları  $-8$  ile  $+8$  arasında değerler atanarak gürültüde temizlenmiş optimum sinyal elde edilmeye çalışılmaktadır.

### 3. Harmoni Arama (HS) Algoritması

Geem ve arkadaşları tarafından 2001 yılında tanıtılan HS algoritması doğal süreçlerden esinlenilerek geliştirilen meta-sezgisel algoritmalar arasındadır [19]. İnsan kulağına hoş gelen ezgilerin bestelenme süreçlerini oluşturduğu matematiksel model ile taklit ederek optimum sonuca "besteleme" yöntemiyle yaklaşmak amaçlanmaktadır. Harmoniler besteleyerek optimum sonuca ulaşmayı hedefleyen algoritmada, harmoni belleği adı verilen sabit boyutlu bir liste oluşturulmakta ve o ana kadar elde edilen en kaliteli harmoniler, kalite sırasına göre iyiler yukarıda olacak şekilde bu listede saklanılmaktadır. Başta rastgele olarak üretilen harmoniler kullanılarak doldurulan harmoni belleği algoritmanın ilerleyen adımlarında yeni harmoniler bestelenirken veya mevcut harmoniler güncellenirken başvurulan bir kaynak özelliği taşımaktadır. HS algoritmasının akış diyagramı Şekil-1'de gösterilmiştir.

Algoritma çalıştırılmaya başlanıldığında harmoni belleği rastgele olarak oluşturulduktan sonra yeni harmonilerin bestelenmesi sırasında *Harmoni Belleğinden Faydalanma Oranı (HMCR)* ve *Ton Ayarlama Oranı (PAR)* adı verilen parametrelerden faydalanılır. 0 ile 1 arasında sabit bir değer alabilen *HMCR* değeri, 0 ile 1 arasında rastgele olarak üretilen bir değerden daha büyük ise yeni harmoninin üretimi sırasında harmoni belleğindeki mevcut harmonilerden faydalanılır, değilse yeni üretilecek harmoni algoritmanın ilklenme aşamasında olduğu gibi rastgele olarak üretilir. Ardından *PAR* değeri kullanılarak yeni üretilen harmonide ton ayarlama yapılıp yapılmayacağına karar verilir. Rastgele olarak 0 ile 1 arasında üretilen bir sayı, *PAR* parametresinin değerinden daha küçükse yeni üretilen harmoni komşu harmoni değerlerinden faydalanılarak güncellenir, değilse mevcut harmoni olduğu gibi kullanılır. *PAR* değeri 0.20 ise %10 ihtimalle üst komşu harmoninin değerinden, %10 ihtimalle alt komşu harmoniden faydalanılarak mevcut harmoni güncellenecektir. *HMCR* ve *PAR* parametreleri kullanılarak üretilen bu yeni harmoni eğer ki harmoni belleğindeki en başarısız harmoniden daha iyi bir değer elde ediyorsa harmoni belleğindeki en başarısız harmoni bellekten çıkarılarak yerine yeni üretilen harmoni eklenir. Bu şekilde harmoni belleğinin boyutu sabit tutulmuş olmakla beraber bellekteki harmonilerin ortalama kalitesi de yükselmiş olur. Bellekte eleman değişikliği olması durumunda

harmoni belleğindeki harmoniler kalite değerlerine göre en iyi harmoni en yukarıda olacak şekilde sıralanır. Bu sayede harmoninin en kaliteli elemanının hep en üstteki harmoni olduğu, yeni üretilen bir harmoninin de en kötü harmoni ile karşılaştırılması için her zaman için en alttaki harmoni ile karşılaştırılması yeterli olur.



Şekil-1: HS Algoritması Akış Şeması

#### 4. Kaynak Bağlantılı HSA (slinkHSA)

HS algoritması, oluşturduğu harmoni belleği ile birlikte HMCR ve PAR parametrelerini kullanarak birçok optimizasyon problemi için optimum değerlere yakınsama başarımı olarak diğer meta-sezgisel algoritmalarla benzer başarımlar göstermekte, bazı problemlerde diğer birçok sezgisel algoritmadan daha iyi başarımlar sergileyebilmektedir. Abbass ve arkadaşları tarafından önerilen EEG sinyali tabanlı büyük veri optimizasyonu problemlerinde, problemin parametre sayısının yüksek olması, gürültülü sinyali temiz sinyal ve gürültülü sinyal olarak parçalama işleminin doğrusal olmayan bir problem ortaya koyması nedeniyle iyi bilinen bazı sürü zekâsı tabanlı ve evrimsel algoritmaların istenilen başarımları ortaya koyamamasına neden olmaktadır. Problemin bu yönleri göz önünde bulundurulduğunda HS algoritmasının da problemin mantığına göre güncellenmesinin daha iyi sonuçlar elde etmeyi kolaylaştıracağı öngörülebilir.

Abbass ve arkadaşları tarafından tanıtılan altı farklı alt problemden her birisi 256 Hz sıklığa sahiptir. 4 kanallı D4 ve D4N örnekleri için her bir çözüm için 1024 parametre, 12 kanallı D12 ve D12N örnekleri için her bir aday çözüm için 3072 parametre, 19 kanallı D19 ve D19N örnekleri için her bir aday çözüm için 4864 adet parametre üretilmesi

