



T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



YAPAY ARI KOLONİSİ ALGORİTMASI İLE
ÖZELLİK SEÇİMİ

Zehra KIRAN

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Ocak-2023
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Zehra KIRAN tarafından hazırlanan ‘‘Yapay Arı Kolonisi Algoritması ile zellik Seimi’’ adlı tez alıřması 04/01/2023 tarihinde ařađıdaki jri tarafından oy birliđi ile Konya Teknik niversitesi Lisansst Eđitim Enstits Bilgisayar Mhendisliđi Anabilim Dalı’nda Yksek Lisans Tezi olarak kabul edilmiřtir.

Jri yeleri

Başkan

Do.Dr. Mehmet Akif řAHMAN

Danışman

Do.Dr. Ahmet BABALIK

ye

Do.Dr. İsmail BABAOĐLU

İmza

.....

.....

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Saadettin Erhan KESEN
Enstit Mdr

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Zehra KIRAN

04.01.2023

ÖZET

YÜKSEK LİSANS TEZİ

YAPAY ARI KOLONİSİ ALGORİTMASI İLE ÖZELLİK SEÇİMİ

Zehra KIRAN

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç.Dr. Ahmet BABALIK

2023, 48 Sayfa

Jüri

Doç.Dr. Ahmet BABALIK
Doç.Dr. Mehmet Akif ŞAHMAN
Doç.Dr. İsmail BABAĞLU

Veri madenciliği ve makine öğrenmesi alanında özellik seçimi boyut indirgemek amacıyla yapılmaktadır. Bu sayede makine öğrenmesi yöntemleri ile işlenen veriler üzerinde hem işlem süresi azalmakta hem de gereksiz özelliklerin silinmesiyle makine öğrenmesi yöntemlerinin sınıflandırma başarıları artmaktadır. Özellik seçiminde veri kümesindeki özelliklerin sayısına bağlı olarak seçilebilecek özellik alt kümelerinin sayısı üstel şekilde artmakta ve bundan dolayı olası bütün çözümlerin denenerek optimum alt kümenin bulunması uzun süreler alabilmektedir. Tez kapsamında öncelikle özellik seçme problemi ikili optimizasyon problemi olarak ele alınmış ve her özelliğin seçilip seçilmediği $\{0,1\}$ kümesi ile temsil edilmiştir. Bir özellik için 0 değeri özelliğin seçilmediğini ve 1 değeri ise özelliğin seçildiğini temsil etmektedir. İkili optimizasyon ile modellenen problemdeki çok zaman tüketebilen bu sürecin üstesinden gelmek amacıyla literatürde özellik seçimi için evrimsel hesaplama ve/veya sürü zekâsı algoritmalarına sıklıkla başvurulmaktadır. Bu algoritmalar ikili, ayrık, sürekli vb. herhangi bir optimizasyon probleminde optimum çözümü garanti etmese de makul sürelerde optimum veya yakın optimum çözümleri elde edebilmektedir. Bu tez çalışmasında sürü zekâsı alanının önde gelen üyelerinden biri olan yapay arı kolonisi algoritmasının ikili versiyonları üzerinde durulmuştur. 8 farklı ikili ABC algoritması 5 farklı veri setinde özellik seçme amacıyla kullanılmıştır. Bu algoritmalarda üretilen çözümlerin değerlendirilmesi amacıyla amaç fonksiyon olarak aşırı öğrenme makinesi ile eğitilmiş yapay sinir ağı kullanılmıştır. Bir çözümün değerlendirilmesi için çözümdeki özellikler ile veri kümesi indirgenmekte ve elde edilen indirgenmiş veri kümesi eğitim ve test olarak ikiye ayrılmaktadır. Eğitim veri seti üzerinde aşırı öğrenme makinesi ile yapay sinir ağı eğitilmekte ve eğitilmiş ağı ile test veri seti sınıflandırılmaktadır. Sınıflandırma hatası ise bu amaç fonksiyondan döndürülmekte ve çözümün kıyaslanması amacıyla kullanılmaktadır. Yöntemlerin veri kümelerine uygulanması sonucunda indirgenmiş veri kümeleri ile yapılan sınıflandırma işleminde hem eğitim başarısı hem de test başarılarının indirgenmemiş veri kümesine göre daha yüksek başarı elde ettiği görülmüştür. Ayrıca ikili ABC algoritmaların karşılaştırılmasında ise genetik operatörler kullanılarak ikili optimizasyon amacıyla uygulanan ABC algoritmasının diğer algoritmalarından daha başarılı sonuçlar ürettiği görülmüştür.

Anahtar Kelimeler: Aşırı öğrenme makinesi, ikili optimizasyon, özellik seçimi, sürü zekâsı, veri madenciliği, yapay arı kolonisi algoritması

ABSTRACT

MS THESIS

FEATURE SELECTION USING ARTIFICIAL BEE COLONY ALGORITHM

Zehra KIRAN

**Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering**

Advisor: Assoc.Prof.Dr. Ahmet BABALIK

2023, 48 Pages

Jury

**Assoc.Prof.Dr. Ahmet BABALIK
Assoc.Prof.Dr. Mehmet Akif ŞAHMAN
Assoc.Prof.Dr. İsmail BABAÖĞLU**

The feature selection is performed for the dimension reduction of the data in the fields of data mining and machine learning. Not only processing time of the machine learning methods is decreased, but also their classification success is improved on the data by removing irrelevant features from the dataset. By depending on the number of the features on the dataset, the number of feature subsets exponentially increases, and finding the optimum subset by evaluating all the possible solutions can consume much more time. In this thesis, the feature selection problem is considered as a binary optimization problem, and whether each feature is selected or not is represented with the set of $\{0,1\}$. 0 and 1 represent the unselected and selected feature, respectively. In order to cope with this time-consuming process of the problem modelled as a binary optimization problem, the evolutionary computing or swarm intelligence algorithms are often used for feature selection in the literature. These algorithms do not guarantee the optimum solution for the problem, but they guarantee to achieve the optimum or near-optimum solutions in a reasonable time. The artificial bee colony algorithm, which is a prominent member of swarm intelligence algorithms, is studied in this thesis. Eight different binary artificial bee colony algorithms are used for feature selection on five datasets. In order to evaluate the solutions produced by these algorithms, the artificial neural network trained by extreme learning machine is used. The dataset is reduced with the feature subset represented in a solution, and the resulting reduced dataset is divided into training and test sets for evaluating this solution. The artificial neural network is trained with the extreme learning machine on the training dataset and the test dataset is classified with the trained network. Classification error is returned from this objective function and used to compare the solutions. In order to evaluate a possible solution, the data are reduced by the feature subset, the extreme learning machine is trained on this reduced dataset, and the reduced dataset is classified with this trained network. Classification error is returned by this objective function, and it is used for comparing the solutions. As a result of the application of the methods to the datasets, it is seen that both the training and test successes in the classification process with the reduced datasets achieved higher success than the non-reduced dataset. Moreover, when the binary versions of ABC are compared, the binary ABC that modified with the genetic operators produced more successful results than the compared binary ABC variants.

Keywords: Artificial bee colony, binary optimization, data mining, extreme learning machine, feature selection, swarm intelligence

ÖNSÖZ

Yüksek lisans öğrenimi sürecimde yapay arı kolonisi algoritması ve aşırı öğrenme makinesi üzerine çalışmayı isteyerek hem sürü zekâsı hem de makine öğrenmesi üzerine bilgi birikimimi artırmaya çalıştım. Bu tez çalışmasında farklı veri setlerine özellik seçimi amacıyla yapay arı kolonisi algoritmasının ikili versiyonları uygulanmıştır. Çözümlerin değerlendirilmesi içinse aşırı öğrenme makinesi ile eğitilmiş yapay sinir ağı kullanılmıştır. Öncelikle yöntemlerin performansları bu amaçla geliştirilmiş kodlar ile değerlendirilmiş ve test edilmiştir. Ayrıca farklı disiplinlerden kullanıcıların ve araştırmacıların kullanımına sunulabilmesi amacıyla örnek bir uygulama geliştirilmiştir. Örnek uygulamanın geliştirilme amacı bilişim bilimleri üzerine eğitim almamış kullanıcılar tarafından da veri madenciliği, özellik seçimi ve makine öğrenmesi yöntemlerinin kullanılabilmesidir. Tez için gerek literatür özetinin hazırlanması gerekse yöntemlerin özellik seçme için kodlanması ve uyarlanması ve örnek uygulamanın geliştirilmesi oldukça uzun ve yorucu bir süreçti. Ayrıca tezin yazımı, kontrolü ve tekrar düzenlenmesi için de oldukça yoğun bir emek harcandı. Bu süreçlerin tümünde yol gösteren ve yardımcı olan danışman hocam Doç. Dr. Ahmet BABALIK Bey'e özellikle teşekkür ederim. Özellik seçme üzerine literatür özetinde yardımcı olan Arş.Gör. Aybüke BABADAĞ'a teşekkürlerimi sunarım. Ayrıca kodlama konusunda bana yardımcı olan hayat arkadaşım Prof. Dr. Mustafa Servet KIRAN'a ve yorulduğum zamanlarda tekrar çalışma isteği kazanmamı sağlayan oğullarım Eren ve Mehmet'e sevgilerimi sunarım.

