



ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ

INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)

URL: <https://dergipark.org.tr/ij3dptdi>

RASGELE ORMAN VE İKİLİ PARÇACIK SÜRÜ ZEKÂSI YÖNTEMİYLE KALP YETMEZLİĞİ HASTALIĞINDAKİ ÖLÜM RİSKİNİN TAHMİNLENMESİ

FORECASTING THE RISK OF DEATH IN HEART
FAILURE DISEASE BY USING RANDOM FOREST AND
BINARY PARTICLE SWARM INTELLIGENCE

Yazarlar (Authors): Osamah Khaled Musleh Salman *, Bekir Aksoy 


Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article): Salman O.K.M., Aksoy B., “Rasgele Orman ve İkili Parçacık Sürü Zekâsı Yöntemiyle Kalp Yetmezliği Hastalığındaki Ölüm Riskinin Tahminlenmesi” *Int. J. of 3D Printing Tech. Dig. Ind.*, 6(3): 416-428, (2022).

DOI: 10.46519/ij3dptdi.982670

Araştırma Makale/ Research Article

Erişim Linki: (To link to this article): <https://dergipark.org.tr/en/pub/ij3dptdi/archive>

RASGELE ORMAN VE İKİLİ PARÇACIK SÜRÜ ZEKÂSI YÖNTEMİYLE KALP YETMEZLİĞİ HASTALIĞINDAKİ ÖLÜM RİSKİNİN TAHMİNLENMESİ

Osamah Khaled Musleh Salman^a , Bekir Aksoy^a 

^a Isparta Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Mekatronik Mühendisliği Bölümü, TÜRKİYE

* Sorumlu Yazar: osamahkms1994@gmail.com

(Received: 13.08.2021; Revised: 29.09.2021; Accepted: 05.12.2022)

ÖZ

Son yıllarda kalp hastalıklarının hızla artışı ile birlikte bu hastalıklardan ölen insanların sayısı da hızla artmaktadır. Bu nedenle hastalıkların erken teşhisi, insan yaşam ömrünü arttırmada oldukça önemlidir. Çalışmada 2015 yılı nisan ve aralık aylarını kapsayan kalp yetmezliği hastalarına ait toplam 299 klinik hastaya ait kayıtlar açık erişimli internet sitesinden alınarak kullanılmıştır. Açık erişimli internet sitesinden elde edilen veriler veri ön işleme aşamasından geçirilerek yapay zekâ ile analiz için hazır hale getirilmiştir. Veri seti ilk olarak Rasgele Orman algoritmasının hiper parametreleri belirlenmiştir. Elde edilen hiper parametreler, ikili parçacık sürü zekâ yöntemi kullanılarak veri seti üzerinde global bir özellik seçimi gerçekleştirilmiştir. Elde edilen sonuçlar tekrar rasgele orman algoritması ve 10 kat 10 iterasyonu ile çapraz doğrulama ile eğitilerek ölüm risk durumunu gösteren en uygun model elde edilmiştir. Elde edilen modelin 100 eğitimin ortalama doğruluk oranı ise %79.66 olarak belirlenmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, Rasgele Orman Algoritması, Özellik Seçimi, Optimizasyon, İkili Parçacık Sürü Optimizasyonu.

FORECASTING THE RISK OF DEATH IN HEART FAILURE DISEASE BY USING RANDOM FOREST AND BINARY PARTICLE SWARM INTELLIGENCE

ABSTRACT

With the rapid increase of heart diseases in recent years, the number of people who die from these diseases is also increasing rapidly. Therefore, early diagnosis of diseases is very important in increasing human life span. In the study, records of 299 clinical patients with heart failure, covering the months of April and December 2015, were taken from the open access website and used. The data obtained from the open access website were made ready for analysis with artificial intelligence by passing through the data preprocessing stage. Firstly, the data set was determined by hyper parameters of the Random Forest algorithm. Using the obtained hyper parameters, binary particle swarm intelligence method, a global feature selection was made from the data set. The obtained results were trained again with random forest algorithm and 10 times 10 iteration cross validation, and the most appropriate model showing death risk situation was obtained. The average accuracy rate of 100 trainings of the model obtained was determined as 79.66%.

Keywords: Heart Failure, Machine Learning, Random Forest, Feature Selection, Optimization, BPSO.

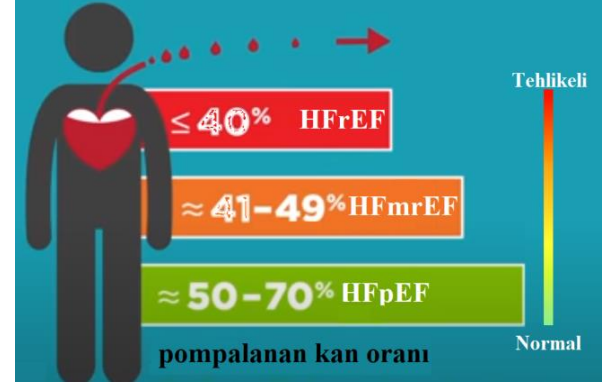
1. GİRİŞ

Bulaşıcı olmayan kronik hastalıklar, 21. yüzyılda toplumun ekonomisi, sağlığı ve yaşam kalitesi üzerinde büyük bir etkisi olan yaygın sorun haline gelmiştir [1]. Dünyadaki tüm ölümlerin %63'ü bulaşıcı olmayan hastalıklardan kaynaklanmaktadır [2]. Kronik bir hastalık olan kalp yetmezliği (KY)'de hastaların yaşam kalitesi üzerinde olumsuz etkilere neden olan önemli bir sağlık sorunudur. KY hastalığı, yaşam boyu tedavi gereksinimi, hastanede uzun süreli kalma, karmaşık ve oldukça pahalı cihaz tedavisi uygulamaları sebebiyle sağlık ekonomisi ve aile bütçeleri üzerinde büyük mali yüklereden neden olan bir hastalıktır. [3,4]. KY dünya çapında yaklaşık olarak 26 milyon insanı etkileyen küresel bir hastalıktır ve görülme sıklığı gün geçtikçe artmaktadır [1]. KY, gelişmiş ülkelerde erişkin nüfusunun %1-2 oranında görülmekte ve özellikle 70 yaş ve üzerindeki kişilerde ise bu oran %10'un üzerine çıkmaktadır [5]. Türkiye'de Kardiyoloji Derneğinin yapmış olduğu çalışmada ülkemizde 2 milyonun üzerinde insan KY hastalığı ile yaşamaktadır [6]. KY hastalığı morbidite ve mortalitede artışa sebep olması ve dolayısıyla uzun süreli hastanelerde kalma ve bunun sonucu olarak da artan maliyet oranı nedeniyle erken dönemde teşhis edilmesi oldukça önemlidir [2].

KY hastalığı, yüksek juguler venöz basınç, pulmoner ve periferik ödem fonksiyonel kalp anormalliklerinin gibi belirtiler ile birlikte nefes darlığı, ayak bileği şişmesi ve yorgunluk gibi tipik semptomlarla karakterize edilen klinik bir sendromdur [7]. Diğer bir ifadeyle, diyabet, yüksek tansiyon gibi rahatsızlıklarından dolayı kalbin vücuda yeterince kan pompalamadığında ortaya çıkan bir hastalıktır [8].

KY hastalığını tanımlamak için kullanılan ana terminoloji tarihseldir ve sol ventrikül ejeksiyon fraksiyonu (SVEF) ölçümüne dayanmaktadır. Şekil 1'de KY hastalığı, SVEF değerine göre üç sınıfa ayrıldığı görülmektedir. Bu sınıflardan ilki olan SVEF değeri (\geq 50), korunmuş enjeksiyon fraksiyonu (İng. Heart Failure with Preserved Ejection Fraction HFpEF) olarak adlandırılan kalp yetmezliğidir. İkinci değer olan SVEF değeri ($<$ 40), azaltılmış enjeksiyon fraksiyonu (İng. Heart Failure with Reduced Ejection Fraction HFrEF) kalp yetmezliğidir. Son sınıf olan SVEF değeri (% 40-49 aralığında), orta enjeksiyon fraksiyonu (

İng. Heart Failure with Mid-range Ejection Fraction HFmrEF) olarak tanımlanan sınıftır. 1990 yılından sonra yayımlanan klinik çalışmaların çoğunda SVEF temelli hastaları seçilmiştir. Bu çalışmalarda HFrEF'li hastalarda tedavilerin hem morbidite hem de mortaliteyi azalttığı görülmektedir [9-13].