gerekmektedir. Problemlerin tamamında sırasıyla 1024, 3072 ve 4864 parametre için seçilebilecek rastgele değer  $-8.0$  ile  $+8.0$  değer aralığındadır. HS algoritmasında başlangıçtaki harmoni belleğinin ilklenmesi, harmonilerin güncellenmesi ve yeni harmoni üretimi sırasında problemin her bir parametresi için parametrenin alabileceği alt değer ile üst değer arasında rastgele olarak seçilen değerler atanmaktadır. El Majdouli ve arkadaşlarının geliştirdiği kaynak bağlantılı teknikte ise, EEG sinyalleri tabanlı büyük veri optimizasyon probleminde çözüm oluşturma aşamasında çözüm parametrelerine değer ataması yapılırken yine kaynak EEG sinyalindeki ilgili parametre değerinden faydalanılmaktadır. Bu yeni teknik kullanılarak parametre üretilmesi sırasında yeni bir parametre üretilirken normalde  $-8.0$  ile  $+8.0$  arasında olan değer aralığı kullanılmak yerine kaynak sinyali temsil eden  $S$  matrisinde ilgili parametreye denk gelen kaynak değer baz alınmaktadır. Yeni üretilen değer, çok daha küçük bir değer aralığı olan  $-1e - 05$  ile  $+1e - 05$  aralığında seçilen rastgele bir değer kaynak değere eklenmesiyle elde edilmektedir. Sinyalde meydana gelen gürültülerin sinyal üzerindeki olası etki miktarını içeren bu aralıkta yapılan seçimler sayesinde, algoritmanın harmoni belleğini oluşturma ve yeni harmoni üretimi sırasında kaynak sinyal baz alınarak yapılan küçük değer değişimleri elde edilmektedir. Rastgele olarak  $-8.0$  ile  $+8.0$  aralığında yapılan değer seçimi nedeniyle kaynak sinyal ile ilişkisi olmayan çözümler üzerinde hesaplamalar yapılması durumu engellenerek yanlış aralıklarda çözüm arayışının önüne geçmektedir. Daha önce farklı algoritmaların probleme özgü olarak kaynak sinyal değerlerini baz alarak optimize edildiğinde çok daha başarılı sonuçlar vermesinden yola çıkılarak HS algoritması kullanılarak oluşturulan ve Kaynak Bağlantılı HSA (slinkHSA) olarak adlandırılan bu yeni yöntemde harmoni belleğinin oluşturulma işlem adımları *Algoritma 1*'de gösterilmiştir.

#### Algoritma 1 slinkHSA Harmoni Belleği Oluşturma

- 1:  $HMS \leftarrow$  Harmoni bellek boyutu
- 2:  $N \leftarrow$   $S$  matrisindeki sütun sayısı
- 3:  $M \leftarrow$   $S$  matrisindeki satır sayısı
- 4: **for**  $i \leftarrow 1 \dots HMS$  **do**
- 5:     **for**  $j \leftarrow 1 \dots N$  **do**
- 6:         **for**  $k \leftarrow 1 \dots M$  **do**
- 7:              $u \leftarrow$  rastgele( $-1e - 05 \dots +1e - 05$ )
- 8:              $p \leftarrow j \times M + k$
- 9:              $X_{i,p} \leftarrow S_{j,p} + u$
- 10:         **end for**
- 11:     **end for**
- 12: **end for**

#### 5. Deneysel Çıktılar

HS algoritması harmoni üretme stratejisi değiştirilmemiş hali kullanılarak ve slinkHSA kullanılarak D4, D4N, D12, D12N örnekleri üzerinde optimizasyon çalışması yapılmıştır. Çeşitli HMCR ve PAR değerleri ile yapılan denemeler sonucunda

algoritmanın geleneksel hali ile ve slinkHSA ile en optimum değerleri ürettiği değer aralıkları saptanmaya çalışılmıştır. Bu çalışmalar sonucunda HMCR parametresinin 0,90 ve 0,95 değerlerini, PAR parametresinin de yine 0,90 ve 0,95 değerlerini aldığı anda probleme en uygun çözümlerin elde edildiği görülmüştür. D4, D4N, D12, D12N örnekleri için HMCR parametresinin 0,90 ve 0,95 değerleri ile, PAR parametresinin de 0,90 ve 0,95 değerleri ile koşurulmasıyla meydana gelen her bir kanal için 4, toplamda 24 farklı kombinasyonda hem

standart HS algoritması ile hem de bu çalışmada önerilen slinkHSA ile harmoni bellek boyutu 50 olarak belirlenen koşmalar yapılmıştır. Değerlendirme sayısı 10.000 olarak belirlenerek yapılan koşmalar 30'ar defa tekrarlanmış, elde edilen sonuçlar D4 ve D4N için Çizelge 1'de, D12 ve D12N için Çizelge-2'de gösterilmiştir. Çizelgelerde her bir kanal sayısı ve gürültü değeri için elde edilen en başarılı ortalama ve en iyi sonuçlar koyu yazı tipiyle belirtilmiştir.

**Çizelge-1: D4 ve D4N Örnekleri için HS Algoritması ve slinkHSA ile Elde Edilen Sonuçlar**

Parametreler	D4 - HSA		D4 – slinkHSA		D4N - HSA		D4N – slinkHSA	
	Ortalama	En İyi	Ortalama	En İyi	Ortalama	En İyi	Ortalama	En İyi
HMCR: 0.90 PAR: 0.90	1.2561E+01	1.2043E+01	<b>8.0326E-02</b>	<b>7.9366E-02</b>	1.2563E+01	1.2442E+01	<b>7.8013E-02</b>	<b>7.7657E-02</b>
HMCR: 0.90 PAR: 0.95	1.2573E+01	1.2213E+01	8.1061E-02	8.0278E-02	1.2437E+01	1.2306E+01	7.8519E-02	7.8230E-02
HMCR: 0.95 PAR: 0.90	8.6611E+00	8.4072E+00	8.4404E-02	8.3280E-02	8.7110E+00	8.3265E+00	8.2299E-02	8.1562E-02
HMCR: 0.95 PAR: 0.95	8.6819E+00	8.3662E+00	8.5361E-02	8.4540E-02	8.7059E+00	8.2811E+00	8.2995E-02	8.1355E-02