Zehra KIRAN

KONYA-2023

İÇİNDEKİLER

ÖZET	iv
ABSTRACT	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	viii
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	3
2.1. Özellik Seçimi Üzerine Kaynak Araştırması	3
2.2. İkili ABC Algoritması Üzerine Kaynak Araştırması	8
3. MATERYAL VE YÖNTEM	13
3.1. Temel ABC Algoritması	13
3.2. Mod Tabanlı İkili ABC Algoritması	16
3.3. Açık Modülasyonu İkili ABC Algoritması	16
3.4. Xor Tabanlı İkili ABC Algoritması	17
3.5. Bitisel Operatör Tabanlı İkili ABC Algoritması	18
3.6. Farklılık Tabanlı İkili ABC Algoritması	19
3.7. Genetik Operatör Tabanlı İkili ABC Algoritması	20
3.8. Sigmoid İkili ABC Algoritması	21
3.9. Stigmerji-Xor Tabanlı İkili ABC Algoritması	22
3.10. Aşırı Öğrenme Makinesi	23
3.11. Deneysel Çalışmada Kullanılan Veri Setleri	24
4. DENEYSEL SONUÇLAR	25
5. GELİŞTİRİLEN ÖRNEK UYGULAMA	33
6. SONUÇLAR VE ÖNERİLER	38
6.1 Sonuçlar	38
6.2. Öneriler	39
KAYNAKLAR	40

SİMGELER VE KISALTMALAR

Simgeler

\emptyset_i^j	i. bireyin j. boyutu için [-1,1] aralığında üretilen rastgele sayı
C_{01}^j	j. boyutta 0'dan 1'e geçiş sayısı
C_{10}^j	j. boyutta 1'den 0'a geçiş sayısı
H^+	H dikdörtgen matrisin (gizli katman çıkış) Moore-Penrose kaba tersi
P_{01}^j	j. boyutun 0'dan 1'e geçiş ihtimali
P_{10}^j	j.boyutun 1'den 0'a geçiş ihtimali
V_i^j	i. aday çözümün j. boyutu
X_i^j	i. bireyin j. boyutu
X_j^{max}	j. boyutun üst sınırı
X_j^{min}	j. boyutun alt sınırı
X_k^j	k. bireyin j. boyutu
f_i	i. bireyin amaç fonksiyon değeri
fit_i	i. bireyin uygunluk değeri
p_i	i. bireyin gözcü arılar tarafından seçilme ihtimali
r_i^j	i. bireyin j. boyutu için üretilen rastgele sayı
y_i	Sürekli çözüm vektörü
z_i	İkili çözüm vektörü
\wedge	Lojik "ve" operatörü
\vee	Lojik "veya" operatörü
\oplus	Lojik "özel veya" operatörü
mod2	2'ye göre mod alma
D	Problemin boyutu (karar değişkeni sayısı)
W	Yapay sinir ağında gizli katman ile çıkış katmanı arası ağırlık matrisi
X	Popülasyon matrisi
Y	Verilerin etiket sütun vektörü
abs	mutlak değer fonksiyonu
$round$	En yakın tam sayıya yuvarlama
$sig()$	Sigmoid fonksiyonu
φ	%50 ihtimalle tersleme

Kısaltmalar

AABC:	Açı modülasyonlu ikili ABC algoritması
ABC:	Yapay arı kolonisi algoritması
ACO:	Karınca kolonisi optimizasyonu
ALO:	Karınca aslanı optimizasyon algoritması
BABC:	Bitsel operatör tabanlı ikili ABC algoritması
BBA:	İkili yarasa algoritması
BT:	Bilgisayarlı tomografi
BT-SFS:	Geri izleme sıralı özellik seçimi
CFS:	Korelasyon tabanlı özellik seçimi
DT:	Karar ağacı
DE:	Farksal gelişim algoritması
EEFS:	Birleşik gömülü özellik seçimi
ELM:	Aşırı öğrenme makinesi
FABC:	Farklılık (dissimilarity) tabanlı ikili ABC algoritması

FOA: Orman optimizasyon algoritması
GA: Genetik algoritma
GABC: Genetik operatörler tabanlı ikili ABC algoritması
HPSO-LS: Hibrit PSO lokal arama algoritması
ImRMR: Gelişmiş minimum fazlalık maksimum ilişki
iBPSO: İyileştirilmiş ikili parçacık sürü optimizasyon algoritması
kNN: k-en yakın komşuluk algoritması
MABC: Mod tabanlı ikili ABC algoritması
MECY-FS: Çok amaçlı evrimsel algoritma – özellik seçimi
MLP: Çok katmanlı algılayıcı
MNB: Çok terimli naif bayes
MRDC: Çok değişkenli görelî ayırım kriteri
NP: Deterministik olmayan polinomsal
PSO: Parçacık sürü optimizasyon algoritması
RDC: Görelî ayırım kriteri
SABC: Sigmoid tabanlı ikili ABC algoritması
SFLA: Kurbağa sıçrama algoritması
SPEA2: Güçlü pareto evrimsel algoritma
SVM: Destek vektör makineleri
SXABC: Stigmerji ve özel veya tabanlı ikili ABC algoritması
UCI: Kaliforniya Üniversitesi, Irvine
UPFS: Danışmansız olasılıksal özellik seçimi
WOA: Balina optimizasyon algoritması
XABC: Özel veya tabanlı ikili ABC algoritması
XOR: Özel veya lojik operatörü
YSA: Yapay sinir ağı

1. GİRİŞ

Son yıllarda dijital verinin artması ile verilerin işlenmesi konusunda da çeşitli ilerlemeler kaydedilmektedir. Verilerin depolanması, istendiğinde elde edilmesi, görselleştirilmesi gibi konuların yanı sıra bu verilerin makine öğrenmesi yöntemleriyle işlenmesi ve kararlar üretilmesi amacıyla da teknolojide ve literatürde çalışmalar sürdürülmektedir. Bu çalışmalar genellikle iki kategoride ele alınmaktadır. Verilerin önceden referans sınıflara ayrılması ile eğitim yapılması ve yeni elde edilen verilerin bu sınıflardan hangisine dahil olduğunun belirlenmesi veya eldeki verilerin belirli ölçekler kullanılarak kümelenmesi olarak karşımıza çıkmaktadır. Sınıflandırma ve kümeleme şeklinde kategorize edilen bu verilerin anlamlandırılması veya kıymetlendirilmesi sürecinde önemli bir etken verinin dikey (hacimsel) ve yatay ölçekte miktarıdır. Verinin dikey ölçekte artması genellikle farklı örneklerin makine öğrenmesi yöntemlerinin eğitim aşamasında önemli bir etken iken yatay ölçekte artması farklı özneliklerin de eğitimde kullanılması anlamına gelebilmektedir. Ayrıca makine öğrenmesi yöntemlerinin eğitim ve test sürecindeki performansı verinin miktarına da bağlıdır. Az sayıda veri ile makine öğrenmesi yönteminin eğitimi yapılabilir olmasına rağmen farklı örnekler ile karşılaşıldığında yöntemin test başarısı düşebilmektedir. Fazla sayıda veri ile yapılan eğitimlerde de eğitim süresi artmakta ve çevrimiçi makine öğrenmesi yöntemlerine uygun olmamaktadır. Bu noktada eldeki verinin yapay zekâ veya makine öğrenmesi yöntemleri ile işlenmeden önce ön işlemden geçirilmesi gerekmektedir. Tekrar eden verilerin çıkarılması, verilerin ayrıklaştırılması, özellik azaltma ve özellikle seçme veri ön işleme adımlarından bazılarıdır. Bu veri ön işleme hem zamansal açıdan avantaj sağlayan hem de yöntemlerin eğitim/test başarısını artıran önemli bir adımdır. Bu tez çalışmasında makine öğrenmesi yöntemlerinin performansının artırılması amacıyla özellik seçme üzerinde durulmuştur. Özelliklerin seçilmesi ikili optimizasyon problemi olarak ele alınmış ve sürü zekâsına dayanan ikili optimizasyon algoritmaları ile özelliklerin seçimi yapılmış ve seçilen özelliklerin etkinliğinin ölçülebilmesi amacıyla makine öğrenmesi yöntemlerinden bir tanesi ile yapay sinir ağı eğitimi ve testi yapılmıştır. Sürü zekâsı algoritması olarak basit, açıklanabilir, sürekli ve ikili optimizasyonda yüksek başarı göstermesi sebebiyle yapay arı kolonisi algoritması (artificial bee colony – ABC), (Karaboga, 2005) seçilmiştir. Yapay sinir ağının eğitimi amacıyla da aşırı öğrenme makinesi (extreme learning machine-ELM) (Huang ve ark., 2004) algoritması kullanılmıştır. Kısacası özellik seçme için ikili yapay arı kolonisi algoritması, seçilen

özelliklerin değerlendirilmesi amacıyla aşırı öğrenme makinesi ile eğitilmiş yapay sinir ağı kullanılmıştır.

Özellik seçme problemi ikili bir optimizasyon problemi olarak modellenebilir ve her bir özelliğin seçilip seçilmediği $\{0,1\}$ değerleri ile temsil edilebilir. 1 özelliğin seçildiği, 0 ise seçilmediği durumu temsil etmektedir. Bir veri setinde özellik sayısının n varsayıldığında seçilebilecek özellik sayısı 2^n tane olarak karşımıza çıkmaktadır. En az bir özelliğin bulunduğu durumda ise 2^n-1 adet özellik kümesi içerisinde eğitim ve test hatasını minimize eden özellik kümesinin bulunması süreci bir ikili optimizasyon problemidir. Bu problem literatürde minimum özellik altkümüresi seçim (minimum feature subset selection) problemi olarak tanımlanır ve olası çözüm sayısı özellik sayısına bağlı olarak üstel şekilde arttığından dolayı NP-zor bir problemdir ve NP-zorluğu (Chen ve ark., 1997) çalışmasında ispatlanmıştır. Problemin karakteristiğinden dolayı özellik sayısının yüksek boyutlara ulaşması neticesinde makul zamanda bir çözüm henüz bulunmadığından dolayı evrimsel hesaplama veya sürü zekâsı algoritmaları bu problemin çözümü için literatürde sıklıkla uygulanmıştır. Bu tez kapsamında da sürü zekâsı algoritmaları içerisinde basit, açıklanabilir ve sürekli ve ikili optimizasyonda yüksek performans sergileyen yapay arı kolonisi algoritması problemin çözümü için uygulanmıştır (Karaboga ve Akay, 2009) (Kiran, 2021) . Uygulanan algoritma ile elde edilen özellik altkümelerinin uygunluğunun ölçülmesi amacıyla özelliği indirgenmiş veri setinin sınıflandırılması için aşırı öğrenme makinesi kullanılmıştır. Aşırı öğrenme makinesi genelleştirme kabiliyeti ve hızlı eğitim süreci (çevrimsiz - iterasyonsuz) nedeniyle literatürde yapay sinir ağı eğitiminde son zamanlarda yoğun olarak kullanılmış bir eğitim algoritmasıdır (Ding ve ark., 2015).