Şekil 1. SVEF değerine göre kalp yetmezliği sınıfları.

KY hastalığının erken evrelerde teşhisi oldukça zordur. Her ne kadar belirtiler hastayı tıbbi yardım almaya yönlendirse de KY hastalığının belirtileri tansiyon, diyabet gibi hastalıklar ile beraber ortaya çıktığı için teşhisi oldukça zordur. Çizelge 1'de KY hastalığı için ortopne, paroksizmal nokturnal dispne gibi belirtileri ait bilgiler verilmiştir [13-14].

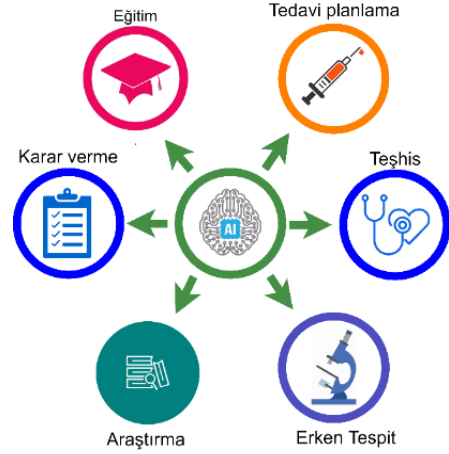
Çizelge1. Kalp yetmezliğin tipik semptomları ve bulguları [13-14]

Tipik Semptomlar	Bulgular
Nefes darlığı	Juguler venöz basıçta artış
Ortopne	Hepatojuguler reflü
Paroksizmal nokturnal dispne	Üçüncü kalp sesi (Gallop ritmi)
Azaltılmış egzersiz kapasitesi	Apikal impulsun laterale yer değiştirmesi
Halsizlik, yorgunluk	Kalp Seslerinde üfürüm
Ayak bileğinde şişlik	

Son yıllarda bilgi ve iletişim teknolojilerinde yaşanan hızlı gelişim ile birlikte birçok alanda geleneksel yöntemlerin yerini dijital sistemler almaktadır. Sağlık sektörü bu uygulamaların kullanıldığı önemli alanlardan birisidir. Özellikle sağlık işlemlerinde bilgisayarlı kontrol sistemine geçiş ile birlikte sağlık hizmetlerinin önemli bir hız artışı olmuştur. Verilerin fiziksel sınırlarının ortadan kalkması ile birlikte, sağlık hizmetlerinin etkinliği daha da artmıştır [15]. Elektronik Sağlık Kaydı

(ESK), uzman hekimlere hastaların durumunu değerlendirme ve tedavi planı geliştirilmesinde, yardımcı olan verilerin saklandığı dijital bir platformdur. Elde edilen veriler hastaların geçmiş tıbbi öykülerini, şikâyetlerini, fiziksel muayene ve bulgularını içermektedir. ESK ile birlikte; koruyucu önlemler, akut hastalıkların tedavisi ve kronik hastalıklarda yaşam boyu sağlık hizmetinin etkin bir şekilde kullanılması sağlanmaktadır [16]. Ayrıca, ESK, karar destek sistemleri ve klinik uyarı sistemlerinin yardımıyla, tıbbi hataların ve yan etkilerin azaltılmasına yardımcı olmaktadır. Buna ek olarak, tanı ve tedavi yöntemlerinin tekrarlanmasını engelleyerek zaman ve maliyet açısından önemli katkılar sağlamaktadır [17]. Sağlık verileri toplandığında, doğru tedavi planlamasının yapılması için verilerin analiz edilmesi oldukça önemlidir. Verilerin analizi için Yapay Zekâ (Artificial Intelligence AI) yöntemleri sıklıkla kullanılmaktadır.

AI, laboratuvar testlerini, röntgen, Bilgisayarlı Tomografi (BT) taramaları ve veri girişi gibi birçok tıbbi görevleri analiz etmede sıklıkla kullanılan bir yöntemdir. [18]. Şekil 2’de AI’nın sağlık sektöründe eğitim, tedavi, teşhis, karar verme, araştırma ve erken teşhis gibi birçok alanda kullanıldığı görülmektedir. AI birçok farklı tanımı olmasına rağmen temel olarak; insan gibi düşünme karar verme yeteneğine sahip olan donanımsal ve yazılımsal sistemler bütünü olarak ifade edilmektedir. AI’nın alt dallarından biri olan makine öğrenmesi; herhangi bir problemin o probleme ait verilere göre modelleyen bilgisayar algoritmalarının genel adıdır. Mevcut veri seti ve kullanılan algoritma ile en yüksek performansı elde etmek üzere kurulmuş makine öğrenmesi yöntemlerinden bazıları; k-en yakın komşu algoritması, Naive Bayes algoritması, karar ağaçları, destek vektör makineleri, rastgele orman ve yapay sinir ağlarıdır. Bu algoritmalar ile tahminleme, sınıflandırma ve kümeleme işlemleri yapılabilmektedir [19-21].



Şekil 2. Yapay zekâ sağlıkta kullanım alanları.

Sınıflandırma ve regresyon gibi makine öğrenmesi problemlerinde, hedef verisi (çıkış parametreleri) tanımlamak için veri setine çok sayıda özellik eklenmiş olabilir. Ancak çok sayıda özellik, sınıflandırma ve regresyon işlemlerinde büyük bir engel olan “the curse of dimensionality” olarak bilinen soruna neden olur. Özellik seçimi, büyük veri boyutu ile ortaya çıkan sorunu çözmek için temel ve yaygın olarak kullanılan bir tekniktir. Özellik seçimi ile ilgisiz ve gereksiz özellikler eğitim veri setinden çıkartılarak özelliklerin sayısını azaltmaktadır. Bu nedenle özellik seçimi eğitim başarısının artması ve işlem zamanının azaltılmasında önemli bir rol oynamaktadır [22]. Özellik seçim algoritmaları, makine öğrenmesinin performansını iyileştirmesi için farklı özellik kombinasyonlarını arama alanını araştırmaktadır. Özellik seçiminde “grid search” gibi farklı arama teknikleri kullanılmaktadır. Ancak bu tekniklerde yerel optimal ve/veya yüksek hesaplama maliyeti sorunları oluşabilmektedir [23,24]. Bu sorunları çözmeye global arama yeteneklerine sahip olan evrimsel hesaplama teknikleri kullanılmaktadır [22]. Bu teknikler arasında Genetik Algoritma (Genetic Algorithm -GA), Evrimsel Algoritma (Evolutionary Algorithm - EA), Karınca Kolonisi Optimizasyonu (Ant Colony Optimization - ACO), Yapay Arı Kolonisi (Artificial Bee Colony - ABC), Parçacık Sürüsü Optimizasyonu (Particle Swarm Optimization - PSO) ve Diferansiyel Evrim (Differential Evolution - DE) gibi farklı evrimsel hesaplama tekniklerinin özellik seçiminde başarı gösterdiği görülmüştür [25-28].

Çalışmada, 2015 yılı nisan ve aralık aylarına ait

KY hastalığına ait klinik kayıtlar alınarak hastaların ölüm riskleri yapay zekâ yöntemleri ile belirlenmeye çalışılmıştır. İlk olarak RF algoritmasındaki hiper parametreler değiştirilerek en uygun hiper parametre değerleri belirlenmiştir. İkinci aşamada ise belirlenen hiper parametre değerlerine göre BPSO algoritması kullanılarak KY hastalığına ait en önemli özellikler belirlenmiştir. Son aşamada ise belirlenen önemli özellikler tekrar RF algoritması ile eğitilerek KY hastalığını %80 doğruluk oranında doğru tahminleyen bir model önerilmiştir.

2. MATERYAL VE METOT

2.1. Materyal

Çalışmada açık kaynak erişimli internet sitesinden (UCI machine learning repository data sets) alınan ve toplam 299 kayıt içeren “Heart failure clinical record” isimli veri seti kullanılmıştır. Kullanılan veri seti, RF ve BPSO algoritmaları kullanılarak giriş verilerine bağlı olarak ölüm olasılığı sınıflandırılarak uygun bir Karar Destek Sistemi önerilmiştir.

