

(Araştırma Makalesi)**Koroner Arter Hastalığı Sınıflandırılmasında Destek Vektör Makinelerinin Gri Kurt Optimizasyonuna Dayalı Özellik Seçim Yöntemi ile Geliştirilmesi****Büşra ER^{*1}, Uğur FİDAN²**¹ Afyon Kocatepe Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği Bölümü, 03030, Afyon,ORCID No : <http://orcid.org/0000-0001-9682-8651>² Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, Biyomedikal Mühendisliği Bölümü, 03030, Afyon,ORCID No : <http://orcid.org/0000-0003-0356-017X>**Anahtar Kelimeler:**

Gri Kurt Optimizasyonu,
Öznitelik Çıkarımı,
Destek Vektör Makinesi,
Koroner Arter Hastalığı,
Çekirdek Matris

Özet: Makine öğrenmesi yöntemleri, büyük veri kümelerinin analiz edilmesine olanak sağlayarak koroner arter rahatsızlığı ve/veya buna benzer hastalık ve durumların tespit edilmesinde kullanılan etkili bir araçtır. Büyük veri kümelerinde işlem hızını ve sınıflandırma başarımını olumsuz etkileyen veriler bulunabilmektedir. Özellik seçim tekniklerinin uygulanması gereksiz verilerin ortadan kaldırılmasına olanak sağlamaktadır. Bu çalışmada, koroner arter hastalığını teşhis etmek amacıyla destek vektör makineleri ve gri kurt optimizasyon algoritmasının birlikte kullanıldığı bir yöntem önerilmektedir. Önerilen yöntem, öznitelik seçimi ve sınıflandırma olmak üzere iki ana aşamadan oluşmaktadır. Yöntemin performans doğrulaması için Cleveland kalp hastalığı veri seti kullanılmıştır. Yöntemin ilk aşamasında kullanılan veri setinden en uygun özellik alt kümesini belirlemek üzere gri kurt optimizasyon algoritması (GWO) kullanılmıştır. İkinci aşamada, GWO'nun uygunluk fonksiyonu, destek vektör makinesi (SVM) sınıflandırıcısı kullanılarak değerlendirilmiştir. İlk aşamadan elde edilen sonuçlara bakıldığında kullanılan veri setinde bulunan 13 parametre arasından belirlenen 7 parametre seçilmiş ve sınıflandırma işlemi bu parametreler üzerinden gerçekleştirilmiştir. Çalışmada belirlenen uygunluk fonksiyonları SVM'de kullanılan çekirdek matrislerin farklı varyasyonları ile değerlendirilmiştir. Bu aşamada en yüksek doğruluk elde edilen çekirdek matris belirlenmiştir. Deneysel sonuçlar, önerilen GWO-SVM'nin lineer çekirdek matris kullanılarak %95.91 doğrulukta, %95.64 duyarlılıkta ve %91.66 başarı ile mevcut çalışmalara kıyasla daha yüksek başarımlar sağlandığını göstermiştir.

(Research Article)**Development of Support Vector Machines in The Classification of Coronary Artery Disease with Gray Wolf Optimization Based Feature Selection Method****Keywords:**

Grey Wolf Optimization,
Feature Selection,
Support Vector Machine,
Coronary Artery Disease,
Kernel Matrix

Abstract: Machine learning methods are an effective tool used to detect coronary artery disease and/or similar diseases and conditions by allowing the analysis of large data sets. Large data sets may contain data that negatively affects processing speed and classification performance. Applying feature selection techniques allows the elimination of unnecessary data. In this study, a method that uses support vector machines and gray wolf optimization algorithm together to diagnose coronary artery disease is proposed. The proposed method consists of two main stages: feature selection and classification. The Cleveland heart disease dataset was used for performance validation of the method. The gray wolf optimization algorithm (GWO) was used to determine the most appropriate feature subset from the dataset used in the first stage of the method. In the second stage, the fitness function of GWO was evaluated using the support vector machine (SVM) classifier. Considering the results obtained from the first stage, 7 parameters were selected among the 13 parameters in the data set used and the classification process was carried out on these parameters. The fitness functions determined in the study were evaluated

with different variations of the kernel matrices used in SVM. At this stage, the core matrix with the highest accuracy was determined. Experimental results showed that the proposed GWO-SVM achieved higher performance compared to existing studies, with 95.91% accuracy, 95.64% sensitivity and 91.66% success using the linear kernel matrix.

1. GİRİŞ

Son yıllarda, tıbbi teşhis ve karar verme süreçlerinde makine öğrenimi ve veri madenciliği yöntemleri sıklıkla kullanılmaya başlanmıştır [1]. Tıp endüstrisinde [2], kalp hastalıkları [3-5] beyin tümörü [6-8] cilt kanseri [9,10], meme kanseri [11] vb. gibi çeşitli hastalık tahmin modellerinde veri madenciliği teknikleri kullanılmaktadır. Yüksek mortalite oranı ve hızı nedeni ile kalp hastalıkları en sık çalışılan konular içerisinde yer almaktadır. Kalp hastalıkları içerisinde en fazla prevalansa sahip olan rahatsızlıklar kardiyovasküler hastalıkları (KDH), koroner kalp hastalığı (KKH), koroner arter hastalığı (KAH), romatizmal kalp hastalığı, serebrovasküler hastalıklar (inme), hipertansiyon ve kalp yetmezliğidir [12]. Belirtilen kalp rahatsızlıkları arasındaki KAH, kalbe giden gerekli kan akışını bloke eden atardamarların iç duvarında yağ ve kolesterol birikmesi nedeniyle yaygın olarak ortaya çıkan bir kalp rahatsızlığıdır [13].