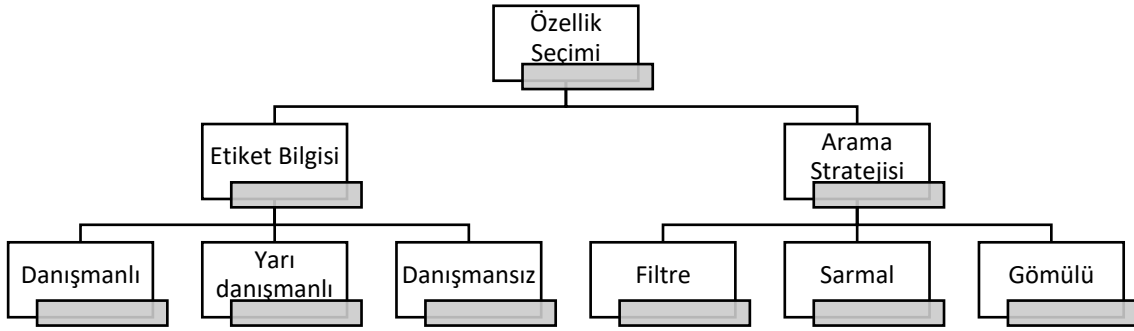
Tezin organizasyonu şu şekildedir: 1. Bölümde teze giriş yapılmış ve 2. Bölümde ikili yapay arı kolonisi algoritması ve sürü zekâsı ile özellik seçimi üzerine literatür araştırması verilmiştir. 3. Bölümde kullanılan yöntemler ve özellik seçimi için kullanılacak veri setleri kısaca açıklanmıştır. 4. Bölümde deneysel çalışmalar yapılmış ve farklı ikili yapay arı kolonisi algoritmaları ile yapılan özellik seçme uygulamalarının eğitim ve test başarıları sunularak karşılaştırmalı bir analiz yapılmıştır. 5. Bölümde geliştirilen uygulamanın kullanımı sunulmuş ve 6. Bölümde sonuçlar ve öneriler verilmiştir.

2. KAYNAK ARAŞTIRMASI

Kaynak araştırması bölümü, özellik seçimi problemi ve bu problemi çözmek için uyguladığımız yöntem olan ikili ABC algoritmasının özellik seçimi problemine uygulanmasını içeren çalışmalar olmak üzere iki bölüme ayrılmıştır. Birinci bölümde özellik seçimi ve özellik seçiminde kullanılan yöntemler üzerinde durulmuştur. Bu yöntemlerin de ABC algoritmasının ait olduğu sürü zekâsı algoritmaları olmasına dikkat edilmiştir. İkinci bölümde ise özellik seçimi problemi çözmek amacıyla geliştirilmiş ve/veya uygulanmış ikili ABC algoritmaları hakkında bir kaynak araştırması verilmiştir. İkinci bölüm ABC algoritmasının geliştirildiği yıl olan 2005 yılında başlayarak 2021 yılına kadar yapılmış ve Google Scholar’da “artificial bee colony feature selection” arama terimi ile yapılmıştır. 2005 yılında 2021 yılına kadar toplam 121 uluslararası araştırma makalesi ve/veya konferans bildirisi ve 5 literatür özeti gözden geçirilmiştir. İkinci bölümün detaylandırılmasındaki amaç ise bu tez çalışmasında özellik seçimi için uygulanan yöntemin ve amaç fonksiyon olarak kullanılan ELM’nin önceki çalışmalardan farklı olarak yenilik içerdiğini vurgulamak için yapılmıştır.

2.1. Özellik Seçimi Üzerine Kaynak Araştırması

Özellik seçimi, bir boyut indirgeme yöntemidir. Veri kümesini etkili bir şekilde temsil edemeyen özelliklerin, özellik kümesinden çıkarılmasıyla boyutun azaltılması amaçlanmaktadır. Özellik seçimi, veri kümesinin boyutunun azalması sayesinde hesaplama maliyetinin düşürülmesini ve veri kümesini en iyi ifade eden minimum sayıda özelliğin kullanılmasıyla öğrenme başarısının artmasını sağlayabilmektedir (Miao ve Niu, 2016; Rostami ve ark., 2021). Özellik seçimi yöntemleri, verilerin ait olduğu sınıf bilgilerinin varlığı ve gerçekleştirilecek arama stratejisine göre iki alt kategoriye ayrılabilir. Etiket bilgisine göre özellik seçimi yöntemleri, tüm verilerin ait olduğu sınıf bilgisinin bilindiği danışmanlı, verilerin bazılarının sınıf bilgilerinin bilindiği yarı danışmanlı ve sınıf bilgisinin bilinmediği danışmansız özellik seçimi yöntemleri olarak üç alt başlıkta incelenebilir. Arama stratejisine göre yöntemler ise sınıflandırıcı kullanan sarmal yöntemler, özelliklerin ilgisini değerlendiren filtre yöntemler ve ikisinin birleşimi olan gömülü yöntemler olmak üzere üç alt kategoride incelenebilir (Miao ve Niu, 2016). Sürü zekâsına veya evrimsel hesaplama algoritmalarının kullanıldığı yöntemler genellikle arama stratejisi alt kategorisinde olduğundan dolayı bu çalışmada bu alt kategoriyi odaklanılmıştır.



Şekil 2.1. Özellik seçimi yöntemlerinin sınıflandırılması (Miao ve Niu, 2016)

-Filtre yöntemler

Wang ve ark. (2015b) özellik seçimi için sınıf bağımlı fazlalığın kullanıldığı çok amaçlı evrimsel algoritma (multi-objective evolutionary algorithm using class-dependent redundancy for feature selection - MECY-FS) yöntemini önermişlerdir. Hedef sınıfa göre özellikler arasındaki ortak bilgilerinin doğru tahmin edilmesi için yeni bir fazlalık ölçütü önermişlerdir. İlişkinin maksimizasyonu ve fazlalığın minimizasyonu olmak üzere iki amacı olan çok amaçlı evrimsel algoritma (Strength Pareto Evolutionary Algorithm - SPEA2) ile ürettikleri aday özellik kümesinin değerlendirilmesinde önerdikleri fazlalık ölçütünü ve ilişki ölçütünü kullanmışlardır.

Wang ve ark. (2015a) danışmansız özellik seçimi için maksimum izdüşüm minimum fazlalık adını verdikleri yeni bir kriter önermişlerdir. Özellik seçimi problemini, izdüşüm matrisi kullanılarak formüle etmiş ve önerdikleri kriteri denk matris ayrışımı (matrix factorization) problemine dönüştürmüşlerdir. Bu problemin çözümü için ise bir iteratif güncelleme algoritması ve bir açgözlü algoritma önermişlerdir. Önerdikleri yöntemi 4 danışmansız özellik seçimi yöntemiyle karşılaştırmış ve 6 veri setinde önerdikleri yöntemin üstünlüğünün gözlemlendiğini belirtmişlerdir.

Dadaneh ve ark. (2016) makalelerinde karınca kolonisi optimizasyonu (ant colony optimization - ACO) ile danışmansız olasılıksal özellik seçimi (UPFS) yöntemini önermişlerdir. Öncelikle çözüm uzayını, ACO ile arayıp karıncaların seçtiği özellik alt kümesiyle seçilmeyen özellikler arasındaki korelasyondan faydalanarak belli özellikleri seçmişlerdir. Ardından özellikleri, olasılık fonksiyonuna göre derecelendirerek nihai özellik alt kümesini elde etmişlerdir.

Zhang ve ark. (2018) global optimizasyon problemine dönüştürdükleri özellik seçimi problemini, gelişmiş minimum fazlalık maksimum ilişki (improved minimum-

redundancy maximum-relevancy - ImRMR) tabanlı filtre kullanarak kurbağa sıçrama algoritması (shuffled frog leaping algorithm - SFLA) ile çözmeyi hedeflemişlerdir. Önerdikleri algoritmanın optimizasyon aşamasında bazı kısıtlamalar getirerek aday çözümlerin uygunluğunu sağlayıp böylece hesaplama süresini kısalttığını ancak özellik seçiminden önce son özellik alt kümesinin boyutunun belirlenmesi gerektiğini belirtmişlerdir.

Labani ve ark. (2018) metin sınıflandırma için çok değişkenli görelî ayırım kriteri (Multivariate Relative Discrimination Criterion - MRDC) adını verdikleri yeni bir filtre yöntemi önermişlerdir. Geliştirdikleri yöntemde minimum fazlalık maksimum ilişki (minimum redundancy – maximum relevance) kavramı kullanılarak alakasız özelliklerin azaltılmasını hedeflemişlerdir. Önerdikleri yöntemde özellikleri iki aşamada değerlendirmişlerdir. İlk aşamada maksimum ilişkili özellikleri RDC kriteri ile seçip ikinci aşamada ise özellikler arasındaki korelasyona göre fazlalığı tespit etmiş ve sonuç özellik kümesini oluşturmuşlardır. Önerdikleri yöntemi, çok katmanlı algılayıcı (Multilayer Perceptron-MLP), çokterimli naif Bayes (Multinomial Naive Bayes-MNB) ve karar ağacı (Decision Tree-DT) sınıflandırıcılarını kullanarak farklı yöntemlerle üç veri seti üzerinde karşılaştırmışlardır. Önerdikleri yöntemin kesinlik, duyarlılık ve f-ölçütünde en iyi performansa sahip olduğunu belirtmişlerdir.

Solorio-Fernández ve ark. (2020) spektral özellik seçimi ve fazlalık analizinin birlikte kullanıldığı danışmanlı filtre özellik seçimi yöntemi önermişlerdir. 40 gerçek dünya veri seti üzerinde önerdikleri yöntemi test etmiş ve farklı danışmanlı filtre özellik seçimi yöntemleriyle karşılaştırmışlardır. Elde ettikleri sonuçlara istinaden önerdikleri yöntemin karşılaştırdıkları yöntemlerden genel olarak daha iyi sonuçlar verdiğini belirtmişlerdir.