Çizelge 2. KY Veri setinin özellikleri ve istatistiksel bilgileri.

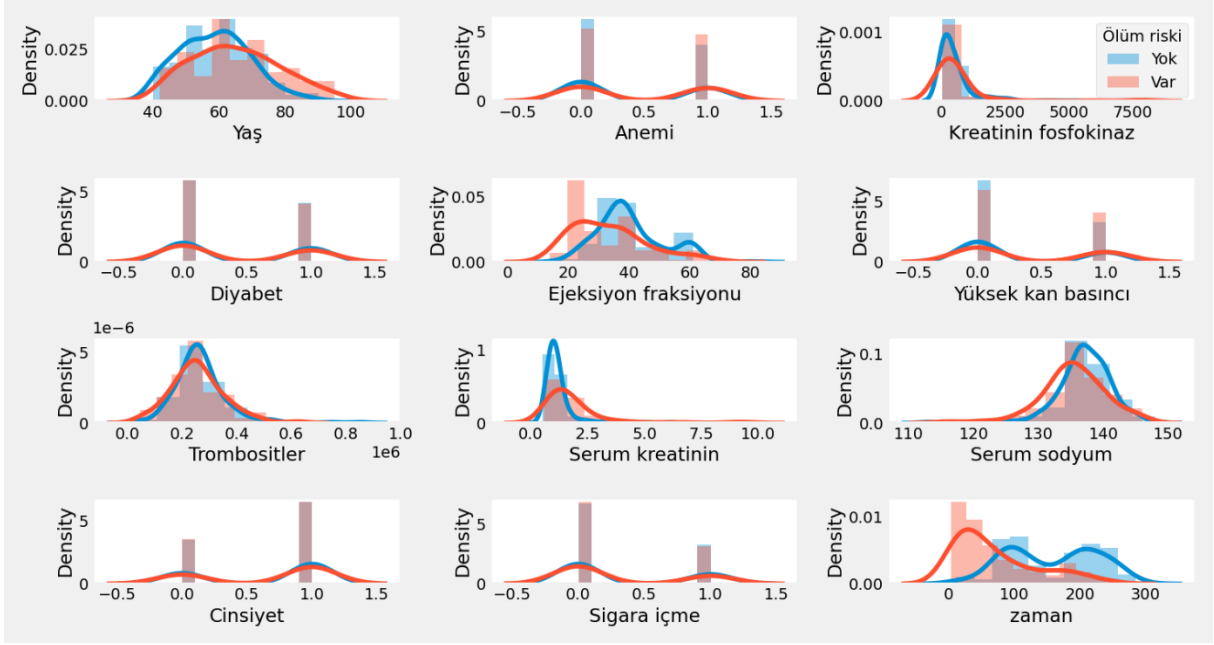
Özellik (indeks)	Anlamı	Türü	Aralığı	Ortalama	Standart sapma
Yaş (0)	Hastanın yaşı	Tam	[40,95]	60.83	11.895
Anemi (1)	Kırmızı kan hücreleri veya hemoglobin azalması	İkili	{0,1}	0.431	0.496
Kreatinin fosfokinaz (2)	Kandaki CPK enziminin seviyesi	Tam	[23,7861]	581.84	970.29
Diyabet (3)	Şeker hastalığın durumu	İkili	{0,1}	0.418	0.494
Ejeksiyon fraksiyonu (4)	Pompalanan kan oranı	Tam	[14,80]	38.08	11.84
Yüksek kan basıncı (5)	Hipertansiyonu durumu	İkili	{0,1}	0.351	0.478
Trombositler (6)	Kandaki trombositler	Tam	[25100, 850000]	263358	97804
Serum kreatinin (7)	Kandaki serum kreatinin seviyesi	Ondalık	[0.5,9.4]	1.394	1.03
Serum sodyum (8)	Kandaki serum sodyum seviyesi	Tam	[113,148]	136.63	4.41
Cinsiyet (9)	Erkek - kadın	İkili	{0,1}	0.649	0.478
Sigara içme (10)	Sigara içme durumu	İkili	{0,1}	0.321	0.468
Zaman (kullanılmadı)	Takip süresi	Tam	[4,285]	130.26	77.61
Ölüm riski (hedef)	Ölüm durumu	İkili	{0,1}	0.32	0.468

Çizelge 2 incelendiğinde zamansal olarak hastaların ortalama 4 ile 285 gün arasında hastane kaldıkları görülmektedir. Ancak hastanede kalma süresi KY hastalarındaki ölüm riski inceleneceği için oluşturulan AI modeli zamandan bağımsız olarak modellenmiştir. Böylece zaman parametresi oluşturulan modelin doğruluğunu arttırmasına rağmen kullanılması mantıksal olarak yanlış bir parametredir. Diğer bir ifadeyle zaman

2.1.1. Veri seti

Çalışmada kullanılan veri setinde KY hastalarının ölüm riskine karşılık gelen özelliklerin tespiti için ‘UCI Heart failure clinical records’ veri seti kullanılmıştır. Kullanılan veri setinde hastaların takip dönemlerinde toplanan 299 KY hastasına ait veri bulunmaktadır. Veri setinde Çalışmada kullanılan veri setinde KY hastalarının ölüm riskine karşılık gelen özelliklerin tespiti için ‘UCI Heart failure clinical records’ veri seti kullanılmıştır. Kullanılan veri setinde hastaların takip dönemlerinde toplanan 299 KY hastasına ait veri bulunmaktadır. Veri setinde her hastaya ait 13 tıbbi klinik kayıtları içermektedir. Bu veri seti Faisalabad Kardiyoloji Enstitüsü ve Faisalabad'daki (Punjab, Pakistan) Allied Hospital'da 2015 yılı nisan – aralık aylarına aittir. [29-30]. Çizelge 2' de veri setinin özellikleri ve istatistiksel bilgileri detaylı bir şekilde verilmiştir.

parametresi kullanıldığında model hastanın kalma süresine bağlı olarak ölüm riskini sınıflandırır. Yeni kalp yetmezliği geçiren bir kişinin modele göre daha riskli olarak sınıflandırılacaktır. Bu durum engellemek için zaman parametresi eğitimde kullanılmamıştır. Şekil 3'de hedef parametresi olan ölüm riski için giriş parametrelerinin dağılımına ait grafikler verilmiştir.



Şekil 3. Ölüm riskine göre verilerin dağılımı.

Şekil 3 incelendiğinde anemi, diyabet, yüksek kan basıncı, cinsiyet, kreatin fosfokinaz ve sigara içme değerlerine bağlı KY hastalığı için ölüm riski sonucu üzerinde etkisinin az olduğu görülmektedir. Ancak yaş, serum kreatin değerlerine bağlı KY hastalığındaki ölüm riskinde etkili olduğu ve parametre değeri arttıkça ölüm riskinin arttığı görülmektedir. Enjeksiyon fraksiyonu, serum sodyum, zaman ve Trombosit parametre değerlerine bağlı KY hastalığındaki ölüm riskinde etkili olduğu ve parametre değeri arttıkça ölüm riskinin azaldığı görülmektedir.