Konu ile ilgili olarak literatürde yapılan çalışmalar incelendiğinde; Abdar ve arkadaşları koroner arter hastalığı teşhisi için 10 farklı makine öğrenim tekniğini denemiş ve bunlardan en iyi sonuç veren 3 destek vektör makine algoritmasını sınıflandırma için seçmiştir. Sınıflandırmada optimizasyonu sağlamak için parçacık sürü optimizasyonu (Particle swarm optimization-PSO) ve genetik algoritmadan (Genetic algorithm-GA) oluşan sarmal bir yapı kullanılmıştır. Çalışmada %93.08 doğruluk ve %91.51 F1 skoru elde edilmiştir [14]. Al-Tashi ve arkadaşları tarafından 2019 yılında yapılan diğer bir çalışmada Cleveland kalp hastalığı veri seti üzerinde gri kurt optimizasyon algoritması (Grey wolf optimization-GWO) ile öznelik çıkarımı yapıp destek vektör makinesi (Support vector machine-SVM) ile sınıflandırma yapılmıştır. Çalışmada sonucunda aynı veri setini kullanan diğer çalışmalara göre daha yüksek doğrulukta sınıflandırma başarımı elde edilmiştir [15]. Tama ve arkadaşları tarafından 2020 yılında yapılan çalışmada koroner kalp hastalığı tahmini için farklı veri setleri (Z-Alizadeh Sani, Statlog, Cleveland ve Hungarian) kullanılmıştır. Parçacık sürü optimizasyonu ile öznelik çıkarımlarının gerçekleştirilmesinin ardından sınıflandırma için 3 farklı makine öğrenmesinden oluşan mimari oluşturulmuştur. Bu çalışma o tarihe kadar yapılan diğer çalışmalara kıyasla daha yüksek doğrulukta sonuçlar üretmiştir [16]. Moturi ve arkadaşları tarafından yapılan çalışmada kalp hastalığı ve meme kanseri teşhisinde kullanılmak üzere; GWO ve yusufoçuk algoritması birleştirilmiş ve diğer popüler hibrid algoritmalarla elde edilen sonuçlar karşılaştırılmıştır [17]. Kalp yetmezliği erken teşhis edilebilmesi için 2021 yılında gerçekleştirilen bir başka çalışmada çok katmanlı perceptron mimarisi geliştirilmiştir. Girdi sayısını

azaltmak için PSO-GWO kullanan sarmal yapılı öznelik seçimi önerilmiştir [18]. Yapılan bir diğer çalışmada kalp hastalığı tahmini için diferansiyel evrimli gri kurt optimizasyon algoritması ve ateşböceği algoritması kullanılmıştır. Önerilen sınıflandırma optimum ağırlıkları elde etmek için sinir ağını eğitmekte ve çok sayıda hiperparametreyi verimli bir şekilde ayarlamaktadır. Veri setleri üzerinde denenen sonuçlarda, kalp hastalığı tahmini için uygun hale getiren yöntemin etkinliğini ortaya koyulmuştur [19]. Krishna ve arkadaşları tarafından yapılan çalışmada ise özellik seçimi için hibrit gri kurt optimizasyon algoritması kullanılmıştır. PSO ve GWO'yu birleştiren algoritmalarda öznelik çıkarımı yapılırken k-en yakın komşu sınıflandırıcısı ve Öklid ayırma matrisi kullanılmıştır. Bu seçeneklerde en iyi sonuç veren algoritma GWOPSO-KNN olarak bulunmuştur. [20]. Arabasadi ve arkadaşlarının yaptığı çalışmada, koroner arter hastalığının teşhisi için Cleveland kalp hastalığı veri seti kullanılarak, hibrit bir yöntem önerilmiştir. Bu yöntemde yapay sinir ağının başlangıç ağırlıkları genetik algoritma ile belirlenmiştir. Daha sonra eğitim verileri kullanılarak yapay sinir ağı öğrenimi gerçekleştirilmiştir [21]. Paul ve arkadaşlarının yaptığı çalışmada kalp hastalığının risk düzeyini tahmin etmek için değiştirilmiş dinamik çok kümelili parçacık sürüsü optimizasyonuna (MDMS-PSO) ve genetik algoritmaya (GA) dayalı bir otomatik bulanık teşhis sistemi önerilmiştir [22]. Subanya ve arkadaşlarının yaptığı çalışmada, hastalık tanımlamasında optimal özellikleri elde etmek için yapay arı kolonisi (Artificial bee colony algorithm-ABC) algoritmasına ve SVM'ye dayalı bir özellik seçim yöntemi önerilmiştir. SVM sınıflandırması, ABC'nin uygunluğunu değerlendirmek için kullanılmıştır [23].