Rostami ve ark. (2020) 3 temel aşamadan oluşan yeni bir PSO tabanlı çok amaçlı özellik seçimi yöntemi önermişlerdir. İlk aşamada tüm özellikler çizge ile temsil edilmiştir. İkinci aşamada çizgedeki tüm düğümlerin özellik merkezilikleri hesaplanmış ve üçüncü aşamada ise PSO tabanlı arama gerçekleştirilmiştir. 5 medikal veri seti kullanarak önerdikleri yöntemi diğer filtre tabanlı özellik seçimi yöntemleriyle karşılaştırmış ve önerdikleri yöntemin performansının karşılaştırdıkları diğer yöntemler içinde en iyi performansa sahip olduğunu belirtmişleridir.

-Sarmal yöntemler

Moradi ve Gholampour (2016) özellik seçimi için 7 aşamadan oluşan yerel arama kullanılan yerel aramalı hibrit parçacık sürü optimizasyonu (hybridized PSO with local

search operations HPSO-LS) algoritmasını önermişlerdir. Öncelikle özellik altkümesi sayısını otomatik olarak belirlemiş ardından özellikleri korelasyon bilgisi kullanılarak benzer ve benzer olmayan gruplara ayırmışlardır. Önceden belirledikleri sayıda parçacıklar oluşturularak PSO'yu başlatmışlardır. Parçacıkların konumlarını yerel ve global en iyi konumlarına göre güncellemiş, özelliklerin korelasyon bilgisiyle yerel arama gerçekleştirmişlerdir. Son olarak parçacıkların uygunluklarını hesaplayıp lokal ve global en iyi değerlerini güncellemişlerdir. Önerdikleri yöntemi 13 kıyas sınıflandırma verisi ile test etmişler ve 5 özellik seçimi yöntemiyle karşılaştırmışlardır. Ayrıca HPSO-LS'yi bilinen filtre tabanlı yöntemlerle de karşılaştırmışlardır. Sonuçlara göre önerdikleri yöntemin diğer filtre ve sarmal yöntemlere göre sınıflandırma başarısını geliştirdiğini belirtmektedirler.

Mlakar ve ark. (2017) yüz ifadesi tanıma sisteminde kullanılmak üzere çok amaçlı farksal gelişim (multiobjective differential evolution - DE) algoritması ile özellik seçimi yöntemi önermişlerdir. Algoritmada seçilen özellik alt kümelerinin uygunluğunu destek vektör makinesi (support vector machine-SVM) sınıflandırıcısı kullanılarak değerlendirmişlerdir. Seçilen özellik sayısının minimizasyonunu ve duygu tanıma doğruluğunun maksimizasyonunu hedeflemişlerdir.

Mafarja ve Mirjalili (2018) balina optimizasyon algoritması (whale optimization algorithm-WAO) tabanlı sarmal özellik seçimi yöntemi önermişlerdir. Farklı seçim mekanizmalarının WAO algoritmasına etkisini incelemek amacıyla turnuva ve rulet tekeri seçim mekanizmalarının ve algoritmanın sömürü yeteneğinin geliştirilmesi amacıyla da çaprazlama ve mutasyonun kullanıldığı iki ikili WAO yöntemi önermişlerdir. Uygunluk fonksiyonunda sınıflandırma başarısını ve seçilen özellik sayılarını dikkate almışlardır. Üretilen çözümlerin uygunluğunun ölçülmesi için kullanılmak üzere kNN sınıflandırıcısını tercih etmişlerdir. Önerdikleri yöntemi, standart kıyas veri setleri üzerinde, PSO, genetik algoritma (Genetic Algorithm-GA), karınca aslanı optimizasyon algoritması (Ant Lion Optimizer-ALO) ve 5 standart filtre özellik seçimi yöntemiyle karşılaştırmışlardır.

Yan ve ark. (2018) soğutma havalandırma sistemlerindeki hata tespiti için geri izlem ileri beslemeli özellik seçimi (back-tracing sequential forward feature selection-BT-SFS) algoritmasını önermişlerdir. BT-SFS kullanılarak oluşturdukları özellik alt kümesinin değerlendirilmesi için SVM sınıflandırıcısı kullanmışlardır.

Zhang ve ark. (2019) maliyet duyarlı özellik seçimi probleminin çözümü için iki arşivli çok amaçlı ABC algoritmasını önermişlerdir. Önerdikleri yöntem, parçacık sürü

optimizasyon algoritmasının hızlı yakınsama özelliğini ve farksal gelişim algoritmasının sömürü yeteneğini entegre ederek daha güçlü bir optimizasyon yöntemi elde etmeyi amaçlamışlardır. Algoritmada üretilen çözümlerin sınıflandırma performansını kNN sınıflandırıcısı kullanarak değerlendirmişlerdir. Önerdikleri yöntemi 8 gerçek veri seti üzerinde 2 geleneksel ve 3 metasezgisel yöntemle karşılaştırmışlardır. Elde ettikleri sonuçlara göre önerdikleri yöntemin maliyet duyarlı özellik seçimi probleminin çözümünde kullanılabilir olacak oldukça rekabetçi bir alternatif yaklaşım olduğunu öne sürmüşlerdir.

Naik ve ark. (2020) özellik seçimi için ikili yarasa algoritmasının (Binary Bat Algorithm-BBA) kullanıldığı bir sarmal yöntem önermişlerdir. Özellik alt kümesini BBA kullanılarak oluşturmuşlardır. Oluşturdukları bu özellik alt kümesini tek geçişli genelleştirilmiş sınıflandırıcı yapay sinir ağı (One-Pass Generalized Classifier Neural Network) sınıflandırıcısı ile değerlendirmişlerdir. Özellik alt kümelerinin uygunluklarını hesaplamak için, doğruluk, hassasiyet, özgüllük ve seçilen özellik sayısının kullanıldığı bir uygunluk fonksiyonu önermişlerdir.

Nouri-Moghaddam ve ark. (2021) çalışmalarında arşiv, ızgara ve bölge tabanlı seçim kavramlarının kullanıldığı orman optimizasyon algoritması (Forest Optimization Algorithm-FOA) tabanlı çok amaçlı bir özellik seçimi algoritması önermişlerdir. Önerdikleri algoritmanın sürekli ve ikili gösterimlerinin kullanıldığı iki farklı versiyonunu geliştirmişlerdir. Önerdikleri algoritmaları 9 UCI veri seti ve iki mikrodizi veri seti üzerinde test etmiş ve 7 geleneksel tek amaçlı ve 5 çok amaçlı yöntemle karşılaştırmışlardır. Elde ettikleri sonuçlara göre önerdikleri algoritmanın iki versiyonunun da performanslarının aynı olduğunu ve tek amaçlı yöntemlerden daha üstün olduğunu belirtmişlerdir. Karşılaştırdıkları çok amaçlı yöntemlere göre önerdikleri algoritmanın sürekli temsili kullandıkları versiyonunun daha az sayıda özellik seçtiği ve çoğu durumda da sınıflandırma hatasını azalttığını belirtmişlerdir.

-Gömülü yöntemler

Zhang ve ark. (2015) sürüş örüntüsü tanıma için istatistiksel bilgilerin SVM kullanıldığı gömülü bir özellik seçimi yaklaşımı önermişlerdir. Lu ve ark. (2017) çalışmalarında gen ifadesi verilerinin sınıflandırılması için ortak bilgi maksimizasyonu ve adaptif genetik algoritmanın kullanıldığı gömülü bir özellik seçimi algoritması önermişlerdir. Yaptıkları deneysel çalışmalar neticesinde önerdikleri algoritmayla gen anlatımı verisinin boyutunun önemli ölçüde azaltıldığını ve geleneksel özellik seçimi

algoritmalarıyla karşılaştırıldığında en yüksek sınıflandırma başarısının elde edildiğini belirtmişlerdir. Jain ve ark. (2018) kanser sınıflandırması için korelasyon tabanlı özellik seçimi (Correlation-based Feature Selection - CFS) ve iyileştirilmiş ikili PSO (improved-Binary Particle Swarm Optimization - iBPSO) algoritmasını birleştiren iki aşamalı hibrit bir model önermişlerdir. İlk aşamada sınıfta daha etkin öngörüye sahip düşük boyutlu gen alt kümesini, gereksiz ve alakasız genleri çıkararak korelasyon tabanlı özellik seçimiyle seçmişlerdir. İkinci aşamada ise önceki aşamada bulunan gen alt kümesinden iBPSO ile optimal gen kümesini seçmişlerdir. Her bir gen alt kümesinin sınıflandırma performansını, naif Bayes sınıflandırıcısı ile değerlendirmişlerdir. Önerdikleri yöntemi 11 kıyas kanser mikrodizi veri seti üzerinde filtre, sarmal ve gömülü 7 farklı yöntemle karşılaştırmışlardır. Veri setlerinin çoğunda önerdikleri yöntemin sınıflandırma başarısı ve seçilen gen sayısı açısından diğer yöntemlerden daha iyi bir performansa sahip olduğunu belirtmişlerdir. Maldonado ve López (2018) sınıf dengesizliği probleminin çözümü için maliyet duyarlı SVM ve destek vektör veri açıklayıcısının kullanıldığı ve yüksek boyutluluğa çözüm getirmek amacıyla ölçekleme faktörü kullanılarak özelliklerin seçilmesine ceza uygulanan yeni bir gömülü özellik seçimi yaklaşımı önermişlerdir. Önerdikleri yöntemi 12 dengesiz mikrodizi veri seti üzerinde test etmişlerdir. Selvakumar ve Muneeswaran (2019) ağ saldırı tespiti için ateşböceği optimizasyon algoritması tabanlı özellik seçimi yöntemi önermişlerdir. Özellik seçiminde, ortak bilgi tabanlı bir filtre yöntemden ve C4.5 ve Bayeşçi Ağlar sınıflandırıcı olarak kullanıldığı 2 sarmal yöntemden 3 farklı özellik alt kümesi elde etmişlerdir. Belirledikleri denklemlerle bu özellik kümelerinden oylama yöntemiyle özellik seçimini gerçekleştirmişlerdir. Guo ve ark. (2019) çok etiketli biyoinformatik veri öğrenme problemi için grup tabanlı gömülü özellik seçimi (ensemble embedded feature selection-EEFS) yöntemini önermişlerdir. Önerdikleri yöntemle hem etiketler arasındaki korelasyonu tespit etmiş hem de çok etiketli sınıflandırıcılar ve değerlendirme metrikleriyle bu korelasyondan faydalanmışlardır. Önerdikleri yöntemi 5 çok etiketli biyoinformatik veri seti üzerinde test edip diğer algoritmalarla karşılaştırmış ve önerdikleri yöntemin daha üstün performansa sahip olduğunu belirtmişlerdir.