2.1.2. Random Forest (RF) algoritması

RF, Leo Breiman [31] tarafından oluşturulan topluluk (Ensemble) sınıflandırma yöntemleri arasında yer alan bir makine öğrenme algoritmasıdır [32]. Topluluk sınıflandırma teknikleri, sadece bir tane sınıflandırıcı üretmek yerine birden fazla sınıflandırıcı üretmektedir. Üretilen sınıflandırıcıların tahminlerinden elde edilen sonuçlara göre veriyi sınıflandıran bir algoritmadır [33]. RF algoritmasının çalışma prensibi birden fazla karar ağacı oluşturulup, karar ağacındaki ortalama değerler alınarak sonuç üretilmektedir [34]. Eğitilmiş olan k ağaçları, Denklem 1'de tanımlanan bir RF modelinde toplanmaktadır [35].

$$H(X, \theta_j) = \sum_{i=0}^k h_i(x, \theta_j) \quad , (j=1,2, \dots, m) \quad (1)$$

Denklemden $H(X, \theta_j)$ bir meta karar ağacı sınıflandırıcısıdır. x , eğitim veri kümesinin girdi özelliği vektörünü temsil etmektedir ve θ_j , ağacın büyüme sürecini belirleyen bağımsız ve aynı şekilde dağıtılmış rastgele bir vektördür. Bağımsız olarak rastgele oluşturulan k karar ağaçlarının bir araya gelmesi sonucu RF algoritması oluşmaktadır. Test veri kümesine ait örneklerden her biri ormanı oluşturan karar ağaçları tarafından tahmin edilmektedir. Bu ağaçların elde edilen sonuçlara göre sınıflandırma işlemi gerçekleştirilmektedir [35]. Geleneksel karar ağaçları algoritmasında karşılaşılan en büyük problemler arasında aşırı öğrenme (overfitting) yer almaktadır. RF algoritması, bu problemi ortadan kaldırmak için hem veri setini hem de öznelikleri çok sayıda parçaya bölerek birden fazla ağaçta işlem gerçekleştirmektedir [36]. RF, büyük veri tabanlarında özelliklerin eksiksiz işlenmesinde ve sonuç üzerindeki etkili özelliklerin ön plana çıkarılmasında oldukça verimli çalışmaktadır. Böylece sınıflandırma işleminde düşük hatalar üretilerek doğruluk oranını artırmaktadır[37].

2.1.3. Parçacık sürü optimizasyonu (PSO)

PSO, Eberhart ve Kennedy tarafından geliştirilen popülasyon tabanlı bir optimizasyon algoritmasıdır. Sürü halinde hareket eden kuş, balık ve karınca gibi hayvanların davranışlarından ilham alınarak geliştirilmiştir. [38]. PSO her bir kuşun "parçacık" olarak adlandırıldığı arama uzayında rastgele bir kuş sürüsünün ilk değer ataması yapılarak başlatılır. Her bir parçacık

en iyi konumunu saklar ve küresel optimuma ulaşmak için her adımda konumunu günceller. Bu güncellemeler yapılırken her bir parçacığın elde ettiği en iyi konumu veren p_{best} ve tüm sürünün elde ettiği en başarılı konumu g_{best} değerleri dikkate alınarak işlemler gerçekleştirilir. Bu işlemleri gerçekleştirmek için parçacık popülasyon matrisi kullanılmaktadır. Parçacık popülasyon matrisinde D adet özelliğe sahip i . parçacık $X_i = [X_{i1}, X_{i2}, \dots, X_{iD}]$ şeklinde gösterilir. Her iterasyon için p_{best} ve g_{best} elde edildikten sonra parçacığın konumu ve hızı eşitlik 2 ve eşitlik 3'de verilen matematiksel ifadeler kullanılarak hesaplanmaktadır [39-45].

$$V_{id}^{k+1} = W \cdot V_{id}^k + c_1 \cdot rand_{1i}(p_{best_{id}} - X_{id}^k) + c_2 \cdot rand_{2i}(g_{best_d} - X_{id}^k) \quad (2)$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \quad (3)$$

Eşitliklerde X_{id} konum ve V_{id} hız değerlerini verirken, $rand_1$ ve $rand_2$ değerleri 0-1 arasında rasgele üretilmiş sayılardır. W eylemsizlik ağırlık değerini, c_1 , c_2 ölçeklendirme faktörlerini ve K değeri de iterasyon sayısını göstermektedir. Ölçeklendirme faktörleri olan c_1 , c_2 her parçacığı p_{best} ve g_{best} değerlerine doğru çeker. Eylemsizlik ağırlığı olan W ise her iterasyonda doğrusal olarak azaltılmalı ve birden küçük seçilmelidir [39-45].

1997 yılında Eberhart ve Kennedy tarafından özellik seçimi gibi ayrık arama alan sahip problemler için bir ikili parçacık sürüsü optimizasyonu (BPSO) geliştirilmiştir. BPSO'da, hız eşitlik 2 ve konum eşitlik 4'deki matematik ifadelerine göre güncellenmektedir. BPSO 'da X_{id} , p_{best} ve g_{best} değerleri 0 veya 1 ayrık değeri olarak hesaplanmaktadır [22].

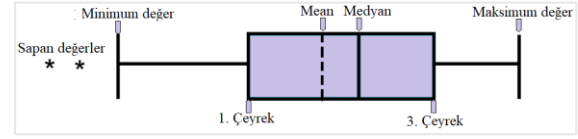
$$X_{id} = \begin{cases} 1, & \text{eğer } rand() < s(V_{id}) \\ 0, & \text{aksi takdirde} \end{cases} \quad (4)$$

Where $s(V_{id}) = \frac{1}{1+e^{-V_{id}}}$

2.1.4. Kutu bıyık grafiği

Kutu bıyık bir verinin dağılımını göstermek için kullanılan bir grafik türüdür. Kutu grafikleri, dağılımın şekli, merkezi eğilimi ve değişkenlerin yayılım düzeyini göstermesi açısından oldukça kullanışlıdır. Şekil 4

görüldüğü gibi kutu bıyık grafiği bir verinin minimum, birinci çeyrek, ortalama, medyan, üçüncü çeyrek ve maksimum değer olmak üzere toplam altı özelliğinden oluşmaktadır.



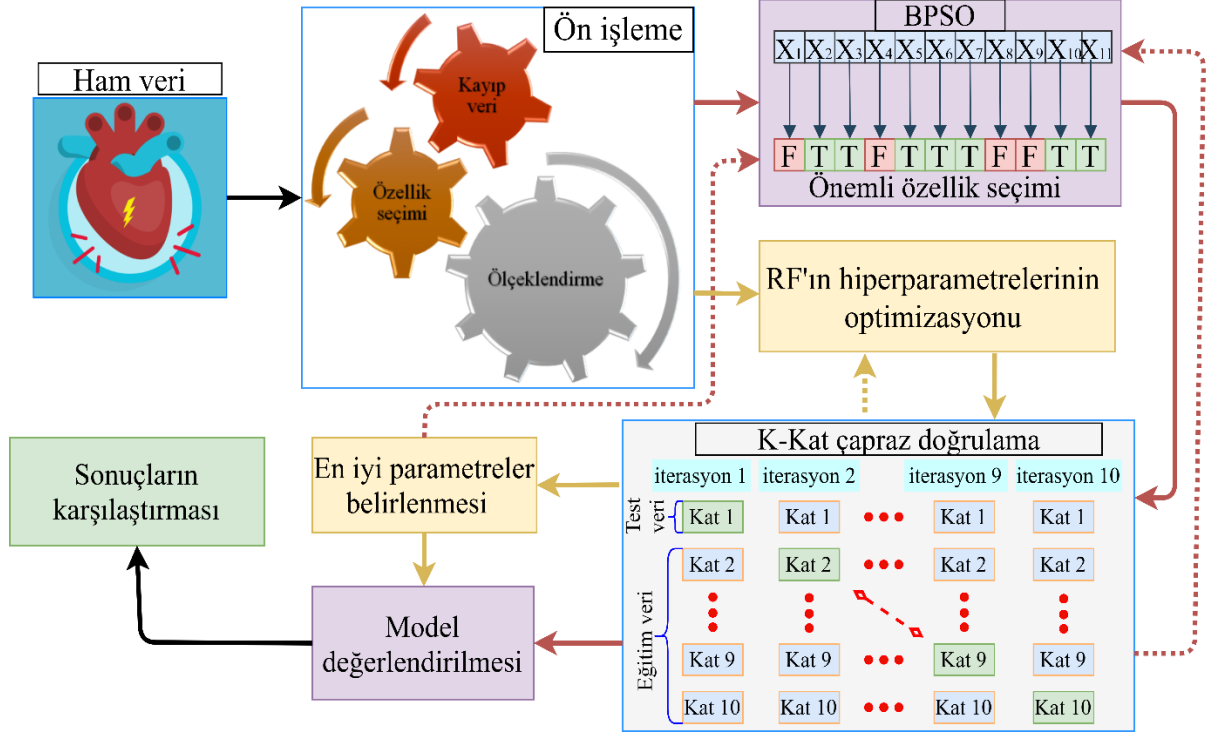
Şekil 4. Kutu bıyık grafiğinin şekli ve iç bilgileri.