Literatürde yapılan çalışmalar incelendiğinde Cleveland veri seti ve farklı veri setleri makine öğrenmesi çalışmalarında kullanılmaktadır. Farklı optimizasyon teknikleri ve farklı sınıflandırıcılarla yapılan çalışmalarda tahmin başarısının ve sistem çalışma hızının arttığı görülmektedir. Bu çalışmada; Cleveland kalp hastalığı veri setinde gri kurt optimizasyon algoritmasıyla özellik seçimi gerçekleştirilecektir. Özellik seçimi öncesinde ve sonrasında SVM ile sınıflandırma yapılacaktır. SVM'de kullanılan çekirdek fonksiyonların da sınıflandırma başarımına olan etkisi incelenecektir. SVM'de kullanılan çekirdek fonksiyonlar değiştirilerek deneyler yeniden gerçekleştirilecek ve en yüksek doğrulukta sonuçlar elde edilmeye çalışılacaktır. Böylece koroner arter hastalığı teşhisinde daha az sayıda parametreler ile daha yüksek tahmin gücüne sahip ve daha hızlı çalışacak algoritmanın geliştirilmesi hedeflenmektedir.

2. MATERYAL VE METOT

2.1. Gri Kurt Optimizasyon Algoritması

Tablo 1’de GWO’nun kaba koduna yer verilmiştir. Gri kurt optimizasyon algoritması (Tablo 1), kurt sürüsünün avlanma davranışından ilham alınarak geliştirilmiş meta-sezgisel optimizasyon algoritmasıdır. Bu algoritma, karmaşık optimizasyon problemlerinin çözümünde etkilidir ve literatürde birçok çalışmada başarılı sonuçlar elde edilmiştir [15,17-19]. Algoritma, kurt sürüsünün davranışlarını taklit ederek çalışmaktadır. Her bir kurt, bir çözüm adayını temsil eder. Başlangıçta, kurtlar rasgele dağılır. Her bir kurt, mevcut çözümünün kalitesini değerlendiren uygunluk fonksiyonuna göre bir uygunluk değeri alır. Daha sonra, kurtlar arasında bilgi paylaşımı gerçekleşir. Algoritma keşif, kaçınma ve takip olmak üzere üç temel davranışı modellemektedir. Keşif davranışı, kurtların çevreyi araştırmak için hareket etmelerini sağlar. Kaçınma davranışı, düşük kaliteli çözümlerden kaçınmak amacıyla yapılan bir davranıştır. Takip davranışı ise yüksek kaliteli çözümleri takip etmek ve bunları geliştirmek için yapılan bir davranışı temsil etmektedir [24].

Tablo 1. Gri kurt optimizasyonu kaba kodu

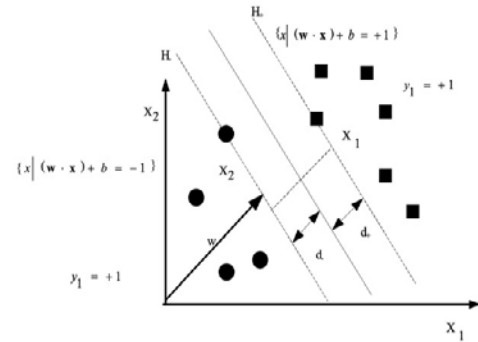
<p>Maksimum yineleme sayısını ayarlayın (L) Popülasyonu başlat X_i ($i = 1, 2, \dots, n$) Başlat α, A, ve C Kurtların uygunluk durumunu hesapla X_α = En iyi arama ajanı X_β = İkinci iyi arama ajanı X_δ = Üçüncü iyi arama ajanı</p> <p>while ($t < L$) do for her aramayı ajan do Geçerli arama aracısının konumunu güncelle end for Başlat α, A, ve C Tüm ajanların uygunluk fonksiyonunu hesapla Güncelle X_α, X_β ve X_δ $T=t+1$</p> <p>end while return X_α</p>

2.1. Çekirdek Matris ve Destek Vektör Makineleri

Çekirdek matris (Kernel matrix), matris çarpımı işlemlerinde kullanılan matematiksel bir kavramdır. Çekirdek matris, genellikle bir matrisin içerdiği bilgileri özetlemek veya işlemek için kullanılmaktadır. Bu işlem veri analizi ve makine öğrenimi uygulamaları için faydalı veri üretimi sağlamaktadır [25, 26] Örneğin veri madenciliği uygulamalarında benzerlik veya uzaklık ölçümlerini temsil etmek için kullanılırken [27] görüntü işleme uygulamalarında nesne tanıma, yüz tanıma ve diğer görüntü analizi işlemlerinde kullanılmaktadır [28].

Destek vektör makineleri istatistiksel öğrenme teorisinde kullanılan yeni bir makine öğrenme tekniğidir. Özellikle non-lineer sınıflandırma problemlerini çözmede etkilidir

[29]. SVM'nin temel mantığı, non-lineer eğitim verilerinin çekirdek matrisler aracılığıyla daha yüksek boyutlu özellik uzayına eşlenmesi işlemidir. SVM algoritmaları, çekirdek matrislerle birlikte sıralı matematiksel işlemlere tabi olurlar. SVM’de Lagrange çarpanları ve çekirdek matrislerle optimizasyon sağlayan diskriminasyon (optimal) hiperdüzlemleri (Şekil 1) oluşturulmaktadır. Lagrange çarpanları ve çekirdek matrisleri yanlış sınıflandırma olasılığı bulunmayan durumlarda lineer ayrılabilir hiperdüzlemler oluşturabilir. SVM’ler aşırı takılmayı önlediklerinden dolayı meydana gelen sorunları göz ardı etme yeteneğine sahiptir.