**Çizelge-2: D12 ve D12N Örnekleri için HS Algoritması ve slinkHSA ile Elde Edilen Sonuçlar**

Parametreler	D12 - HSA		D12 – slinkHSA		D12N - HSA		D12N – slinkHSA	
	Ortalama	En İyi	Ortalama	En İyi	Ortalama	En İyi	Ortalama	En İyi
HMCR: 0.90 PAR: 0.90	1.6477E+01	1.6111E+01	<b>1.7381E-02</b>	<b>1.7027E-02</b>	1.6464E+01	1.6053E+01	<b>1.7227E-02</b>	<b>1.6783E-02</b>
HMCR: 0.90 PAR: 0.95	1.6428E+01	1.5902E+01	1.8344E-02	1.7975E-02	1.6505E+01	1.6236E+01	1.8263E-02	1.7981E-02
HMCR: 0.95 PAR: 0.90	1.3430E+01	1.3099E+01	2.8163E-02	2.7566E-02	1.3467E+01	1.3057E+01	2.8104E-02	2.7444E-02
HMCR: 0.95 PAR: 0.95	1.3444E+01	1.3145E+01	2.9547E-02	2.8620E-02	1.3443E+01	1.3000E+01	2.9469E-02	2.8490E-02

Çizelge-1'de görülebileceği üzere slinkHS algoritması 4 kanallı gürültülü ve gürültüsüz sinyaller üzerinde, HMCR parametresinin 0,90 ve PAR parametresinin 0,90 olarak ayarlandığı durumda hem HS algoritmasının hem de slinkHS algoritmasının diğer parametrelerle aldığı sonuçlardan daha iyi sonuçlar almıştır. Çizelge-2'de de aynı şekilde 12 kanallı gürültülü ve gürültüsüz sinyaller üzerinde HMCR parametresinin 0,90 ve PAR parametresinin 0,90 olarak belirlendiği durumda en başarılı sonuçlar elde etmiştir. Çizelge-3'te de 19 kanallı gürültülü ve gürültüsüz sinyaller üzerinde HMCR parametresinin 0,90 ve PAR parametresinin 0,90 olarak belirlendiği durumda en başarılı sonuçlar elde edilmiştir.

Standart HS ve slinkHS algoritması ile alınan sonuçlar diğer bilinen optimizasyon algoritmalarının elde ettiği yöntemlerle karşılaştırılmıştır. EEG sinyali tabanlı büyük veri optimizasyonu problemlerinde D4, D4N, D12, D12N örnekleri için 30 koşma sonucu elde edilen ortalama değerler Genetik Algoritma (GA) [20], Diferansiyel Gelişim (DE) Algoritması [21], Yapay Arı Koloni (ABC) Algoritması [22], Parçacık Sürü Optimizasyonu (PSO) [23] ve İmmun Plazma Algoritması (IPA) [24] ile alınan sonuçlar Çizelge-3'te HS algoritması ve slinkHS algoritması ile karşılaştırılmıştır. Çizelgede her bir kanal ve gürültü örneği için elde edilen değerlerden diğer algoritmalar içerisinde en iyi değere sahip olanlar koyu olarak yazılmıştır. slinkHS algoritması D4, D4N, D12 ve D12N kaynak sinyallerinin optimizasyonunda en başarılı sonuçları elde etmiştir.

**Çizelge-3: slinkHSA ve HS Algoritmalarının Diğer Meta-Sezgisel Algoritmalarla Sonuç Karşılaştırması**

Algoritma	D4	D4N	D12	D12N
slinkHSA	<b>8.0326E-02</b>	<b>7.8013E-02</b>	<b>1.7381E-02</b>	<b>1.7227E-02</b>
HSA	8.6611E+00	8.7059E+00	1.3444E+01	1.3443E+01
GA	2.1300E+00	2.1491E+00	2.8016E+00	2.7769E+00
PSO	7.8788E+00	7.8962E+00	1.0634E+01	1.0613E+01
DE	1.4653E+01	1.6886E+01	2.2396E+01	2.2439E+01
ABC	2.0042E+01	2.0101E+01	2.1958E+01	2.1954E+01
IPA	1.6599E+00	1.6989E+00	1.8370E+00	1.8359E+00

Daha önce bahsedildiği üzere EEG sinyallerinin gürültü içeren ölçüm sonuçları S matrisinde depolanmaktadır. S matrisi gürültülü sonuçları içeriyor olsa dahi aynı matrisin gürültüden filtrelenmiş karşılığı yine S matrisine oldukça yakın olacaktır. Başlangıç çözümlerinin S matrisinin elemanları dikkate alınarak belirlenmesi, bir diğer ifade ile S matrisinin elemanlarına oldukça yakın değerler içeren çözümlerin başlangıç popülasyonunda bulunması, HS algoritmasının kaliteli çözümler ile optimizasyon işlemine başlamasına olanak sağlamıştır. Ayrıca, HS algoritmasının elde ettiği aday çözümleri popülasyondaki daha kötü çözümler yerine tercih ederek muhafaza etmesi hem popülasyonun genel kalitesini artırmış hem de sonraki adımlarda yeni eklenen çözümlerden faydalanılmasına katkı sunmuştur.