2.2. İkili ABC Algoritması Üzerine Kaynak Araştırması

ABC algoritması 2005 yılında Karaboga (2005) tarafından sürekli optimizasyon için önerilmiş bir sürü zekâsı algoritmadır. Algoritmanın gelişim safhasında 2010 yılına kadar herhangi bir özellik seçiminde kullanılmadığı bu yıla kadar performans

karşılaştırmaları, uygulamaları veya performansının sürekli optimizasyonda iyileştirilmesi üzerinde durulduğu görülmüştür (Karaboga ve ark., 2007; Karaboga ve Basturk, 2007; 2008; Abu-Mouti ve El-Hawary, 2009; Benala ve ark., 2009; Karaboga ve Akay, 2009; Karaboga ve Ozturk, 2009; Karaboga, 2009; Mala ve Mohan, 2009; Marinakis ve ark., 2009; Omkar ve Senthilnath, 2009; Pulikanti ve Singh, 2009; Singh, 2009). 2010 yılında özellik seçimi için ABC algoritmasının uygulandığı 3 çalışma bulunmaktadır. Bu yıldan sonra algoritmanın ikili optimizasyon ve özellik seçimi probleminin çözümü için uygulanması artarak devam etmiştir. Bilindiği üzere özellik seçimi bir ikili optimizasyon problemidir ve temel ABC algoritması sürekli optimizasyon için önerilmiştir. Bundan dolayı özellik seçimi problemine uygulanan ABC varyantları ikili optimizasyon algoritması olarak ele alınmaktadır. Özellik seçimi için uygulanmış olan ABC algoritmaları literatür özetinin kalan kısmında yıl bazında kronolojik sırada verilmiştir.

Wang ve ark. (2010) ağıdaki saldırı tespitinde destek vektör makineleri ile ABC algoritmasını kullanmıştır. Bu çalışmada temel ABC algoritması öncelikle SVM'nin parametrelerini optimize etmek amacıyla ve ikili ABC algoritması ise özellik seçimi amacıyla kullanılmıştır. Temel ABC algoritmasını özellik seçimi için bir transfer fonksiyonu (sigmoid) kullanarak değiştirmişlerdir. Bir başka çalışmada ABC algoritması denetimli bulanık C-ortalama algoritması ile kümeleme amacıyla kullanılmıştır. Bu ikili ABC varyantında işçi arılar tarafından kovana getirilen çözümler nektar kalitesine bağlı olarak (özellik alt kümesi üzerindeki kümeleme başarısı) belirli sayıda gözcü arıya dağıtılmakta ve gözcü arılarda bu çözümleri iyileştirmeye çalışmaktadır (Shokouhifar ve Farokhi, 2010). Aynı algoritma yapay sinir ağları ile sınıflandırma performansının artırılması amacıyla özellik seçimi için kullanılmıştır (Shokouhifar ve Sabet, 2010).

2011 yılında bir çalışmada SVM algoritması ile ABC algoritması özellik seçimi amacıyla birlikte kullanılmış ve seçilen özelliklerin sınıflandırıcı performansını artırması üzerine durulmuştur. Bu çalışmada ABC algoritmasının ikilileştirilmesi için Wang ve ark. (2010) tarafından yapılan çalışmadaki ile aynı şekilde sigmod transfer fonksiyonu kullanılmıştır (Prasartvit ve ark., 2011).

2012 yılında ABC algoritması bir grup sınıflandırıcı (destek vektör makineleri, naif bayes ve karar ağaçları) algoritmalarıyla birleştirilmiş ve temel ABC algoritmasıyla özellik seçimi yapılmış ve grup sınıflandırıcının performansı artırılmaya çalışılmıştır (Palanisamy ve Kanmani, 2012a). Tuş vuruş dinamikleri kullanıcıyı tespit etmek amacıyla kullanılan bir tekniktir. Bu teknik özellik çıkarma, seçme ve sınıflandırmayı da

içermektedir. Özellik seçme amacıyla ABC, sınıflandırma amacıyla ise hata geri yayılım algoritmasıyla eğitilmiş yapay sinir ağı kullanılmıştır (Akila ve ark., 2012). ABC algoritması bir başka çalışmada karar ağaçlarıyla birlikte UCI veri tabanındaki 10 veri setindeki özellikleri seçmek amacıyla kullanılmıştır (Palanisamy ve Kanmani, 2012b). Kütle spektrometri verilerinden elde edilen özellik kümesinden özellik seçimi amacıyla ABC algoritması ve SVM birlikte kullanılmıştır. ABC algoritmasındaki çözüm güncelleme kuralı doğrudan farklı bir kaynağın boyutunun aday çözüme kopyalanması ile değiştirilmiştir (Syarifahadilah ve ark., 2012).

ABC algoritması üzerine ilgi 2013 yılından itibaren oldukça artmış ve araştırmacılar tarafından ABC algoritması birçok iyileştirme ve uygulamanın konusu olmuştur. Bu bağlamda ikili ABC üzerine literatür özeti tablo şeklinde sunulmuştur.

Cizelge 2.1. ABC algoritmasının özellik seçimi amacıyla geliştirildiği veya uygulandığı çalışmalar

Referans	Yöntem	Amaç Fonksiyon	Problem
(Mohammadi ve Abadeh, 2013)	Sürekli ABC	SVM	Stenografik analiz
(Rakshit ve ark., 2013)	ABC-Kümeleme	Global doğruluk indeksi	Beyin-Bilgisayar Arayüzü, EEG verileri
(Uzer ve ark., 2013)	ABC-kümeleme	SVM	UCI veri setleri
(Chahkandi ve ark., 2013)	Kaotik ABC- Yer değiştirme	kNN	UCI veri setleri
(Shunmugapriya ve ark., 2013)	Sürekli ABC	Karar Ağacı	UCI veri setleri
(Schiezaro ve Pedrini, 2013)	İkili ABC – Rastgele ekleme	SVM	UCI veri setleri
(Mohammadi ve Abadeh, 2014)	Sürekli ABC	SVM	Stenografik analiz
(Uzer ve ark., 2014)	ABC-kümeleme	YSA	UCI Veri seti
(Yusoff ve ark., 2014)	ABC-DE yer değiştirme	SVM	Kütle spektrometri verileri
(Subanya ve Rajalaxmi, 2014)	Sürekli ABC	SVM	UCI veri seti
(Sivakumar ve Chandrasekar, 2014)	ABC yer değiştirme	kNN	Akciğer BT tarama görüntüleri
(Kaushik ve Keswani, 2014)	ABC-PSO eşik değer ile dönüşüm	N/A	UCI veri setleri
(Hancer ve ark., 2015a)	İkili ABC – Benzerlik indeksi	kNN	UCI veri setleri
(Chen ve ark., 2015)	Sürekli ABC – eşik değer	SVM	UCI veri setleri
(Hancer ve ark., 2015b)	Çok amaçlı ABC – genetik operatörler	kNN	UCI veri setleri
(Zorarpacı ve ark., 2015)	İkili ABC – lojik operatör	Karar Ağacı	UCI veri setleri
(Lin ve Hsieh, 2015)	ABC-PSO, eşit değer	SVM	UCI veri setleri
(Agrawal ve Chandra, 2015)	ABC, uygunluk tabanlı yer değiştirme	kNN, SVM	Rahim ağız kanseri BT Görüntüleri
(Ye ve Chen, 2015)	İkili ABC, lojik operatör	Minimum öznelik azaltma	UCI veri setleri
(Zorarpacı ve Özel, 2016)	ABC-DE, lojik operatör	Karar ağacı	UCI veri setleri
(Yavuz ve Aydın, 2016)	Sürekli ABC, açılı modülasyonu	SVM, kNN	UCI veri setleri