Şekil 1. Lineer regresyon grafiksel gösterimi [30]

SVM eğitimi her zaman evrensel olarak optimize bir çözüm uzayı arayışındadır. Lineer ayrılabilir veri uzaylarında, hiperdüzlem fonksiyonu Denklem (1)’deki gibidir.

$$w \bullet x + b = 0 \quad w \in \mathbb{R}^N, \quad b \in \mathbb{R} \quad (1)$$

Denklem (2)’de, optimize lineer bölme işlemi için iki sınıfı ayıran bir hiperdüzlem oluşturulur.

$$y_i (w \bullet x + b) \geq 1, \quad i = 1, \dots, N \quad (2)$$

SVM yaklaşımı, bu şartlarda $\|w\|$ ’yi en aza indirgeyerek en uygun hiperdüzlemi bulmaya çalışır. Burada $\|w\|$, w vektörünün Öklidyen normunu ifade eder ve hiperdüzlem ile her sınıfın en yakın veri noktaları arasındaki mesafe $1/\|w\|$ ’dir. Lagrange çarpanları α_i olarak belirlenmesiyle, SVM eğitim işlemi konveks kare problemini çözmekle eşdeğerdir. Denklem (3) evrensel optimizasyon çözümünü vermektedir.

$$w = \sum_i^N \alpha_i y_i x_i \quad (3)$$

Denklem 3’te, α_i 0’den farklı bir sayı olmak üzere x_i destek vektörü olarak adlandırılır. SVM eğitiminde her yeni x yalnızca destek vektör $\{x_i\}$ ile karşılaştırılarak karar alınır. ($\{x_i\}, i \in SV$) (Denklem (4))

$$y = \text{sign}(\sum_{i \in SV} \alpha_i y_i) (x_i \bullet x^T) + b \quad (4)$$

SVM, ayrılabilir non-lineer olan durumlarda, giriş vektörü x ’i, verilerin çekirdek işlevi tarafından belirlenen daha yüksek boyutlu bir özellik uzayına non-lineer olarak eşlemek için kullanılır. SVM algoritmaları farklı çekirdek fonksiyonlarıyla kullanılabilir. Lineer, non-lineer,

radyal baz fonksiyonu, sigmoid fonksiyon ve polinom fonksiyonları şeklinde farklı çekirdek matris fonksiyonu bulunmaktadır. Veri sınıflandırmasında yaygın olarak kullanılan tipik çekirdek işlevleri Tablo 2'de listelenmiştir.

Tablo 2. Çekirdek matrislerin formülasyonları

Çekirdek (Kernel)	$K(x, x_i)$
Lineer	$x^T \cdot x_j$
Polinom	$(x^T \cdot x_j + 1)^d$
Radyan Tabanlı Fonksiyon (RBF)	$exp(- x - x_j ^2 / 2\sigma^2)$

Çekirdek matrislerin seçimi verilere bağlı olarak yapılmaktadır. Optimum sınıflandırma sonucu elde etmek için farklı çekirdek yapısı kullanılabilir.

2.3. Önerilen Yöntem

Bu çalışma, koroner arter hastalığını teşhis etmek için etkili bir özellik seçme yöntemi olan GWO-SVM'yi önermektedir. İlk aşamada GWO yöntemi, Cleveland koroner arter veri setindeki en iyi özellikleri arayarak gereksiz ve ilgisiz özellikleri ortadan kaldırmak için kullanılır. GWO ilk olarak, popülasyon başlangıç konumlarını üretir ve ardından ayırık arama uzayında popülasyonun mevcut konumlarını günceller. GWO, popülasyonun mevcut konumlarını güncelledikten sonra, her bir bireyin uyum değerlerini değerlendirir ve en iyi performans gösteren bireyleri belirler. Bu aşamada, bireyler arasındaki benzerlikleri ve farklılıkları analiz ederek, popülasyonun genetik çeşitliliğini artırmak ve daha etkili bir arama stratejisi oluşturmak için özellik çıkarımı yapılır. İkinci aşamada, birinci aşamada elde edilen optimal özellik alt kümesine dayalı olan özellikler destek vektör makineleri ile sınıflandırılır. Destek vektör makinelerinde kullanılmış olan çekirdek fonksiyonları değiştirilerek ve GWO ile özellik seçimi yapılarak sisteminin doğruluğu ve hassasiyetinin artırılması bu çalışmada amaçlanmıştır. Özellik seçimi ile sistemin daha hızlı yanıt verebilmesi ve gereksiz verilerden kaynaklı gürültünün engellenmesi sağlanmaktadır. Önerilen yöntem Python dilinde Jupyter derleyicisinde gerçekleştirilmiştir.

Veri seti olarak UCI üzerinden edinilen Cleveland koroner arter hastalığına ait veriler kullanılmıştır. Veri setine ait özniteliklerin açıklaması Tablo 3'te yer almaktadır.

Tablo 3. Cleveland veri kümesi özellikleri

No	Özellikler	Değerler
1	Age	Yaş
2	Sex	0 Kadın/ 1 erkek
3	Cp	Göğüs ağrısı tipi
		Durum 1: Tipik anjina
		Durum 2: Atipik anjina
		Durum 3: Anjinal olmayan ağrı
4	Trestbps	Durum 4: Asemptomatik
		Kan şekeri
		Kolestrol
		(Açlık kan şekeri > 120 mg/dl)
6	Fbd	(1=Doğru; 0= Yanlış)
7	Rest-Ecg	EKG
8	Thalach	Maksimum kalp atışı
9	Exang	Egzersize bağlı angina

10	Oldpeak	Dinlenmeye göre egzersizin neden olduğu ST düşümü
11	Slope	Egzersiz sırasındaki ST'nin eğim ve tepe değeri
12	Ca	Floroskopide renklenmiş damar sayısı
13	Thal	Kusur tipi

Tablo 4'te önerilen yöntemde kullanılan parametrelerin değerleri belirtilmiştir. Uygunluk fonksiyonu için α ve β parametreleri bildirilmiştir.