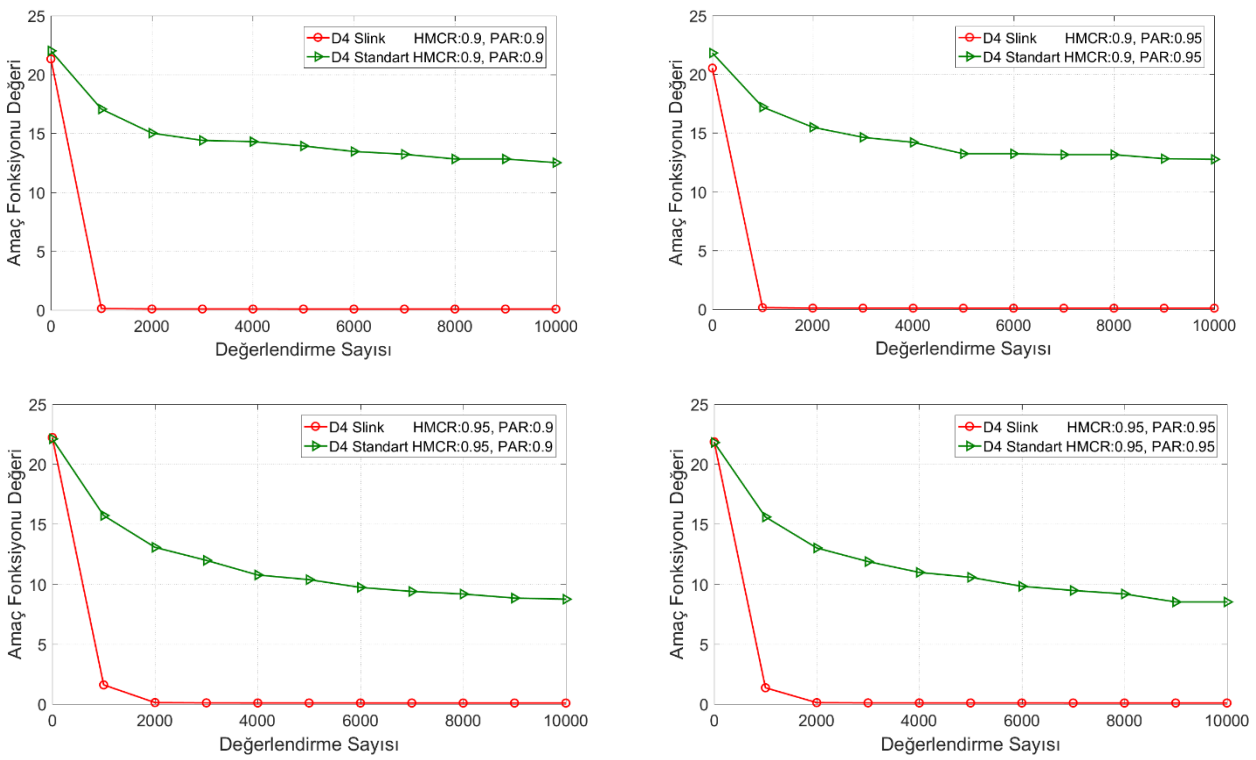
EEG sinyallerinde gürültüyü en aza indirmeyi gerektiren büyük veri optimizasyon probleminde D4 ve D4N problem örnekleri 1024, D12 ve D12N problem örnekleri 3072, D19 ve D19N problem örnekleri ise 4864 parametre içermektedir. Problem örneklerinin içerdiği parametre sayıları dikkate alındığında



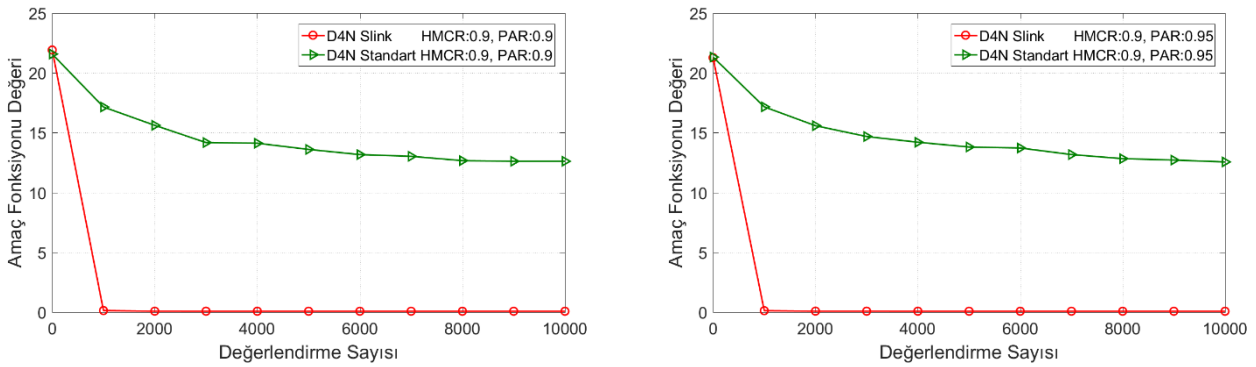
hem başlangıç çözümlerinin hem de aday çözüm üretmek üzere belirlenecek stratejinin problemin bazı özelliklerine bağlı yapılmasının büyük veri optimizasyonu için meta-segzisel algoritmaların performansını doğrudan artıracığı ön görülmektedir. En iyi değerleri belirlenecek az sayıda parametreye sahip optimizasyon problemleri için başlangıç çözümlerinin rastgele üretilmesi, araştırma uzayının taraması, çözüm çeşitliliğinin sağlanması ve kaliteli başlangıç çözümlerinin bulunması açısından katkı sunabilir. Ancak, bahsedilen büyük veri optimizasyon problemindeki örnekler için sayıları 1024 ile 4864 arasında değişen parametrelerin başlangıç değerlerinin -8 ile +8 arasında rastgele belirlenmesi, parametre sayılarının fazla olması sebebi ile kaliteli başlangıç çözümlerinin elde edilmesi ihtimalini azaltır. Eğer, slinkHS algoritmasında olduğu gibi başlangıç çözümlerini temsil eden