2.2. Metot

Çalışmada KY hastalığına ait ölüm riski için oluşturulan AI modelinin iş akışı diyagramı şekil 5'de verilmiştir. İlk olarak açık erişimli internet sitesinden KY hastalığının kinlik kayıtlarının ham veri seti alınmıştır. İkinci aşamada ise veri seti üzerinde özellik seçimi, kayıp veri analizi ve ölçeklendirme ön işlemleri gerçekleştirilmiştir. Ön işleme aşamalarından olan kayıp veri analizi yapıldığında veri setinde herhangi bir kayıp veri görülmemiştir. Özellik seçiminde ise zaman giriş parametrelerinden çıkartılarak model de oluşabilecek olan mantıksal hata giderilmiştir. Ölçeklendirme aşamasında ise AI modelindeki doğruluğu arttırabilmek için sürekli giriş değişkenlerin değerleri 0 ile 1 aralığında ölçeklendirilmiştir. Çalışmanın üçüncü aşamasında ise RF algoritmasının hiper parametreleri olan ağaç sayısı değerleri 0 ile 100 arasında beşer beşer artırılmış, ağacın maksimum derinliği ise 2 ile 16 arasında ikişer ikişer artırılarak optimize edilmiştir. Optimizasyon işlemi 0 ile 100 arasında olan toplam 20 adet ağaç sayısının 2 ile 16 arasında değişen toplam 8 ağacın maksimum derinlik değerleri üzerinde gerçekleştirilmiştir. Böylece tüm parametrelerin arama uzayında test ederek toplam $20 \times 8 = 160$ adet 10 kat çapraz doğrulama ile eğitim gerçekleştirilerek en uygun hiper parametreler tespit edilmiştir. Çapraz doğrulama işleminde veriler 10 kate bölünüp 10 iterasyon işlemi gerçekleştirilmiştir. Böylece her bir eğitim için toplam 10 alt eğitim işlemi gerçekleştirilmiştir. Bir sonraki aşamada ise sınıflandırma işleminin doğruluğunun arttırmak adına BPSO arama algoritması ile belirlenen RF hiper parametreler ve veri ön işleme aşamasından elde edilen sonuçlar kullanılarak özellik seçimi yapılmıştır. Uygulanan BPSO algoritmasında maksimum hız değeri (V_{max}) 4, minimum hız değeri (V_{min}) -4, eylemsizlik ağırlığı (W) 0,5 ve ölçeklendirme faktörleri (c_1 , c_2) ise 2 olarak alınmıştır. BPSO algoritmasında parçacık boyutu 0 ile 100

arasında beşer beşer artırılarak, iterasyon sayısı ise 10 ile 200 arasında onar onar artırılarak en uygun özellik seçimi yapılmıştır. Elde edilen özellik seçimine göre tekrar RF ve 10 kat çapraz

doğrulama ile eğitim yapılmıştır. Elde edilen tüm modeller değerlendirilerek en uygun model ve KY hastalığında ölüm riskini etkileyen en önemli özellikler belirlenmiştir.



Şekil 5. İş akış diyagramı.

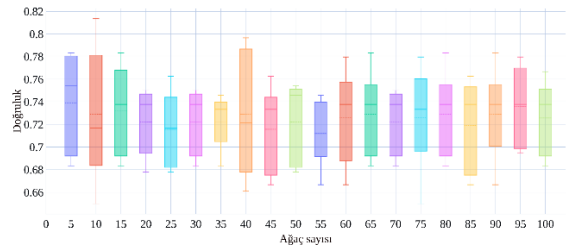
3. ARAŞTIRMA BULGULARI

Çalışmada kullanılan RF ve BPSO algoritmalarının hiper parametre aralıkları çizelge 3'de verilmiştir.

Çizelge 3. RF ve BPSO algoritmalarının hiper parametrelerinin değerleri

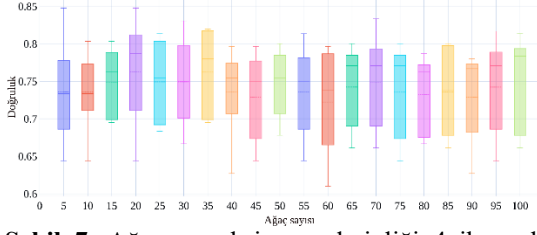
Algoritma	Hiper parametre	Değer aralığı
RF	Ağaç sayısı	[0,5,10,15,...,90,95,100]
	Maksimum derinlik	[2,4,6,8,10,12,14,16]
BPSO	İterasyon sayısı	[10,20,30,...,180,190,200]
	Parçacık boyutu	[0,5,10,15,...,90,95,100]

Çizelge 3'de belirlenen hiper parametre aralıklarına göre çalışmada kullanılan RF algoritması ile ilk olarak ağacın maksimum derinliği 2 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 6'da verilmiştir.



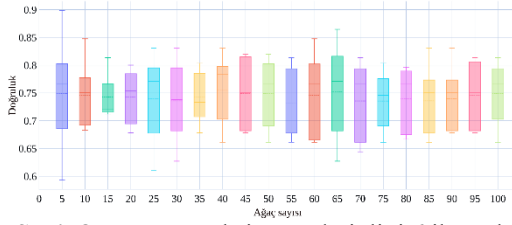
Şekil 6. Ağacın maksimum derinliği 2 iken elde edilen modellerin performans karşılaştırılması.

Şekil 6 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 75 olduğu modelde 0.7124 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 95 olduğu modelde 0.7358 olarak elde edilmiştir. İkinci aşamada ise maksimum derinliği 4 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 7'de verilmiştir.



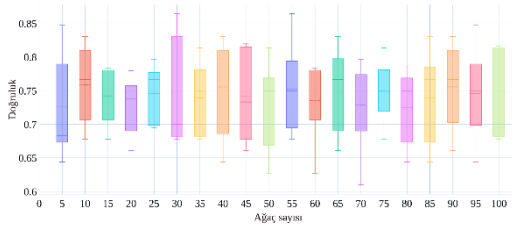
Şekil 7. Ağacın maksimum derinliği 4 iken elde edilen modellerin performans karşılaştırılması.

Şekil 7 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 60 olduğu modelde 0.7222 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 20 olduğu modelde 0.7624 olarak elde edilmiştir. Üçüncü aşamada ise maksimum derinliği 6 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 8'de verilmiştir.



Şekil 8. Ağacın maksimum derinliği 6 iken elde edilen modellerin performans karşılaştırılması.

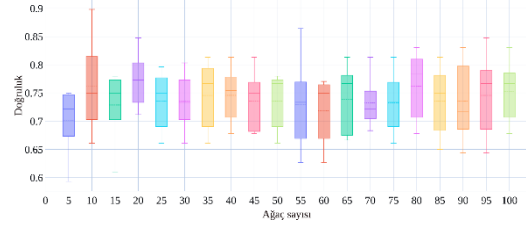
Şekil 8 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 55 olduğu modelde 0.7323 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 40 olduğu modelde 0.7557 olarak elde edilmiştir. Dördüncü aşamada ise maksimum derinliği 8 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 9'da verilmiştir.



Şekil 9. Ağacın maksimum derinliği 8 iken elde edilen modellerin performans karşılaştırılması.

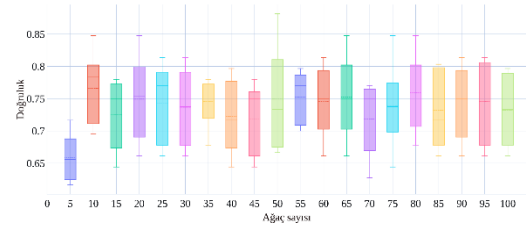
Şekil 9 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 80 olduğu

modelde 0.7254 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 10 olduğu modelde 0.7590 olarak elde edilmiştir. Beşinci aşamada ise maksimum derinliği 10 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 10'da verilmiştir.



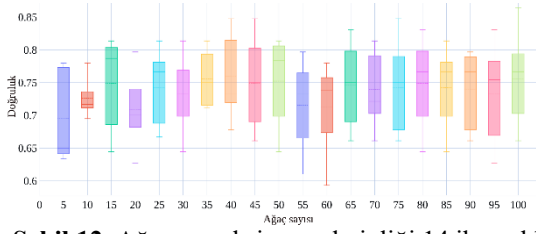
Şekil 10. Ağacın maksimum derinliği 10 iken elde edilen modellerin performans karşılaştırılması.