Tablo 4. Cleveland veri kümesi özellikleri

Parametre	Değer
İterasyon Sayısı	100
Kurt Sayısı	5
Boyut Sayısı	13
Arama Alanı	[0,1]
Uygunluk fonksiyonundaki α	0.99
Uygunluk fonksiyonundaki β	0.01

2.3. İkili sınıflandırma için performans ölçütleri

Önerilen yöntemin performansı, doğru pozitif (TP), doğru negatif (TN), yanlış negatif (FN) ve yanlış pozitif (FP) terimlerini kullanan duyarlılık, özgüllük ve doğruluk testlerine dayalı olarak değerlendirilmiştir. Bu performans ölçütleri Denklem (5), Denklem (6) ve Denklem (7)'deki gibi hesaplanır [31].

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

$$\text{Özgüllük} = \frac{TN}{FP+TN} \times 100\% \quad (6)$$

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (7)$$

3. BULGULAR

Çalışma sırasında SVM'de kullanılan farklı çekirdek fonksiyonları ile deneyler gerçekleştirilmiştir. Çalışma sırasında destek vektör makinelerinde kullanılan çekirdek matrislerde lineer matris, polinom matrisi ve radyan tabanlı fonksiyon matrisleri kullanılmıştır.

Her matrisle deney 10 kez gerçekleştirilmiş ve en iyi sınıflandırma sonuçlarının lineer matrisin kullanıldığı destek vektör makineleri sisteminde elde edildiği görülmüştür. Çalışmada özellik seçimlerinde 7 ayrı özellik belirlenmiştir. Bu özelliklerle yapılan farklı çekirdek matrisler kullanılarak yapılan çalışmanın sonuçları Tablo 5'te verilmiştir.

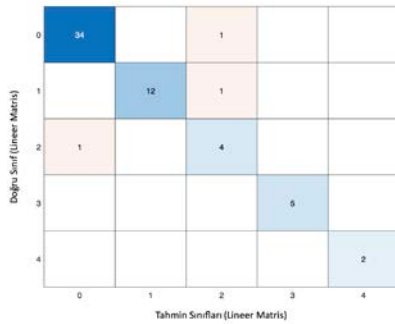
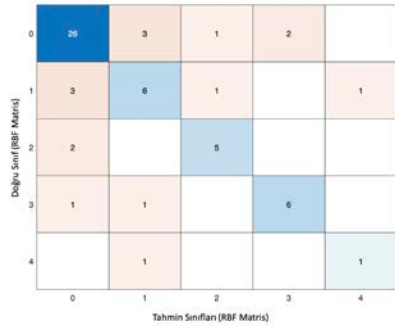
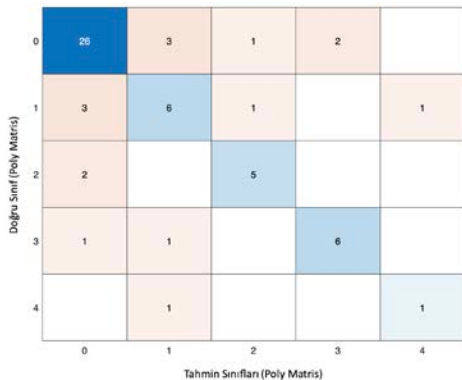
GWO-SVM'nin lineer çekirdek matris kullanılarak veri setinde kullanılan 7 ayrı özellikle en doğru sonucun elde edildiği bulunmuştur. Bu deney sonucunda veri seti teşhis sınıflandırmasının %95.36 doğruluk, %95.64 duyarlılık ve %95.91 oranında özgüllük oranına sahip olduğu görülmüştür.

GWO'da özellik seçimi gerçekleştirilmeden yalnızca SVM kullanılarak ve SVM'de çekirdek matris olarak lineer matris kullanıldığı durumda sınıflandırmadan elde edilen doğruluk oranı %59.83 olarak bulunmuştur. GWO ile özellik seçim parametrelerinin tamamlanması ardından tekrar destek vektör makineleri ile sınıflandırma gerçekleştirilmiş ve doğruluk oranı %95.36'ya yükselmiştir.

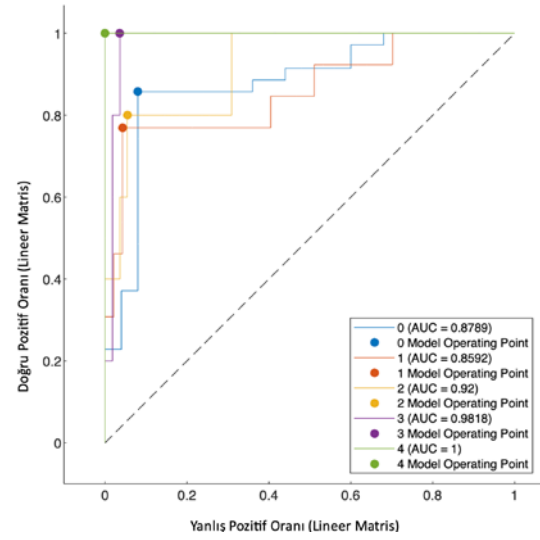
Tablo 5. Cleveland veri kümesi özellikleri