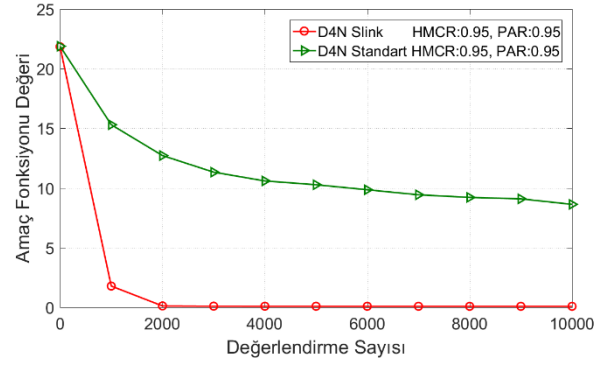
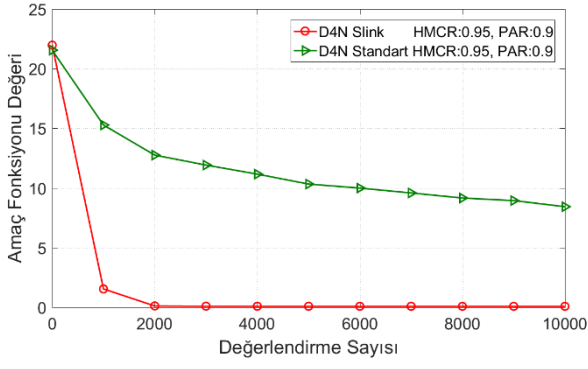
armonilerin üretilmesi problemin özellikleri dikkate alınarak gerçekleştirilir ve algoritmaya özgü kontrol parametrelerine uygun değerler atanır ise oldukça başarılı çözümlere ulaşılabildiği görülür.

Optimizasyon algoritmalarının performansını belirlemeye yönelik en önemli göstergelerden bir tanesi de algoritmanın yakınsama performansdır. Geliştirilen yeni yöntemin algoritmanın değiştirilmemiş hali ile optimum sonuca yakınsama hızlarını karşılaştırabilmek adına algoritmaların aşamaları bazlı olarak elde ettiği sonuçlar da kayıt altına alınmıştır. HS algoritması ve slinkHSA ile elde edilen yakınsama grafiklerinin karşılaştırılması Şekil 2-7'de gösterilmiştir.

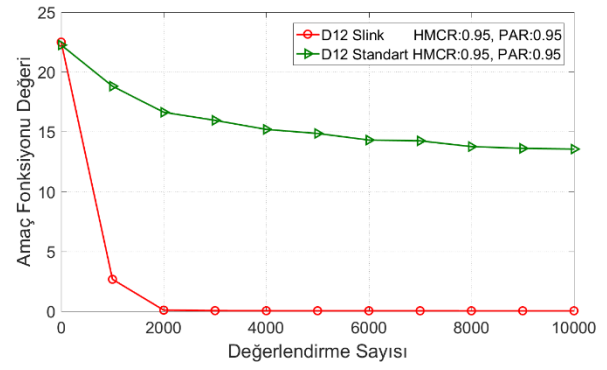
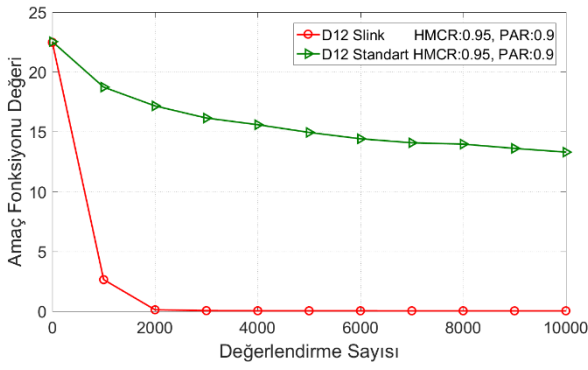
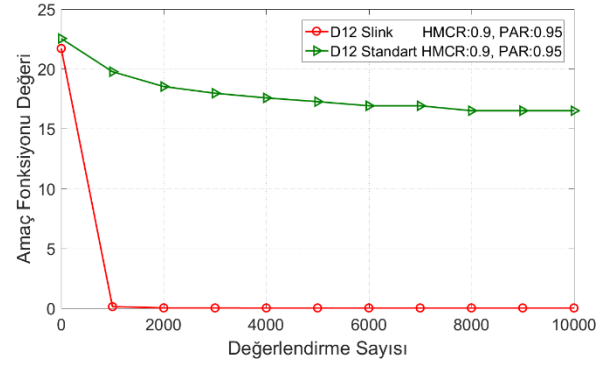
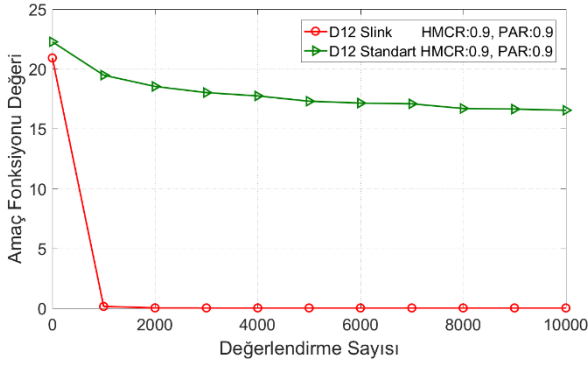


Şekil-2: D4 için HS Algoritması (Standart) ile slinkHS Algoritmasının 10.000 İterasyondaki Yakınsama Grafikleri