Şekil 10 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 5 olduğu modelde 0.7021 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 20 olduğu modelde 0.7624 olarak elde edilmiştir. Altıncı aşamada ise maksimum derinliği 12 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 11'de verilmiştir.



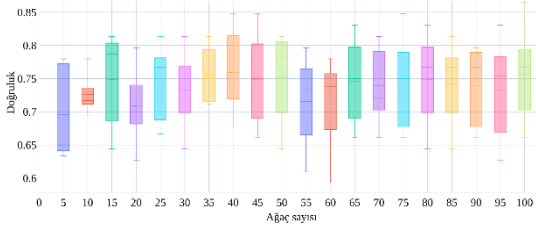
Şekil 11. Ağacın maksimum derinliği 12 iken elde edilen modellerin performans karşılaştırılması.

Şekil 11 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 5 olduğu modelde 0.6588 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 10 olduğu modelde 0.7659 olarak elde edilmiştir. Yedinci aşamada ise maksimum derinliği 14 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilyik grafiği şekil 12'de verilmiştir.



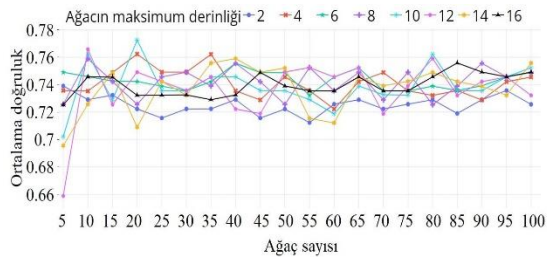
Şekil 12. Ağacın maksimum derinliği 14 iken elde edilen modellerin performans karşılaştırılması.

Şekil 12 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 5 olduğu modelde 0.6955 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 40 olduğu modelde 0.7590 olarak elde edilmiştir. Sekizinci aşamada ise maksimum derinliği 16 olarak belirlenip ağaç sayısı 0 ile 100 arasında değiştirilerek toplam 20 model eğitilmiştir. Eğitilen modellerin kutu bilye grafiği şekil 13'de verilmiştir.

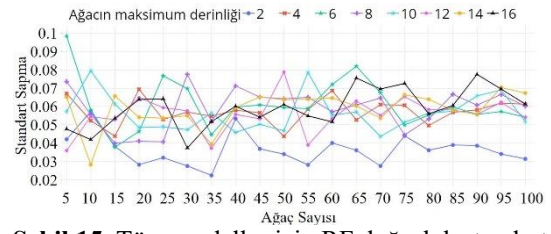


Şekil 13. Ağacın maksimum derinliği 16 iken elde edilen modellerin performans karşılaştırılması.

Şekil 13 incelendiğinde ağaç sayısı 0 ile 100 arasında değiştirildiğinde modelin en düşük ortalama doğruluk ağaç sayısının 5 olduğu modelde 0.7254 olarak elde edilirken en büyük ortalama doğruluk ise ağaç sayısının 80 olduğu modelde 0.7559 olarak elde edilmiştir. Elde edilen tüm sonuçlara ait ortalama sonuç standart sapma grafikleri şekil 14 ve şekil 15'de verilmiştir.

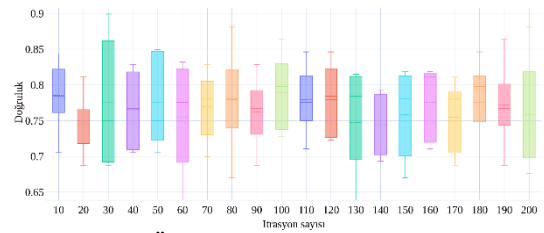


Şekil 14. Tüm modeller için RF ortalama doğruluk grafikleri.



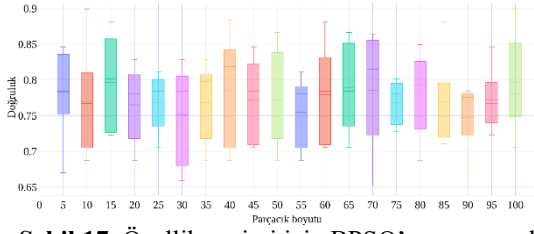
Şekil 15. Tüm modeller için RF doğruluk standart sapma grafikleri.

Şekil 14 incelendiğinde RF modeli için ağacın maksimum derinliği 10 ve ağaç sayısı 10, 20 ve 80 olduğunda doğruluk değerinin 0.76 üzerine üç kez çıktığı görülmektedir. Diğer ağaç değerleri ise en fazla bir kez bu değer üzerine çıkabilmiştir. Bu nedenle ağacın maksimum derinliği 10 olarak belirlenmiştir. Ağacın maksimum derinliği 10 iken en yüksek ortalama doğruluk veren ağaç sayısı 20 olduğu tespit edilmiştir. Şekil 15'de RF modellerin 10 kat çapraz doğrulamasından elde edilen sonuçlar arasında standart sapma değerleri 0.0225 ile 0.0984 arasında değiştiği görülmektedir. Model için belirlenen maksimum derinlik 10 ve ağaç sayısı 20 iken elde edilen standart sapma değeri 0.0486 ve doğruluk değeri ise 0.7624 olarak tespit edilmiştir. Şekil 16'de RF hiper parametreleri kullanılarak gerçekleştirilen özellik seçimine ait iterasyon grafiği verilmiştir.



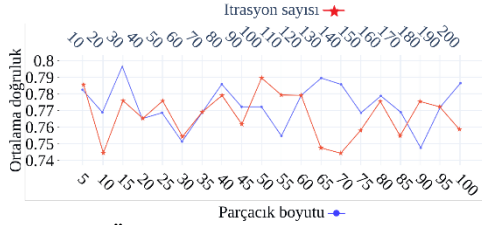
Şekil 16. Özellik seçimi için BPSO iterasyon değişim grafiği.

Şekil 16 incelendiğinde iterasyon sayısı 140 olduğunda en küçük ortalama doğruluk değeri 0.7442 olarak belirlenmiştir. Bu doğruluk değerine ait seçilen özelliklerin indeks değerleri [3, 4, 5, 6, 7, 8, 9, 10]'dir. En yüksek ortalama doğruluk değeri ise iterasyon sayısı 100 iken doğruluk değeri 0.7896 olduğu belirlenmiştir. Bu doğruluk değerine ait seçilen özelliklerin indeks değerleri [1, 2, 4, 6, 7, 8, 9] olarak belirlenmiştir. Ayrıca bu eğitim esnasında BPSO'nun parametresi olan parçacık boyutu değeri 50 olarak alınmıştır. Şekil 17'da belirlenen 100 iterasyon sayısı için; BPSO'nun 0 ile 100 arasında beşer beşer artırılarak değişen parçacık boyutuna göre elde edilen grafik verilmiştir.



Şekil 17. Özellik seçimi için BPSO'nun parçacık boyutu değişimi.

Şekil 17 incelendiğinde parçacık boyutu 90 olduğunda en küçük ortalama doğruluk değeri 0.7475 olarak belirlenmiştir. Bu doğruluk değerine ait seçilen özelliklerin indeks değerleri [4, 6, 7, 9, 10] olarak tespit edilmiştir. En yüksek ortalama doğruluk değeri ise parçacık boyutu 15 iken doğruluk değeri 0.7966 olduğu belirlenmiştir. Bu doğruluk değerine ait seçilen özelliklerin indeks değerleri [0, 2, 3, 4, 6, 7, 9] olarak belirlenmiştir. Şekil 18 ve şekil 19'de özellik seçimi için BPSO parçacık boyutu ve iterasyon için ortalama ve standart sapma değerleri gösterilmiştir.



Şekil 18. Özellik seçimi için BPSO'nun parçacık boyutu ve iterasyon sayısına doğruluk değişimi.



Şekil 19. Özellik seçimi için BPSO'nun parçacık boyutu ve iterasyon sayısına standart sapma değeri grafiği.