SVM (Test) Çekirdek Matris	Doğruluk	Özellik Seçimi (GWO)	SVM (Test) Çekirdek Matris	Doğruluk	Duyarlılık	Özgüllük
Lineer	%59,83	1,2,3,6,7,8,9 (Tablo 3)	Lineer	%95,36	%95,64	%95,91
Polinom	%59,83	1,2,3,6,7,8,9 (Tablo 3)	Polinom	%76,22	%79,54	%80
RBF	%63,93	1,2,3,6,7,8,9 (Tablo 3)	RBF	%87,70	%83,45	%82,92

Tablo 5'te lineer matristen elde edilen sonuçların yanı sıra, polinom matris ve RBF matris kullanılarak elde edilmiş SVM sınıflandırma sonuçları görülmektedir. Şekil 2, şekil 3 ve şekil 4'te sırasıyla lineer matris, RBF matris ve poly matrisin kullanıldığı destek vektör makinelerinden elde edilen sonuçlara ait hata matrislerine (Confusion matrix) yer verilmiştir. Şekillerde yer 0,1,2,3 ve 4 ile tanımlanmış sınıflarda; 0-sağlıklı bireyi, 1,2,3 ve 4 ise kalp hastalığı varlığını tanımlamaktadır.

**Şekil 2.** SVM sonunda elde edilen test verilerine ait hata matrisi (Lineer Kernel)**Şekil 3.** SVM sonunda elde edilen test verilerine ait hata matrisi (RBF Kernel)**Şekil 4.** SVM sonunda elde edilen test verilerine ait hata matrisi (Poly Kernel)

Şekil 6'da lineer kernel kullanılan SVM'den elde edilen ROC eğrisine ait grafik sunulmuştur.

**Şekil 5.** Lineer kernel kullanılmış SVM'ye ait ROC eğrisi

4. TARTIŞMA VE SONUÇ

Önerilen yöntem ve literatürde yapılan çalışmaların karşılaştırılması Tablo 6'da özetlenmiştir. Yöntemin tarafsız olarak karşılaştırılabilmesi için Cleveland veri setini kullanılan çalışmalar referans olarak seçilmiştir. Tablo 6'ya bakıldığında, bu çalışmada gerçekleştirilen GWO-SVM (Lineer), GWO-SVM (RBF) yöntemiyle elde edilen doğruluk değerlerinin Subanya ve arkadaşları tarafından önerilen ABC-SVM yönteminden elde edilen doğruluk değerlerine göre sırasıyla %11,18 ve %2,26 oranında artmıştır [23]. Bu çalışmada doğruluk değerleri hesaplanmadığı için duyarlılık ve özgüllük değerleri kıyaslanamamıştır.

Tablo 6. Referans çalışmalar ve bu çalışmadan elde edilen sonuçlar

Referans	Yöntem	Doğruluk	Duyarlılık	Özgüllük
[23]	ABC-SVM	%86,76	N/A	N/A
[22]	GA-FBSS	%80,00	%84,00	%75,00
[21]	GA-NN	%89,40	%88,00	%91,00
[15]	GWO-SVM	%89,83	%93,00	%91,00
Bu çalışma	GWO-SVM (Poly)	%76,22	%79,54	%80,00
Bu çalışma	GWO-SVM (Lineer)	%95,36	%95,64	%95,91
Bu çalışma	GWO-SVM (RBF)	%87,70	%83,45	%82,92

Paul ve arkadaşları tarafından önerilen GA-FBSS yöntemde doğruluk, duyarlılık ve özgüllük değerleri bu çalışmada gerçekleştirilen GWO-SVM (Lineer) yöntemine göre daha düşük başarımları vermiştir [22]. Bu çalışma ile GWO-SVM (Lineer) yöntemi kıyaslandığında; doğruluk değeri %19.2, duyarlılık değeri %13.86 ve özgüllük değeri %27.88 arttığı görülmektedir. Paul ve arkadaşları tarafından yapılan çalışmada elde edilen performans değerleri ve bu çalışmada GWO-SVM (RBF) değeri ile elde edilen sonuçlar karşılaştırıldığında bu çalışmada elde edilen doğruluk değeri %9.62, özgüllük değeri %10.56 oranında artmış ve duyarlılık değeri % 0.65 oranında azalmıştır. Paul ve arkadaşları tarafından gerçekleştirilen çalışma ve bu yöntemle gerçekleştirilen GWO-SVM (Poly) yöntemi kıyaslandığında ise GWO-SVM (Poly) yönteminin performans metriklerinin daha düşük olduğu görülmektedir.

Arabasadi ve arkadaşları tarafından yapılan çalışmada GA-NN yöntemi ile farklı veri setleri ve Cleveland veri seti bulguları değerlendirilmiştir [21]. Arabasadi ve arkadaşları tarafından yapılan çalışmada elde edilen performans metrikleri, bu çalışmada elde edilen GWO-SVM (Poly) ve GWO-SVM (RBF) yöntemine göre daha iyi performans göstermiştir. Fakat bu çalışmada gerçekleştirilen GWO-SVM (Lineer) yöntemi doğruluk, duyarlılık ve hassasiyet metriklerinde sırasıyla %6.67, %6.98 ve %5.40 daha fazla başarımları göstermiştir.