(Özger ve ark., 2016) (Kumar ve Nedunchezian, 2016)	İkili ABC karşılaştırması ABC-BBO, göç tabanlı	kNN SVM	UCI veri setleri Gen seçimi
(Li ve ark., 2016)	Çok amaçlı ikili ABC, göç tabanlı	ELM	Gen seçimi
(Shi ve ark., 2016) (Shunmugapriya ve Kanmani, 2017)	Sürekli ABC, eşik değer ACO-ABC melez	kNN Karar ağacı	UCI veri seti UCI veri seti
(Feng ve ark., 2017) (Nadira ve Rustam, 2018)	ABC, eşik değer ABC-PSO, eşik değer	kNN, karar ağacı SVM	DNA Mikrodizi verileri Kanser verileri
(Hancer ve ark., 2018)	Çok Amaçlı ABC, ikili gösterimde çaprazlama ve mutasyon, sürekli gösterimde eşik değer	kNN	UCI veri setleri
(Keleş ve Kılıç, 2018)	ABC, eklemeli yaklaşım	kNN vd.	Scadi veri seti
(Kiliç ve Keleş, 2018)	ABC, eklemeli yaklaşım	kNN vd.	Z-Alizadeh Sani veri seti
(Kuo ve ark., 2018)	ABC, eşik değer	SVM	UCI veri setleri
(Keleş ve ark., 2018)	ABC, eklemeli yaklaşım	kNN vd.	Beton dayanımı veri seti
(Seth ve Chandra, 2018)	ABC, eşik değer	Karar ağacı	Dos atağı veri seti
(Arslan ve Ozturk, 2019)	ABC, programlama yaklaşımı	N/A	UCI veri setleri
(Santana Jr ve ark., 2019)	İkili ABC, yer değiştirme	kNN	UCI veri setleri
(Zhang ve ark., 2019)	Çok amaçlı ABC, eşik değer	kNN	UCI veri setleri
(Badem ve ark., 2019)	ABC, eşik değer	SVM vd.	Parkinson veri seti
(Kharis ve ark., 2019)	ABC, mod tabanlı (tam sayı öznitelik indeksi)	SVM	Beyin kanseri veri seti
(Ge ve ark., 2019)	GA-ABC, tersini alma	SVM	Kanser veri setleri
(Du ve ark., 2019)	ABC, eşik değer	kNN	UCI veri setleri
(Chu ve ark., 2020a; Chu ve ark., 2020b)	ABC, transfer fonksiyonu	kNN	UCI veri setleri
(Bindu ve Sabu, 2020)	ABC, tersini alma	Rasgele Orman	UCI veri setleri
(Durgut ve ark., 2020)	İkili ABC, lojik operatör, benzerlik, transfer fonksiyonu, yer değiştirme	kNN	Parkinson veri seti
(Mlakar ve Fister, 2020)	İkili ABC, benzerlik indeksi	kNN	UCI veri setleri
(Wang ve ark., 2020a)	ABC, yuvarlama ve birleştirme	SVM	Sinyal veri setleri
(Wang ve ark., 2020b)	Çok amaçlı ABC, eşik değer	k-Ortalamalar	UCI veri setleri
(Punitha ve ark., 2021)	ABC, kaydırma, yer değiştirme, çaprazlama	YSA	UCI veri seti
(Harudin ve ark., 2021)	İkili ABC, lojik operatör	SNR	UCI veri setleri
(Hanbay, 2021)	ABC, entropi tabanlı	SVM	UCI veri setleri
(Lin ve ark., 2021)	ABC, eşik değer, çaprazlama	Regresyon modeli	QSAR
(Zhang ve ark., 2021)	ABC, transfer fonksiyonu	kNN	UCI ve diğer veri setleri
(Bindu ve Sabu, 2021)	ABC, çaprazlama ve mutasyon	N/A	UCI
(Nagarajan ve ark., 2021)	ABC-GA, eklemeli yaklaşım	SVM vd.	UCI veri setleri
(Hijazi ve ark., 2021)	ABC, komşudan eklemeli yaklaşım	SVM vd.	Arapça metin veri setleri
(Rostami ve Kaveh, 2021)	BBO-ABC, komşuluk ve mutasyon	SVM	SAR görüntüsü

Çizelge 2.1'den görüldüğü üzere ABC algoritması veya geliştirilmiş varyantları özellik seçme amacıyla sıklıkla kullanılmıştır. Bunların yanı sıra ABC üzerine literatür değerlendirmeleri de yayınlanmıştır. 2014 yılında tüm ABC çalışmaları üzerine yapılan bir literatür özeti (Karaboga ve ark., 2014) çalışmasında, ABC ile sinyal, görüntü ve video işleme uygulamaları üzerine yapılan çalışmaların literatür değerlendirmesi (Akay ve Karaboga, 2015) çalışmasında, ABC ile kümeleme üzerine yapılmış çalışmaların değerlendirmesi (Kumar ve ark., 2017) çalışmasında, sürü zekâsı algoritmaları ile veri madenciliğinde özellik seçimi üzerine yapılmış çalışmaların genel bir değerlendirmesi ise (Nguyen ve ark., 2020) çalışmasında bulunabilir. Ayrıca 2021 yılında ABC algoritmasının ayrıklı versiyonları üzerine geniş bir literatür çalışması yayınlanmıştır (Akay ve ark., 2021). Bu çalışmada da ikili ABC varyantları ve çözüm için uygulandıđı problemler kategorik olarak değerlendirmeye tabi tutulmuştur.

Literatür özetleri ile bu tez çalışması kesiştirildiğinde bu çalışmadaki katkılar aşağıda madde madde özetlenmiştir.

- ❖ Stigmerji ve xor tabanlı ikili ABC algoritması ilk defa özellik seçimine uygulanmıştır.
- ❖ 8 adet ikili ABC algoritması özellik seçme amacıyla kullanılmış ve karşılaştırma yapılmıştır.
- ❖ Sınıflandırıcı olarak aşırı öğrenme makinesi bu algoritmalar için ilk defa kullanılmıştır.

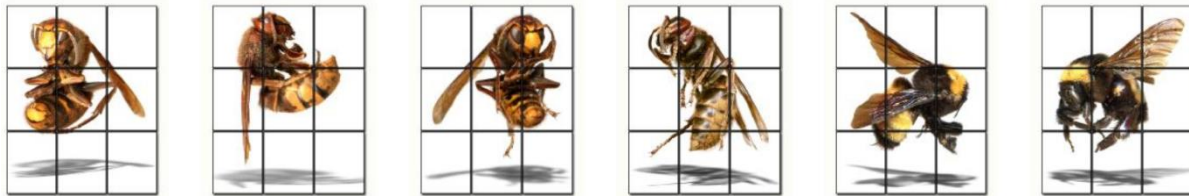
Yukarıda sayılan katkılardan dolayı ikili ABC algoritması varyantlarının diđer sürü zekâsı algoritmalarıyla karşılaştırılması bu tez kapsamına alınmamıştır.

3. MATERYAL VE YÖNTEM

Bu bölümde öncelikle temel ABC algoritması verildikten sonra bu algoritmanın ikili optimizasyon için uyarlanmış 8 farklı varyantı sunulmuştur. Ayrıca elde edilen çözümleri değerlendirmek amacıyla kullanılan yapay sinir ağı modeli ve eğitim algoritması olarak ELM algoritması açıklanmıştır. Önerilen özellik seçme yönteminin performansının test edilmesi amacıyla 5 farklı veri seti kullanılmış olup bu veri setlerine dair açıklamalara da bu bölümde yer verilmiştir.

3.1. Temel ABC Algoritması

ABC algoritması 2005 yılında sürekli optimizasyon problemlerini çözmek amacıyla Türk bilim adamı Prof. Dr. Derviş Karaboga tarafından geliştirilmiştir (Karaboga, 2005). Algoritma, bal arılarının yiyecek arama ve bilgi paylaşımı gibi zeki davranışlarından etkilenerek oluşturulmuştur. Bal arıları kovandan çıktıktan sonra yiyecek kaynağı aramaktadırlar ve buldukları kaynakla ilgili bilgileri (kaynak konumu, kaynak kalitesi vb.) kovana taşımaktadır. Kovanda bilgi paylaşımını ise “waggle” adı verilen bir dans ile diğer arılarla paylaşmaktadırlar. Dansa ait figürler Şekil 3.1’de gösterilmiştir. “Şekil 3.1. doğrudan (Karaboga, 2005) raporundan alınmıştır”.



Şekil 3.1. Yiycek kaynağı bilgilerinin kovanda paylaşımı için yapılan “Waggle” dans (Karaboga, 2005)

Bu dans ile kaynak konum ve kalite bilgisi paylaşılırken bazı arılar (gözcü arılar) bu arı ile etkileşime girmektedir. Bütün arılar konum ve kalite bilgisi paylaştığından dolayı kovan etrafındaki yiyecek kaynaklarının kalite bilgisi kovanda oluşmaktadır. Bu sayede kovana nektar taşıma işi daha verimli hale getirilmeye çalışılmaktadır. Doğadaki bal arılarının zeki davranışlarından esinlenilerek ortaya konulmuş olan ABC algoritması popülasyon temelli iteratif bir algoritmadır ve algoritmanın başlangıcında verilen durdurma kriteri (maksimum çevrim sayısı, uygunluk fonksiyonu değerlendirme sayısı, eşik optimizasyon hatası vb.) sağlanana kadar çalışmaya devam edecektir. Her ne kadar gerçek arı kovanlarında arı sayıları değişebilmekle birlikte algoritmada arı sayıları (popülasyondaki arı sayısı – popülasyon boyutu) sabit bir değer olarak yer almaktadır. Bu başlangıç popülasyondaki arıların yarısı görevli arılar ve yarısı da gözcü arılardan

oluşması durumu da algoritma için bir kabul olarak karşımıza çıkmaktadır. Bir diğer önemli parametre ise herhangi bir arının kâşif olması durumudur. Bir arının kâşif arı olup olmamasına kaynağın durumu ile karar verilmektedir. Algoritma için bu varsayımları akılda tutmak kaydıyla algoritmanın aşamaları aşağıda verilmiştir.

Başlatma

1. Popülasyondaki arı sayısına karar ver (N).
2. Arıların yarısını gözcü arı, diğer yarısını kâşif arı olarak ayarla.
3. Bütün kâşif arılara Denklem 3.1 ile verilen rastgele bir yiyecek kaynağı oluştur.

$$X_i^j = X_j^{\min} + r_i^j \times (X_j^{\max} - X_j^{\min}) \quad i = 1, 2, \dots, NP \text{ and } j = 1, 2, \dots, D \quad (3.1)$$

Denklemdaki X popülasyonu, i popülasyondaki yiyecek kaynağını, j boyutu, NP kâşif arı sayısını, D ise problemin toplam boyutunu gösterir.

4. Kendisi için yiyecek kaynağı oluşturulan bütün kâşif arıları görevli arı olarak kabul et ve kaynakların kalitesini Denklem 3.2 ile hesapla

$$fit_i = \begin{cases} 1/(1+f_i), & \text{if } (f_i \geq 0) \\ 1+abs(f_i), & \text{otherwise} \end{cases} \quad (3.2)$$

Denklemdaki i yiyecek kaynağını, fit yiyecek kaynağının kalitesini, f bu yiyecek kaynağı kullanılarak hesaplanmış amaç fonksiyon değerini (çözülmek istenen problemin amaç fonksiyonu) göstermektedir.