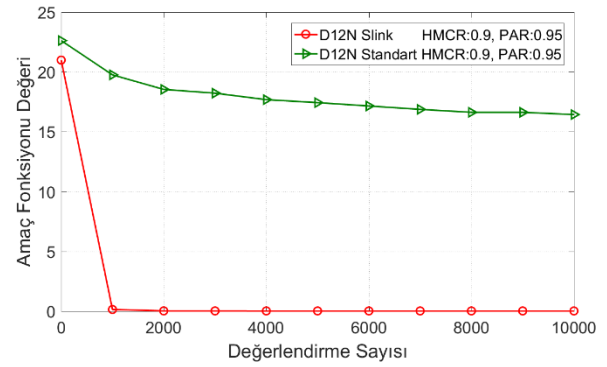
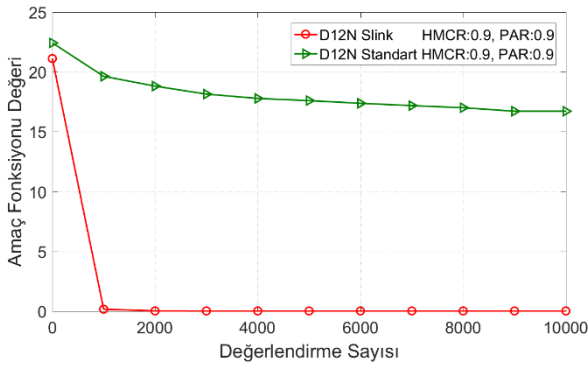


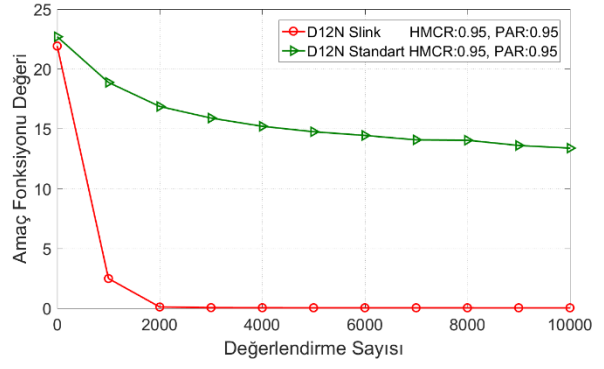
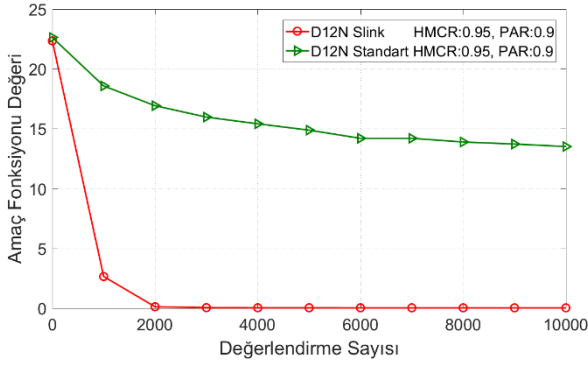


Şekil-3: D4N için HS Algoritması (Standart) ile slinkHS Algoritmasının 10.000 İterasyondaki Yakınsama Grafikleri



Şekil-4: D12 için HS Algoritması (Standart) ile slinkHS Algoritmasının 10.000 İterasyondaki Yakınsama Grafikleri

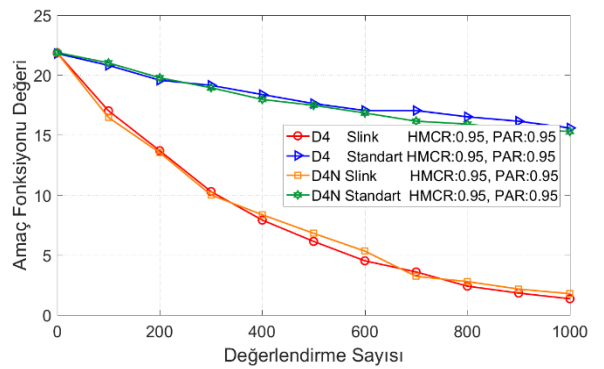
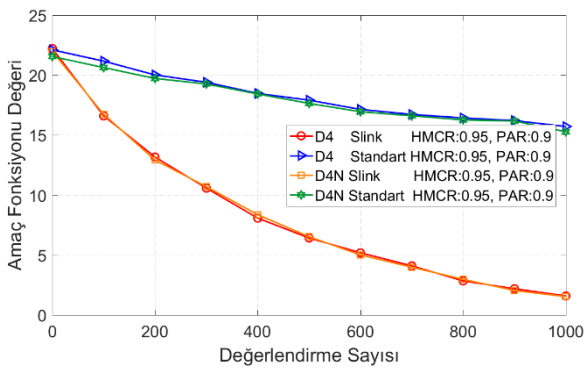
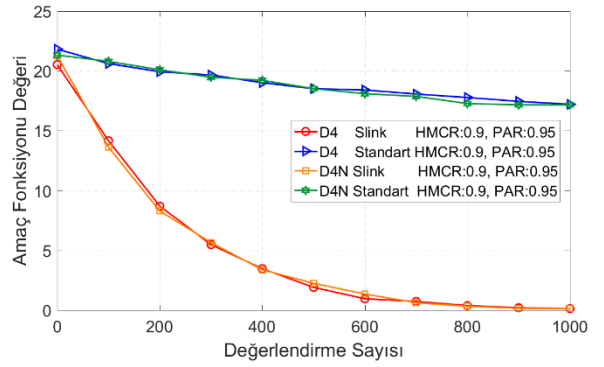
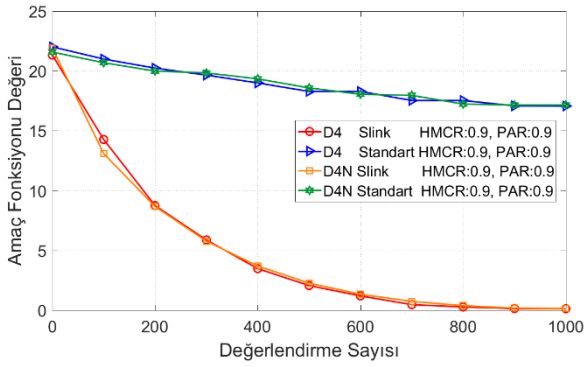