Al-Tashi ve arkadaşları [15] tarafından yapılan GWO-SVM yöntemi bu çalışma ile benzerlik göstermektedir. Bu çalışmada optimizasyon sürecinde SVM'de kullanılan çekirdek matrisler değiştirilerek bulguların performansları karşılaştırılmıştır. Elde edilen sonuçlarda bu çalışmada gerçekleştirilen GWO-SVM (Lineer) yöntemi ile Al-Tashi ve arkadaşları tarafından yapılan çalışmayla arasındaki doğruluk değerinde %6.65, duyarlılık değerinde %2.84 ve hassasiyet değerinde %5.40 daha fazla başarımları sağlamıştır.

Sonuç olarak bu çalışmada gri kurt optimizasyonuna dayalı özellik seçim yöntemi ve destek vektör makineleriyle sınıflandırma işlemi önerilmiştir. Önerilen yöntem veri setinden gereksiz ve ilgisiz özellikleri kaldırmayı sağlamıştır. Cleveland veri setinden elde edilen özellik seçimleri sonucu veri setinde bulunan 13 ayrı parametre arasından, tanı koymaya en uygun 7 parametre belirlenmiştir. Bu parametreler yaş, cinsiyet, göğüs ağrısı tipi, açlık kan şekeri, EKG verileri, maksimum kalp atış hızı ve egzersize bağlı anjina olarak tespit edilmiştir. Deneysel sonuçlar, önerilen yöntemin doğruluk, duyarlılık ve özgüllük açısından performansı arttığını göstermektedir. Gerçekleştirilen çalışmada 7 ayrı parametre ile kalp hasatlığı teşhis yöntemi üzerine yeni bir yöntem önerilmektedir ve önerilen yöntemin optimum teşhis parametresi gerektirdiğinden dolayı emsal yöntemlere göre daha hızlı ve performans metrikleri bakımından yüksek doğrulukta olduğu bulunmuştur.

Gelecekte yapılacak çalışmalarda uygulanan yöntemde bulunan öznelik seçim yöntemiyle farklı makine öğrenmesi yöntemleri birleştirilerek yapay zeka destekli sistemlerin başarımları artırılabilir.

Etik Hususlar

Etik Kurallara Uyum

Bu araştırma, planlamadan uygulamaya, veri toplama sürecinden veri analizine kadar tüm aşamalarda "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" çerçevesinde belirlenen kurallara uygunluk göstermiştir. Yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altındaki kurallarda herhangi bir ihlal gerçekleşmemiştir. Çalışmanın yazım sürecinde bilimsel etik ve alıntı kurallarına tam olarak uyulmuş, toplanan veriler üzerinde herhangi bir manipülasyon yapılmamış, ayrıca bu çalışma, başka herhangi bir akademik yayın ortamında değerlendirme için gönderilmemiştir.

Finansman

Yazarlar, bu çalışma için kamu, ticari veya kâr amacı gütmeyen sektörlerdeki fon kuruluşlarından özel bir hibe almadıklarını beyan ederler.

Çıkar çatışması

Yazarlar olarak, bu çalışma ile ilgili herhangi bir kişi veya kurumla çıkar çatışması bulunmadığını onaylıyoruz.

KAYNAKÇA

- [1] Shouman, M., Turner, T. & Stocker, R. 2012. Using data mining techniques in heart disease diagnosis and treatment. Japan-Egypt Conference on Electronics, Communications and Computers, 173-177. DOI: 10.1109/JEC-ECC.2012.6186978
- [2] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alslibi, A. I., & Gandomi, A. H. 2022. Machine learning in medical applications: A review of state-of-the-art methods. Computers in Biology and Medicine, 145, 105458. DOI: 10.1016/j.compbiomed.2022.105458
- [3] Ahsan, M. M., Siddique, Z., 2022. Machine learning-based heart disease diagnosis: A systematic literature review. Artificial Intelligence in Medicine, 128, 102289. DOI: 10.1016/j.artmed.2022.102289
- [4] Averbuch, T., Sullivan, K., Sauer, A., Mamas, M. A., Voors, A. A., Gale, C. P., Van Spall, H. G. 2022. Applications of artificial intelligence and machine learning in heart failure. European Heart Journal-Digital Health, 3(2), 311-322. DOI: 10.1093/ehjdh/ztac025
- [5] Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., Hamdi, M. 2022. Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148. DOI: 10.22452/mjcs.sp2022no1.10