5. En iyi yiyecek kaynağını ezberle (G).
6. Her bir yiyecek kaynağı için birer sayaç tanımla ve sayaçları sıfırla.
7. Bir sınır (limit) değeri oluştur (bu sınır değerini geçen arılardan biri tekrar kâşif arı olacak)
8. Durdurma kriterini tanımla.

Görevli Arı Aşaması

Her bir görevli arı için;

1. Rastgele bir komşu yiyecek kaynağı seç (k).
2. Rastgele bir boyut seç (j)
3. Denklem 3.3'ü kullanarak yeni bir yiyecek kaynağı üret ($V_i = X_i$)

$$V_i^j = X_i^j + \phi_i^j \times (X_j^i - X_j^k), \quad i \neq k \text{ and } j \in \{1, 2, \dots, D\} \quad (3.3)$$

4. Yeni oluşturulan yiyecek kaynağının Denklem 3.2 ile hesaplanan kalitesi mevcut yiyecek kaynağından daha İYİYESE

- 4.1. Yeni yiyecek kaynağının yerine eskisini yaz ($X_i = V_i$)

4.2. Yiyecek kaynağının sayacını sıfırla

DEĞİLSE

4.1. Yiyecek kaynağının sayacını bir artır.

Gözcü Arı Aşaması

Denklem 3.4 ile her bir yiyecek kaynağının seçilme ihtimalini (p_i) hesapla.

$$p_i = \frac{fit_i}{\sum_{k=1}^{NP} fit_k} \quad (3.4)$$

Her bir gözcü arı için;

1. Uygunluk tabanlı oransal seçim (rulet tekerliği uygulaması) ile bir yiyecek kaynağını seç (seçilen i . yiyecek kaynağı olmak üzere)
2. Rastgele bir komşu yiyecek kaynağı seç (k).
3. Rastgele bir boyut seç (j)
4. Denklem 3.3 kullanarak yeni bir yiyecek kaynağı üret ($V_i=X_i$)
5. Yeni oluşturulan yiyecek kaynağının Denklem 3.2 ile hesaplanan kalitesi mevcut yiyecek kaynağından daha İYİYESE

5.1. Yeni yiyecek kaynağının yerine eskisini yaz ($X_i=V_i$)

5.2. Yiyecek kaynağının sayacını sıfırla

DEĞİLSE

5.1. Yiyecek kaynağının sayacını bir artır.

Kâşif Arı Aşaması

1. En yüksek değere sahip sayacı belirle.
2. Bu değeri “limit” ile kıyasla.
3. Limit değerinden büyükse yiyecek kaynağı yerine Denklem 3.1 ile yeni bir kaynağı oluştur.
4. Denklem 3.2 ile yiyecek kaynağının kalitesini hesapla.
5. Yiyecek kaynağının sayacını sıfırla.

Durdurma

1. En iyi çözümü bul ve G ile kıyasla ve daha iyi olanı G olarak ata.
2. Durdurma kriteri sağlandıysa G 'yi raporla.
3. Durdurma kriteri sağlanmadıysa Görevli Arı Aşaması'na git.

3.2. Mod Tabanlı İkili ABC Algoritması

Mod tabanlı ikili ABC algoritması, kısaca MABC, kapasitesiz tesis yerleşim problemlerinin çözümü üzerinde performans analiz yapılmış bir ikili ABC algoritmasıdır (Kiran, 2015). Bilindiği üzere ikili bir optimizasyon probleminde amaç fonksiyona ikili değerler gönderilmektedir. Temel ABC algoritması ise sürekli değerlere sahip karar değişkenlerinin gönderildiği fonksiyonları optimize etmek için önerilmiştir. Kısacası MABC algoritmasında çözümlerin kalitesinin hesaplanması için kullanılan amaç fonksiyona çözümler değerlendirilmek üzere gönderilmeden önce bir dönüşüm işlemine tabi tutulmaktadır. Bu dönüşüm işleminde sürekli çözümler aşağıdaki denklem ile ikili çözümlere dönüştürülmektedir.

$$z_i = \text{round}(|y_i \bmod 2|) \bmod 2 \quad (3.5)$$

Denklemden z_i ikili çözümü, y_i sürekli çözümü göstermektedir. Bu dönüşüm işlemi ile algoritmanın genel yapısında önemli bir değişiklik yapılmadan temel ABC algoritması ikili optimizasyon problemlerini çözecek şekilde yapılandırılmıştır.

3.3. Açık Modülasyonu İkili ABC Algoritması

Açık modülasyonu ikili ABC algoritması, kısaca AABC, herhangi bir ikili optimizasyon probleminin çözümü için problemin 4 boyutlu sürekli optimizasyon problemine indirgenmesine dayanmaktadır (Pampará ve Engelbrecht, 2011). Bu sayede ikili karar değişkenlerinin sayısından bağımsız olarak 4 boyutlu sürekli problemde uygun değerler bulunduğu taktirde istendiği kadar uzunlukta bit dizisi elde edilebilir. Açık modülasyonu ile ikili karar değişkenleri için değer üretme işlemi Denklem 3.6, 3.7 ve 3.8 ile yapılmaktadır.

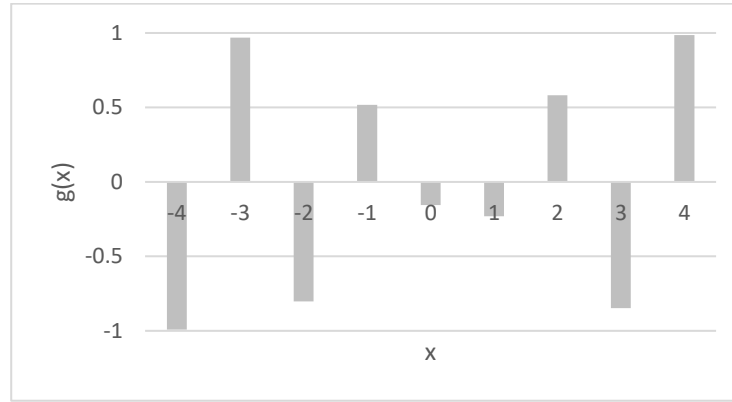
$$g(x) = \sin(F \times b \times \cos(F \times c)) + d \quad (3.6)$$

$$F = 2\pi(x - a) \quad (3.7)$$

$$k(x) = \begin{cases} 1, & \text{eğer } (g(x) > 0) \\ 0, & \text{değilse} \end{cases} \quad (3.8)$$

Denklemlerde, a, b, c ve d ABC ile optimize edilecek katsayılarıdır. a=0,4, b=0,2, c=-0,5, d=0 olduğu durumda x, [-4,4] aralığında 1 aralık ile 9 bit üretilmek istenirse ilgili

denklemlerin çözümlenmesiyle [010100101] ikili dizisi elde edilecektir ve Şekil 3.2’de bu durum gösterilmiştir.



Şekil 3.2. Açık modülasyonu ile üretilen değerler

Şekil 3.2’de üretilen değerler Denklem 3.8 ile ikili hale getirilmektedir. Daha fazla veya az sayıda ikili değer elde edilmek istendiğinde ise x için belirlenen aralık değiştirilerek bu sağlanabilmektedir. AABC algoritması da MABC algoritması gibi temel ABC algoritmasında önemli bir değişikliğe sebep olmamaktadır. Hatta sürekli ABC algoritması 4 sürekli karar değişkeni için optimum değeri ararken amaç fonksiyona çözüm gönderilmeden önce açık modülasyon tekniği ile ikili optimizasyon problemi için istenen sayıda ikili karar değişkeni üretmektedir.

3.4. Xor Tabanlı İkili ABC Algoritması

Özel veya operatörü girişler aynı ise çıkış 0, girişler farklı ise çıkış 1 olacak şekilde çalışan bir mantık kapısıdır. Ayrıca bu kapı ile toplayıcı devreleri tasarlanabilmektedir. Bundan dolayı ABC algoritmasının ikili optimizasyon problemlerini çözecek şekilde yapılandırılması için kullanılabilir. Kiran ve Gündüz (2013) ABC algoritmasında iki değişiklik yaparak önerdikleri algoritmayı, kısaca XABC, kapasitesiz tesis yerleşim problemlerini çözmek için kullanmışlardır. Birinci değişiklik algoritmanın başlatılması sırasında kâşif arılara çözümlerin oluşturulmasında Denklem 3.9’un kullanılmasıdır.

$$X_i^j = \begin{cases} 0, & \text{eğer } (r_{i,j} > 0,5) \\ 1, & \text{değilse} \end{cases} \quad (3.9)$$

Denklemdaki $r_{i,j}$ $[0,1]$ aralığında üretilmiş rastgele bir sayıdır. Bu sayede algoritma sürekli çözüm uzayında değil ikili yapılandırılmış çözüm uzayında başlatılmıştır. İkinci değişiklik ise hem işçi hem de gözcü arılar için çözüm güncelleme denkleminin Denklem 3.10 ile değiştirilmesidir.

$$V_i^j = X_i^j \oplus [\varphi (X_i^j \oplus X_k^j)] \quad (3.10)$$

Denklemdaki \oplus özel veya (xor) operatörüdür ve doğruluk tablosu Çizelge 3.1’de verilmiştir. Aday çözümler ikili uzayda başlatıldığından dolayı özel veya operatörü Denklem 3.10’da kullanılmıştır. φ ise sağdaki terimden elde edilen sonucun %50 ihtimalle tersini almaktadır ve Tablo 3.1’de sonucun tersinin alınması Q’ ile gösterilmiştir.

Çizelge 3.1. Özel veya (xor) operatörünün doğruluk tablosu ve sonucun terslenmesi

A	B	Q	Q'
0	0	0	1
0	1	1	0
1	0	1	0
1	1	0	1

Çizelge 3.1’de A ve B özel veya operatörünü girişlerini ve Q özel veya operatörünün çıkışını göstermektedir.