Şekil 18 incelendiğinde iterasyon sayısı 140 ve parçacık boyutu 50 olduğunda en küçük ortalama doğruluk değeri 0.7442 olarak belirlenmiştir. En yüksek ortalama doğruluk değeri ise parçacık boyutu 15 ve iterasyon sayısı 100 iken doğruluk değeri 0.7966 olduğu tespit edilmiştir. Şekil 19 incelendiğinde en yüksek doğruluk veren modelin standart sapma değeri 0.0634 olarak belirlenmiştir. Çizelge 4'de Chicco D. ve Jurman G. (2020) yılında aynı veri seti üzerinde yapmış oldukları çalışmadan elde edilen sonuçlar ile gerçekleştirilen çalışmadaki sonuçlar kıyaslanmıştır.

Çizelge 4. Kullanılan veri seti üzerinde gerçekleştirilen çalışmanın sonucu ile diğer çalışmalar karşılaştırılması.

Çalışma	Yöntem	Eğitimde kullanılan özelliklerin indeksi	Ortalama doğruluk
[9]	RF	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.740
	Karar ağacı	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.737
	Gradyan Artırma	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.738
	Lineer regresyon	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.730
	One rule	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.729
	Yapay sinir ağları	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.680
	Naive bayes	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.696
	Radyal SVM	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.690
	Lineer SVM	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.684
	k-en yakın komşu	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.624
	RF	[4,7]	0.754
	Gradyan Artırma	[4,7]	0.750
	Radyal SVM	[4,7]	0.720
Çalışmada	RF + hiper parametrelerin ayarı	[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]	0.76
	RF + hiper parametrelerin ayarı + BPSO	[0, 2, 3, 4, 6, 7, 9]	0.80

Çizelge incelendiğinde gerçekleştirilen çalışmanın sonuçları daha başarılı olduğunu görülmektedir. Birinci yöntemde, RF algoritmasının hiperparametrelerin optimize edilerek modelin doğruluğu yaklaşık olarak %2 oranında artmıştır. İkinci yöntemde ise, BPSO ile global özellik seçimi yapılarak modelin doğruluğu yaklaşık olarak %4 oranında artmıştır.

4. SONUÇ

AI yöntemleri birçok farklı alanda olduğu gibi sağlık alanında da sıklıkla kullanılmakta ve birçok hastalığın erken teşhis edilmesinde önemli rol oynamaktadır. İnsan hayatı için risk oluşturan hastalıkların erken ve doğru bir şekilde teşhis edilmesi bu hastalıklardan kaynaklanan ölüm risklerinde azaltılmasında oldukça önemlidir. Çalışmada açık erişimli internet sitesinden (UCI) elde edilen UCI Heart failure clinical records ait toplam 299 veri kullanılmıştır. Elde edilen veri seti üzerinde RF ve BPSO-RF algoritmaları için farklı hiper parametreler seçilerek farklı modeller elde edilmiştir. Elde edilen sonuçlar aşağıda maddeler halinde verilmiştir.

- İlk olarak RF modelinde farklı ağaç sayısı ve ağacın maksimum derinlik değerleri değiştirilerek farklı sonuçlar elde edilmiştir. Elde edilen sonuçlar içerisinde en yüksek doğruluk değerini veren modelin ağaç sayısı 20, ağacın maksimum derinliği 10 ve ortalama doğruluk değeri 0.7624 ve standart sapma değeri ise 0.0486 olarak belirlenmiştir.
- Çalışmanın ikinci aşamasında RF modelinden elde edilen en uygun parametrelere kullanılarak BPSO-RF modeli oluşturulmuştur. Özellik seçimi için BPSO algoritmasının parçacık boyutu ve iterasyon sayısı değiştirilerek en uygun model elde edilmiştir. En yüksek doğruluk veren modelde BPSO için parçacık boyutu 15, iterasyon sayısı 100 olarak tespit edilmiştir. BPSO algoritmasından [0, 2, 3, 4, 6, 7, 9] indeksli özelliklerin seçilmiştir. Eğitim için RF ağaç sayısı 20, ağacın maksimum derinliği 10 ve ortalama doğruluk değeri 0.7966 ve standart sapma değeri ise 0.0634 olarak belirlenmiştir.

Gerçekleştirilen çalışma ile KY hastalığının tespitinde hangi özneliklerin KY hastalığı için

önemli olduğu Binary parçacık sürü optimizasyonu yöntemi ile belirlenmiştir. Ayrıca belirlenen özneliklere göre KY hastalığını %80 doğruluk oranında tahminleyen Random forest modeli geliştirilerek akademik literatüre katkıda bulunulmuştur.

İlerleyen çalışmalarda farklı optimizasyon yöntemleri kullanılarak modelin doğruluğunun artırılması planlanmaktadır. Ayrıca veri seti artırımı yapılarak modelin performansı geliştirilmesi düşünülmektedir.

TEŞEKKÜR

Çalışmada 'Heart failure clinical records Data Set' veri setini açık kaynak erişimli internet sitesine (<https://archive.ics.uci.edu/>) aktaran Davide Chicco ve Giuseppe Jurman teşekkürlerimizi sunarız.

KAYNAKLAR

1. Kanbay, M., Yerlikaya, A., Sag, A. A., Ortiz, A., Kuwabara, M., Covic, A., Wiecek, A., Stenvinkel, P., Afsar B., "A journey from microenvironment to macroenvironment: the role of metaflammation and epigenetic changes in cardiorenal disease", *Clinical Kidney Journal*, Vol. 12, Issue 6 Pages 861-870, 2019.
2. Görgülü, S., "Acil Servis'e Kalp Yetmezliği Semptomları İle Başvuran Hastalara Uygulanan Diüretik Tedavi Neticesinde Karşılaşılabilecek Akut Böbrek Yetmezliği Tablosunu Önceden Belirlemede", *Uzmanlık Tezi, Hacettepe Üniversitesi, Ankara, 2020.*
3. Ekin, H., & Polat, Ü., "Kalp yetmezliği olan hastaların hastalık algısı ve bakımlarına ilişkin görüşlerinin değerlendirilmesi", *Genel Tıp Dergisi*, Cilt 28, Sayı 1, Sayfa14-22, 2018.
4. Çavuşoğlu, Y., Altay, H., Ekmekçi, A., "Kronik kalp yetersizliği tedavisinde pratik yaklaşımlar: Güncel uygulamada sık sorulanlar, gözden kaçanlar, tartışmalı durumlar", *Anatol J Cardiol*, Cilt 15, Sayfa 1-65, 2015.
5. Ponikowski, P., Voors, A.A., Anker, S.D., et al. "ESC Guidelines For The Diagnosis And Treatment of acute and chronic heart failure", *European heart journal*, Vol. 39 Issue 10, Pages 2129, 2018.
6. Değertekin, M., Erol, Ç., Ergene, O., "Türkiye'deki kalpyetmezliği prevalansı ve öngördüğüleri: HAPPY çalışması", *Türk Kardiyol DernArs*, Cilt 40, Sayfa 298-308, 2012.
7. Faris, R.F., Flather, M., Purcell H., Poole-Wilson P. A., Coats A.J., "Diuretics for heart failure",

Cochrane Database of Systematic Reviews, 2012.
<https://doi.org/10.1002/14651858.CD003838.pub3>

8. National Heart Lung and Blood Institute (NHLBI),“Heart failure”,<https://www.nhlbi.nih.gov/health-topics/heart-failure>. Ağustos 20, 2020.

9. Chicco, D., Jurman, G.,“Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone”, BMC medical informatics and decision making, Vol. 20, Issue 1, Pages1-16, 2020.

10. Meng, F., Zhang, Z., Hou, X., Qian, Z., Wang, Y., Chen, Y., Wang, Y., Zhou, Y., Chen, Z., Zhang, X., Yang, J., Zhang, J., Guo, J., Li, K., Chen, L., Zhuang, R., Jiang, H., Zhou, W., Tang, S., Wei, Y., Zou, J.,“Machine learning for prediction of sudden cardiac death in heart failure patients with low left ventricular ejection fraction: study protocol for a retrospective multicentre registry in China”, Br Med J (BMJ) Open, Vol. 9, Issue 5, Pages 1-9, 2019.

11. Nauta, J.F., Jin, X., Hummel, Y.M., Voors, A.A.,“Markers of left ventricular systolic dysfunction when left ventricular ejection fraction is normal”, Eur J Heart Fail, Vol. 20, Issue 12, Pages, 1636-1638, 2018.

12. Nanayakkara, S., Kaye, DM.,“Targets for heart failure with preserved ejection fraction”, Clin Pharmacol Ther. Clinical Pharmacology & Therapeutics, Vol. 102, Issue 2, Pages 228–237, 2017.