- [6] Taha, B., Liza, F. R., Masud, M. A., Bepery, C., Islam, M. T., Samsuzzaman, M. 2023. BrainVisionNet: A Deep Learning-based approach to evaluate the potential of microwave imaging for classification of brain tumors. In 2023 International Conference on Next-Generation Computing, IoT and Machine Learning, 1-6.
DOI: 10.3390/healthcare9020153
- [7] Arabahmadi, M., Farahbakhsh, R., Rezazadeh, J. 2022. Deep learning for smart Healthcare-A survey on brain tumor detection from medical imaging. *Sensors*, 22(5).
DOI: 10.3390/s22051960
- [8] Shoeibi, A., Khodatars, M., Jafari, M., Ghassemi, N., Moridian, P., Alizadesani, R., Gorriz, J. M. 2022. Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: A review. *Information Fusion*.
DOI: 10.1016/j.inffus.2022.12.010
- [9] Mazhar, T., Haq, I., Ditta, A., Mohsan, S. A. H., Rehman, F., Zafar, I., Goh, L. P. W. 2023. The role of machine learning and deep learning approaches for the detection of skin cancer. In *Healthcare*, 11(3), 415.
DOI: 10.3390/healthcare11030415
- [10] Tembhurne, J. V., Hebbar, N., Patil, H. Y., Diwan, T. 2023. Skin cancer detection using ensemble of machine learning and deep learning techniques. *Multimedia Tools and Applications*, 1-24.
DOI: 10.1007/s11042-023-14697-3
- [11] Humayun, M., Khalil, M. I., Almuayqil, S. N., Jhanjhi, N. Z. 2023. Framework for detecting breast cancer risk presence using deep learning. *Electronics*, 12(2), 403.
DOI: 10.3390/electronics12020403
- [12] Krishnaiah, V., Narsimha, G., Chandra, N.S. (2015) Heart disease prediction system using data mining technique by fuzzy K-NN approach. *Emerging ICT for Bridging the Future Proceedings of the 49th Annual Convention of the Computer Society of India (CSI)*, 1, 371–384.
DOI: 10.1007/978-3-319-13728-5_42
- [13] Libby, P., Theroux, P. 2005. Pathophysiology of coronary artery disease. *Circulation*, 111(25), 3481-3488.
DOI: 10.1161/CIRCULATIONAHA.105.537878
- [14] Abdar, M., Książek, W., Acharya, U. R., Tan, R. S., Makarek, V., Pławiak, P. 2019. A new machine learning technique for an accurate diagnosis of coronary artery disease. *Computer methods and programs in biomedicine*, 179, 104992.
DOI: 10.1016/j.cmpb.2019.104992
- [15] Al-Tashi, Q., Rais, H., Jadid, S. 2019. Feature selection method based on grey wolf optimization for coronary artery disease classification. In *Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology*, 257-266.
DOI: 10.1007/978-3-319-99007-1_25
- [16] Tama, B. A., Im, S., Lee, S. 2020. Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *BioMed Research International*, 2020.
DOI: 10.1155/2020/9816142
- [17] Moturi, S., Rao, S. T., Vemuru, S. 2021. Grey wolf assisted dragonfly-based weighted rule generation for predicting heart disease and breast cancer. *Computerized Medical Imaging and Graphics*, 91, 101936.
DOI: 10.1016/j.compmedimag.2021.101936
- [18] Le, T.M., Pham, T.N., Dao, S.V. 2021. A novel wrapper-based feature selection for heart failure prediction using an adaptive particle swarm grey wolf optimization. *Enhanced Telemedicine and e-Health: Advanced IoT Enabled Soft Computing Framework*, 315-336.
DOI: 10.1007/978-3-030-70111-6_15
- [19] Deepika, D., Balaji, N. 2022. Effective heart disease prediction with Grey-wolf with Firefly algorithm-differential evolution (GF-DE) for feature selection and weighted ANN classification. *Computer Methods in Biomechanics and Biomedical Engineering*, 25, 1409 - 1427.
DOI: 10.1080/10255842.2022.2078966
- [20] Krishna, E. R., Devarakonda, N. 2023. Feature selection method based on GWO-PSO for coronary artery disease classification. *Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies*. 1-8.
DOI: 10.1109/ICAECT57570.2023.10118351
- [21] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., Yarifard, A.A. 2017. Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Comput. Methods Programs Biomed.* 141, 19–26
DOI: 10.1016/j.cmpb.2017.01.004
- [22] Paul, A.K., Shill, P.C., Rabin, M.R.I., Akhand, M. A. H. 2016. Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease. *2016 5th International Conference on Informatics, Electronics and Vision*, 145–150
DOI: 10.1109/ICIEV.2016.7759984
- [23] Subanya, B., Rajalaxmi, R. R. 2014. Feature selection using artificial bee colony for cardiovascular disease classification. 2014

International Conference on Electronics and Communication Systems.
DOI: 10.1109/ECS.2014.6892729

- [24] Mirjalili, S., Mirjalili, S. M., Lewis, A. 2014. Grey wolf optimizer. *Advances in engineering software*, 69, 46-61.
DOI: 10.1016/j.advengsoft.2013.12.007
- [25] Lanckriet, G. R., Cristianini, N., Bartlett, P., Ghaoui, L. E., Jordan, M. I. 2004. Learning the kernel matrix with semidefinite programming. *Journal of Machine learning research*, 27-72.
- [26] Weinberger, K. Q., Sha, F., Saul, L. K. 2004. Learning a kernel matrix for nonlinear dimensionality reduction. In *Proceedings of the twenty-first international conference on Machine learning* (106).
DOI: 10.1145/1015330.1015345
- [27] Huang, T. M., Kecman, V., Kopriva, I. 2006. Kernel based algorithms for mining huge data sets (1).
DOI: 10.1007/3-540-31689-2
- [28] Takeda, H., Farsiu, S., Milanfar, P. 2007. Kernel regression for image processing and reconstruction. *IEEE Transactions on image processing*, 16(2), 349-366.
DOI: 10.1109/tip.2006.888330
- [29] Zhang, K., Lan, L., Wang, Z., Moerchen, F. 2012. Scaling up kernel SVM on limited resources: A low-rank linearization approach. In *Artificial intelligence and statistics*, 1425-1434.
- [30] Astuti, W., Fadli, A., Tan, S., Akmeliawati, R. 2019. Brain signal recognition system based on One-Against-One Multiclass Support Vector Machines: a comparison with Multiclass Neural Network. In *Journal of Physics: Conference Series*, 1367(1), 012027.
DOI: 10.1088/1742-6596/1367/1/012027
- [31] Cengil, E., Çınar, A. 2020. Göğüs verileri metrikleri üzerinden kanser sınıflandırılması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(2), 513-519.
DOI: 10.24012/dumf.578606