Şekil-5: D12N için HS Algoritması (Standart) ile slinkHS Algoritmasının 10.000 İterasyondaki Yakınsama Grafikleri

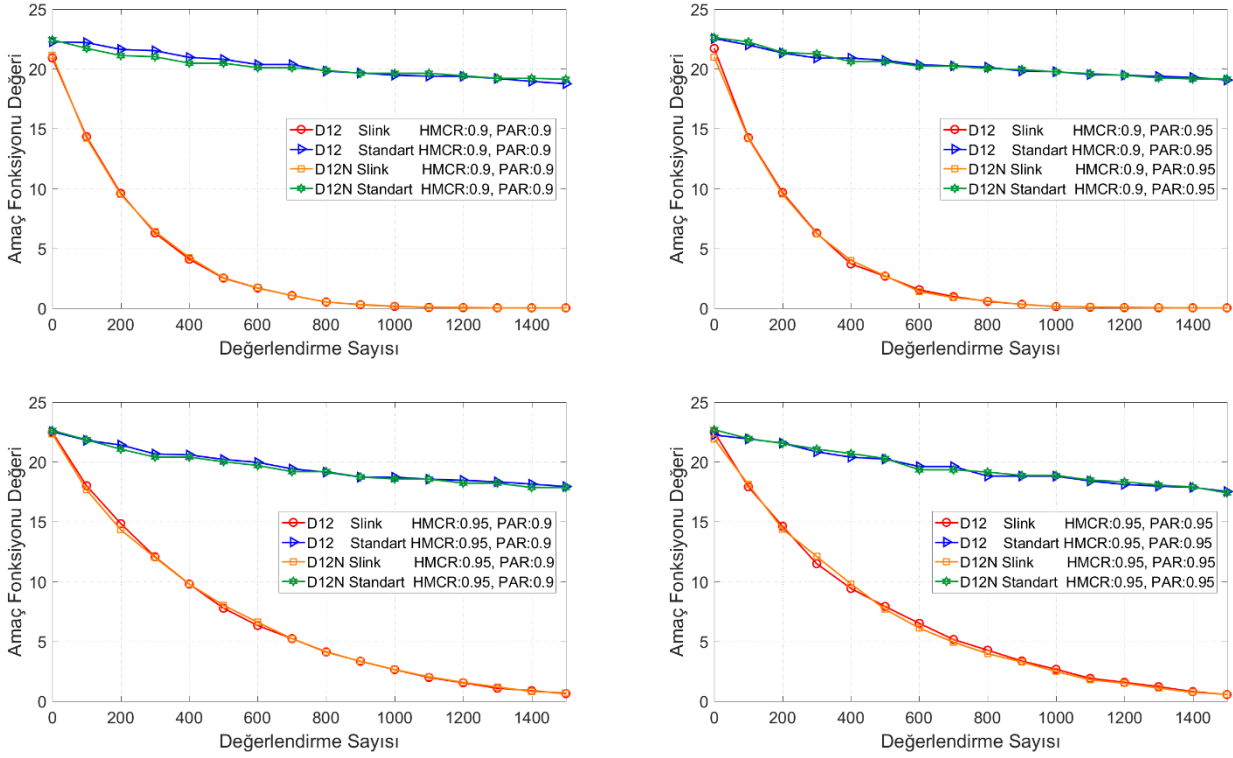
Şekil 2-7'de yapılan çalıştırmalarda durdurma ölçüsü olarak belirlenen 10.000 değerlendirme sayısı üzerinden görselleştirme yapılmıştır. Ancak yakınsama hızındaki başta meydana gelen farklılık nedeniyle ilk sonuçlar yeterince detaylı görülememektedir. Yakınsama eğrilerinin iki algoritma için de belirli bir aşamadan sonra yatay eksende düzleştiği görülmektedir. Grafikte büyük yer kaplayan bu alanlar kırılarak daha detaylı bir görünüm elde edilmiştir.

Şekil-6'de D4 ve D4N için iki algoritmanın ilk 1000 adımdaki yakınsama grafikleri, Şekil-7'da D12 ve D12N için iki algoritmanın ilk 1500 adımdaki yakınsama grafikleri verilmiştir. Grafiklerden rahatça görülebildiği üzere slinkHS algoritması HS algoritmasından çok daha hızlı şekilde optimum sonuca yakınsamakta, gürültü temizleme işlemi daha başarılı şekilde yürütülebilmektedir.