13. McMurray, J. J. V., Adamopoulos, S., Anker, S. D., Auricchio, A., Böhm, M., Dickstein, K., Falk, V., Filippatos, G., Fonseca, C., Gomez-Sanchez, M. A. N., Jaarsma, T., Kober, L., Lip, G. Y. H., Maggioni, A. P. I., Parkhomenko, A., Pieske, B. M., Popescu, B. A., Ronnevik, K., Rutten, F. H., “Akut ve Kronik Kalp Yetersizliği Tani ve Tedavisine Yönelik 2012 ESC Kilavuzu”, Turk Kardiyoloji Dernegi Arsivi, Cilt 40, Sayfa 77-137, 2012.

14. Alkan, B., “Kronik Kalp Yetmezliği Hastalarında Semptom Limitli Merdiven Çıkma Testinin Etkinliğinin Değerlendirilmesi”, Doktora Tezi, Dokuz Eylül Üniversitesi, İzmir, 2018.

15. Paksoy, V. M., “Elektronik sağlık kayıtlarının güvenlik ve mahremiyet uygulamalarının özel hastanelerde değerlendirilmesi: Kayseri ili örneği”, Doktora Tezi, Marmara Üniversitesi, İstanbul, 2019.

16. Hartley, C., Jones, E., “EHR Implementation: A Step-by-Step Guide for the Medical Practice”,

American Medical Association, 2th Edition, Chicago, 2012.

17. Aldosari, B., “Causes of EHR Projects Stalling or Failing: A Study of EHR Projects in Saudi Arabia”, Computers in Biology and Medicine, Vol. 91, Pages 372- 381, 2017.

18. Büyükgöze, S., Dereli, E., “Dijital Sağlık Uygulamalarında Yapay Zeka”, VI. Uluslararası Bilimsel ve Mesleki Çalışmalar Kongresi-Fen ve Sağlık, 07-10, Online, 2020.

19. Cheng, J. C., Chen, W., Chen, K., Wang, Q., “Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms” Automation in Construction, Vol. 112, Pages 1-21, 2020.

20. Kaynar, O., Tuna, M. F., Görmez, Y., Devenci, M. A., “Makine öğrenmesi yöntemleriyle müşteri kaybı analizi”, Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi, Cilt 18, Sayı 1, Sayfa 1-14, 2017.

21. Filiz, E., Karaboğa, H. A., Akoğul, S., “Bıst-50 Endeksi Değişim Değerlerinin Sınıflandırılmasında Makine Öğrenmesi Yöntemleri Ve Yapay Sinir Ağları Kullanımı”, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 26, Sayı 1, Sayı 231-241, 2017.

22. Cervante, L., Xue, B., Zhang, M., & Shang, L., “Binary particle swarm optimisation for feature selection: A filter based approach” In 2012 IEEE Congress on Evolutionary Computation, Sayfa 1-8, Brisbane, 2012.

23. Whitney, A., “A direct method of nonparametric measurement selection,” IEEE Transactions on Computers, vol. C-20, Issue 9, Pages 1100–1103, 1971.

24. Marill, T., and Green, D., “On the effectiveness of receptors in recognition systems,” IEEE Transactions on Information Theory, Vol. 9, Sayı 1, Pages 11–17, 1963.

25. Tao, Z., Huiling, L., Wenwen, W., Xia, Y. “GA-SVM based feature selection and parameter optimization in hospitalization expense modeling”, Applied Soft Computing, Vol. 75, Pages 323-332, 2019.

26. Kushwaha, N., & Pant, M., “Link based BPSO for feature selection in big data text clustering”, Future Generation Computer Systems, Vol. 82, Pages 190-199, 2018.

27. Shunmugapriya, P., Kanmani, S., “A hybrid algorithm using ant and bee colony optimization for

feature selection and classification (AC-ABC Hybrid)”, *Swarm and Evolutionary Computation*, Vol. 36, Pages 27-36, 2017.

28. Manoj, R. J., Praveena, M. A., Vijayakumar, K., “An ACO-ANN based feature selection algorithm for big data”, *Cluster Computing*, Vol.22, Issue 2, Pages 3953-3960, 2019.

29. Ahmad, T., Munir, A., Bhatti, S.H., Aftab, M., Raza, M.A., “Survival analysis of heart failure patients: a case study”, *PLoS ONE*, Vol. 12, Issue 7, Pages 1-8, 2017.

30. UCI Heart failure clinical records Data Set. - <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records>. Ağustos 2020.

31. Breiman, L., “Rastgele Ormanlar”, *Machine Learning*, Vol. 45, Pages, 5-32, 2001.

32. Mursalin, M., Zhang, Y., Chen, Y., & Chawla, N. V., “Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier”, *Neurocomputing*, Vol. 241, Pages 204-214, 2017.

33. Akar, Ö., & Güngör, O., “Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması”, *Jeodezi ve Jeoinformasyon Dergisi*, Cilt 1, Issue 2, 139-146, 2012.

34. Korkmaz, D., Çelik, H. E., & Kapar, M., “Sınıflandırma ve Regresyon Ağaçları ile Rastgele Orman Algoritması Kullanarak Botnet Tespiti: Van Yüzüncü Yıl Üniversitesi Örneği”, *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Vol. 23, Issue 3, Pages 297-307, 2018.

35. Chen, J., Li, K., Tang, Z., Bilal, K., Yu, S., Weng, C., & Li, K., “A parallel random forest algorithm for big data in a spark cloud computing environment”, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 28, Issue 4, Pages 919-933, 2016.

36. Gulia, A., Vohra, R., Rani, P., “Liver patient classification using intelligent techniques”, *International Journal of Computer Science and Information Technologies*, Vol. 5, Issue 4, Pages, 5110-5115, 2014.

37. Sevli, O., “Göğüs Kanseri Teşhisinde Farklı Makine Öğrenmesi Tekniklerinin Performans

Karşılaştırması”, *Avrupa Bilim ve Teknoloji Dergisi*, Sayı 16, Sayfa 176-185, 2019.

38. Doğan, M., Özkan, İ. A., “Sürü zekası yöntemleriyle aşırı öğrenme makinesi’nin öğrenme parametreleri optimizasyonu”, *Selçuk-teknik dergisi*, Cilt 19, Sayı 2, Sayfa 35-50, 2020.

39. Kaya, S., Karaçizmeli, İ. H., Aydılek, İ. B., Tenekeci, M. E., Gümüşçü, A., “Akış tipi çizelgeme problemlerinin hibrit ateşböceği ve parçacık sürü optimizasyonu algoritmasıyla çözümünde başlangıç popülasyonlarının etkileri”, *Pamukkale University Journal of Engineering Science*, Cilt 26, Sayı 1, Sayfa 104-149, 2020.

40. Güney, H. M., “Hibrit parçacık sürü optimizasyonu algoritması ile hidrolojik model kalibrasyonu”, *Yüksek Lisans Tezi*, Balıkesir Üniversitesi, Balıkesir, 2020.

41. Mirjalili, S., “Particle swarm optimisation”, In *Evolutionary Algorithms and Neural Networks*, Pages 15-31, Springer, Cham, 2018.

42. Cui, Z., Zhang, J., Wu, D., Cai, X., Wang, H., Zhang, W., & Chen, J., “Hybrid many-objective particle swarm optimization algorithm for green coal production problem”, *Information Sciences*, Vol. 518, Pages 256-271, 2020.

43. Liang, J., Ge, S., Qu, B., Yu, K., Liu, F., Yang, H., ... & Li, Z., “Classified perturbation mutation based particle swarm optimization algorithm for parameters extraction of photovoltaic models”, *Energy Conversion and Management*, Vol. 203, 112138, 1-20, 2020.

44. Zhang, P., Yin, Z. Y., Jin, Y. F., & Chan, T. H., “A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest”, *Engineering Geology*, Vol. 265, 105328, Pages 1-12, 2020.

45. Elbaz, K., Shen, S. L., Sun, W. J., Yin, Z. Y., & Zhou, A., “Prediction model of shield performance during tunneling via incorporating improved Particle Swarm Optimization into ANFIS”, *IEEE Access*, Vol. 8, Pages 39659-39671, 2020.