Şekil-6: D4 ve D4N için HS Algoritması (Standart) ile slinkHS Algoritmasının 1000 İterasyondaki Yakınsama Grafikleri





Şekil-7: D12 ve D12N için HS Algoritması (Standart) ile slinkHS Algoritmasının 1500 İterasyondaki Yakınsama Grafikleri

## 6. Sonuç

Bu çalışmada EEG sinyalleri temelli büyük veri optimizasyon problemlerinde HS algoritmasının parametrelere değer ataması sırasında kaynak sinyalden faydalanarak gürültüleri daha başarılı temizlemek için geliştirilen slinkHSA algoritması tanıtılmış ve diğer meta-sezgisel algoritmalar ile ve standart HS algoritması ile elde edilen sonuçlar kıyaslanmıştır. Elde edilen sonuçlar göstermiştir ki HS algoritmasını geleneksel haliyle kullanmak yerine kaynak sinyalden faydalanarak çözüm üretmek algoritmanın yakınsama performansını D4 sinyali için %99,18, D4N sinyali için %99,21, D12 sinyali için %99,83, D12N sinyali için %99,83 oranında daha iyi bir hale getirmektedir. slinkHSA'nın optimum sonuca yakınsama hızı

## Kaynakça

- [1] Kambatla, K., Kollias G., Kumar V., Grama, A., *Trends in Big Data Analytics*. Journal of Parallel and Distributed Computing, 74(7):2561–2573, 2014.
- [2] Gudivada, V. N., Baeza-Yates, R., Raghavan, V. V., *Big Data: Promises and Problems*. Computer, 48(3):20–23, 2015.
- [3] Tsai, C. W., Lai, C. F., Chao, H. C., Vasilakos, A. V., *Big data Analytics: A Survey*. Journal of Big Data, 2(1):21, 2015.
- [4] Özköse, H., Arı, E. S., Gencer, C., *Yesterday, Today and Tomorrow of Big Data*, Procedia-Social and Behavioral Sciences, 195, 1042-1050, 2015
- [5] Abbass, H. A., *Calibrating Independent Component Analysis with Laplacian Reference for Real-Time EEG Artifact Removal*. International Conference on Neural Information Processing, pages 68–75, 2014.
- [6] Goh, S. K., Abbass, H. A., Tan, K. C., Al-Mamun, A., *Artifact Removal From EEG Using a Multi-Objective Independent Component Analysis Model*. International Conference on Neural Information Processing, pages 570–577, 2014.
- [7] Goh, S. K., Tan, K. C., Al-Mamun, A., Abbass, H. A., *Evolutionary Big Optimization (bigopt) of Signals*. IEEE Congress on Evolutionary Computation (CEC), pp: 3332–3339. IEEE, 2015.
- [8] Zhang, Y., Zhou, M., Jiang, Z., Liu, J., *A Multi-Agent Genetic Algorithm for Big Optimization Problems*. IEEE Congress on Evolutionary Computation (CEC), pages 703–707. IEEE, 2015.
- [9] Zhang, Y., Liu, J., Zhou, M., Jiang, Z., *A Multi-Objective Memetic Algorithm Based on Decomposition for Big Optimization Problems*. Memetic Computing, 8(1):45–61, 2016.

- [10] Elsayed, S., Sarker, R., *An Adaptive Configuration of Differential Evolution Algorithms for Big Data*. IEEE Congress on Evolutionary Computation (CEC). IEEE, pp: 695–702, 2015.
- [11] Elsayed, S., Sarker, R., *Differential Evolution Framework for Big Data Optimization*. Memetic Computing, 8(1):17–33, 2016.
- [12] El Majdoui, M. A., Bougrine, S., Rbouch, I., El Imrani, A. A., *A Fireworks Algorithm for Single Objective Big Optimization of Signals*. IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), pp: 1–7. IEEE, 2016.
- [13] Loukdache A., El Majdoui, M. A., Bougrine, S., El Imrani, A. A., *A Clonal Selection Algorithm For the Electro Encephalography Signals Reconstruction*. International Conference on Electrical and Information Technologies (ICEIT), pp: 1–6. IEEE, 2017.
- [14] Meselhi, M. A., Elsayed, S. M., Essam, D. L., Sarker, R. A. *Fast Differential Evolution for Big Optimization*. 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), pp: 1–6. IEEE, 2017.
- [15] Wang, H., Wang, W., Cui, L., Sun, H., Zhao, J., Wang, Y., Xue, Y., *A Hybrid Multiobjective Firefly Algorithm for Big Data Optimization*. Applied Soft Computing, 69:806–815, 2018.
- [16] Yi, J. H., Deb, S., Dong, J., Alavi, A. H., Wang, G. G., *An Improved NSGA-III Algorithm with Adaptive Mutation Operator for Big Data Optimization Problems*. Future Generation Computer Systems, 88:571–585, 2018.
- [17] Aslan, S., *An Artificial Bee Colony-Guided Approach for Electro-Encephalography Signal Decomposition-Based Big Data Optimization*. International Journal of Information Technology & Decision Making, 19(02), 561-600, 2020.
- [18] Jiang, X., Bian, G., Tian, Z., *Removal of Artifacts from EEG Signals: A Review*. Sensors, 19(5), 987, 2019
- [19] Geem, Z. W., Kim, J. H., Loganathan, G. V., *A New Heuristic Optimization Algorithm: Harmony Search*. Simulation, 76(2), 60-68, 2001.
- [20] Srinivas, M., Patnaik, L. M., *Genetic Algorithms: A Survey*. Computer, vol. 27, no. 6, pp. 17–26, Jun. 1994.
- [21] Price, K. V., *Differential Evolution*, Handbook of Optimization. Berlin, Germany: Springer, 2013, pp. 187–214.
- [22] Karaboga, D., Basturk, B., *A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm*. J. Global Optim., vol. 39, no. 3, pp. 459–471, Oct. 2007.
- [23] Shi, Y., *Particle Swarm Optimization: Developments, Applications and Resources*, Proc. Congr. Evol. Comput., vol. 1, May 2001, pp. 81–86.
- [24] Aslan, S., Demirci, S., *Immune Plasma Algorithm: A Novel Meta-Heuristic for Optimization Problems*. IEEE Access, vol. 8, pp. 220227-220245, 2020
- [25] İleri, S. C., Aslan, S., Demirci, S., *A Novel Harmony Search Based Method for Noise Minimization on EEG Signals*. 2021 6th International Conference on Computer Science and Engineering (UBMK), 2021, pp. 747-750