3.5. Bitsel Operatör Tabanlı İkili ABC Algoritması

Bitsel operatör tabanlı ikili ABC (Jia ve ark., 2014), kısaca BABC, algoritması da XABC algoritması ile benzer şekilde çalışmaktadır. BABC algoritması da Denklem 3.9 ile başlatılmakta ve ikili yapılandırılmış çözüm uzayında aday çözümler üretmektedir. BABC’nin XABC’den farkı ise çözüm güncelleme denkleminde ve (\wedge), veya (\vee), özel veya operatörlerinin tümünün kullanılmasıdır ve Denklem 3.11’de BABC’nin çözüm güncelleme denklemi verilmiştir.

$$V_i^j = X_i^j \oplus [\phi_{ij} \wedge (X_i^j \vee X_k^j)] \quad (3.11)$$

Denklemda ϕ_{ij} 0 veya 1 değeri alabilen bir değişken olarak kullanılmıştır ve Denklem 3.12’de değişkene yapılacak atamanın nasıl yapılacağı verilmiştir.

$$\phi_{ij} = \begin{cases} 1, rand(0,1) < r \\ 0, rand(0,1) \geq r \end{cases} \quad (3.12)$$

Denklemdaki $rand(0,1)$ $[0,1]$ aralığında üretilmiş rastgele bir sayıyı ifade etmektedir ve r , ϕ_{ij} 'nin 0 ve 1 olma olasılığını kontrol eder. Yukarıdaki açıklamalar altında Denklem 3.11 ile aday çözümün ilgili boyutunun ne olacağına aşağıdaki tablo ile karar verilebilir.

Çizelge 3.2. Denklem 3.11'in çıkış tablosu

X_i^j	X_k^j	ϕ_{ij}	$(X_i^j \vee X_k^j)$	$\phi_{ij} \wedge (X_i^j \vee X_k^j)$	V_i^j
0	0	0	0	0	0
0	0	1	0	0	0
0	1	0	1	0	0
0	1	1	1	1	1
1	0	0	1	0	1
1	0	1	1	1	0
1	1	0	1	0	1
1	1	1	1	1	0

Çizelge 3.2'den görüleceği üzere aday çözümün ilgili boyutu %62,5 ihtimalle 0, %37,5 ihtimalle ile 1 olacaktır ve algoritmada aday çözümler için 0'a doğru sapma (bias) olduğu sonucuna varılabilir.

3.6. Farklılık Tabanlı İkili ABC Algoritması

Farklılık (dissimilarity) tabanlı ikili ABC algoritması (Kashan ve ark., 2012), kısaca FABC, aday çözümü oluşturacak çözümlerin birbirlerine benzerliğine ve kalıtıma dayalı bir çözüm güncelleme mekanizması ile ABC algoritmasının entegre edilmesiyle oluşturulmuş bir ikili optimizasyon algoritmasıdır. İki çözümün birbirine benzemesi Denklem 3.13 kullanılarak hesaplanır.

$$Similarity(X_i, X_k) = \frac{m_{11}}{m_{01} + m_{10} + m_{11}} \quad (3.13)$$

Denklemda m_{11} , m_{01} ve m_{10} ikili çözümlerdeki eş indislerin 1-1, 0-1 ve 1-0 olması durumlarının sayılarak elde edilmesine dayanır. Benzerlik Denklem 3.13 ile gösterildiğine göre farklılık Denklem 3.14 ile elde edilebilir ve $[0,1]$ aralığındadır.

$$Dissimilarity(X_i, X_k) = 1 - Similarity(X_i, X_k) = 1 - \frac{m_{11}}{m_{01} + m_{10} + m_{11}} \quad (3.14)$$

Bu hesaplamalar yapıldıktan sonra ikili uzayda başlatılan çözümlerden yeni yeni çözümlerin elde edilmesi için Denklem 3.15 kullanılır.

$$Dissimilarity(V_i, X_i) \approx \varphi. Dissimilarity(X_i, X_k) \quad (3.15)$$

Denklem 3.15'te amaç V_i çözümünün oluşturulması olduğundan ve sağ tarafı Denklem 3.14 ile hesaplandığında A sonucunu verdiği kabul edilirse;

$$\min \left| 1 - \frac{m_{11}}{m_{11}+m_{10}+m_{01}} - A \right| \quad (3.16)$$

kısıtlar:

$$m_{11} + m_{01} = n_1 \quad (3.17)$$

$$m_{10} \leq n_0 \quad (3.18)$$

$$m_{11}, m_{10}, m_{01} \geq 0 \text{ ve tam sayı} \quad (3.19)$$

tam sayı programlama probleminin çözülmesi ile çözüm ile aday çözüm arasındaki farklılık katsayısı hesaplanabilir. Denklemlerdeki n_1 , X_i çözümündeki birlerin, n_0 ise sıfırların toplam sayısıdır. Bu problemin çözümü $D = n_0 + n_1$ (D ikili karar değişkeni sayısı) olduğundan dolayı sayım tekniğiyle $O(D^2)$ karmaşıklığa sahip olur (Kashan ve ark., 2012).

Yukarıdaki problemin çözümü üretilecek aday çözümdeki birlerin-sıfırların sayısını ve m_{11} , m_{10} ve m_{01} değerlerini vermektedir. Aday çözümün oluşturulması için ise aşağıdaki adımlar uygulanır.

- i) $1 \times D$ dizisini (V) oluştur ve tüm elemanlarına 0 değerini ata.
- ii) (Kalıtım-Inheritance) Herhangi bir mantıkla V aday çözümündeki m_{11} tane elemana X_i 'deki konumlarına (X_i konumu 1 olan) uygun olarak 1 değerini ata.
- iii) (Mirastan Farklılık-Disinheritance) Herhangi bir mantıkla V aday çözümündeki m_{10} tane elemana X_i 'deki konumlarına (X_i konumu 0 olan) uygun olarak 1 değerini ata.

3.7. Genetik Operatör Tabanlı İkili ABC Algoritması

Genetik operatörlerden çift noktalı çaprazlama operatörü ve takas (swap) operatörünün çözüm güncelleme mekanizması olarak kullanımını içeren genetik operatör tabanlı ikili ABC, kısaca GABC, algoritması Ozturk ve ark. (2015) tarafından önerilmiştir. Bu algorithmada aday çözüm üretmek amacıyla seçilen üç çözüm ile sıfır vektörü ve popülasyondaki en iyi çözüm kullanılmaktadır. Birebir ve örten eşleme kuralına göre rastgele eşleştirilen çözümler öncelikle çift noktalı çaprazlama ile çocuklar, çocuklar üzerinde takas operatörü uygulanması ile torunlar üretilmektedir. Çocuklar ve torunlar içerisinde en uygun çözüm seçilerek aday çözüm belirlenmektedir. İkili uzayda

çözümler oluşturulduktan sonra aday çözüm üretmek amacıyla kullanılan algoritma adım adım aşağıda verilmiştir.

1. X_i çözümünün komşuluğunda X_k ve X_s çözümlerini popülasyondan seç. Ayrıca popülasyondaki en iyi çözümü X_{best} 'i belirle ve X_{zeros} çözümünü oluştur ve 5 çözümünün kopyasını oluştur.
2. Birebir ve örten eşleme kuralına göre çözümler ve kopya çözümleri rastgele eşleştir.
3. Çift noktalı çaprazlama ile çocukları oluştur. Çocuklar içerisinde sıfır çözümü varsa belirli sayıda boyutunu 1 yap (sıfır çözümü genellikle amaç fonksiyon değerlendirmesinde kullanılmadığı için bu değişiklik yapılmaktadır ve asıl algoritmada sıfır çözümü ortaya çıkarsa rastgele seçilen 3 boyutu 1 olarak değiştirilmektedir)
4. Çocuklar üzerinde takas operatörü ile torunları oluştur.
5. Çocuklar ve torunlar içerisinde en uygun çözümü belirle ve aday çözüm olarak döndür.

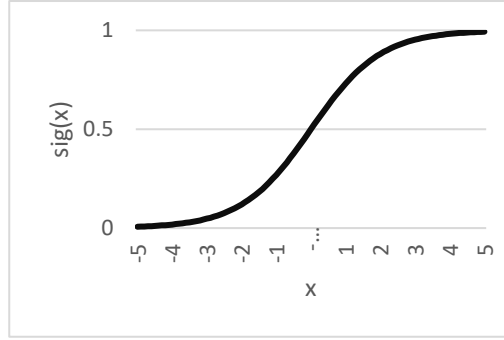
Yukarıdaki algoritmadan görüldüğü üzere bir aday çözüm oluşturabilmek amacıyla öncelikli olarak çaprazlama sonrasında 10, takas operatörü uygulandıktan sonra ise 10 olmak üzere 20 çözüm oluşturulmaktadır. Bu 20 çözümden en iyisi aday çözüm olarak belirlenmektedir.

3.8. Sigmoid İkili ABC Algoritması

Sigmoid fonksiyonu herhangi bir değeri $[0,1]$ arasına haritalamak için kullanılan bir fonksiyondur ve Denklem 3.20'de verilmiştir.

$$sig(x) = \frac{1}{1+e^{-x}} \quad (3.20)$$

Denklemin çıkışı $(0,1)$ arasında olduğundan dolayı 0,5 eşik değeri kullanılarak çıkış 0 veya 1 değerini dönüştürülebilir ve Denklem 3.21'de dönüşüm ve Şekil 3.3'te $sig(x)$ fonksiyon grafiği verilmiştir.



Şekil 3.3. Sigmoid fonksiyon grafiği

$$V_i^j = \begin{cases} 0, & \text{eğer } (r_{i,j} < sig(x)) \\ 1, & \text{değilse} \end{cases} \quad (3.21)$$

Denklemden V_i^j aday çözümü ve $r_{i,j}$ $[0,1]$ aralığında üretilmiş rastgele bir sayıyı temsil eder. Bu ikili ABC algoritmasından deneysel sonuçlarda ve karşılaştırmalarda SABC olarak bahsedilecektir.

3.9. Stigmerji-Xor Tabanlı İkili ABC Algoritması

Stigmerji kavramı ajanlar veya eylemleri ile ortam arasında dolaylı bir koordinasyonun varlığını tanımlamak için kullanılmaktadır (Wikipedia, 2021). Ajanın aksiyonu veya başka bir sebeple ortamın durumu değiştiğinde bu değişiklikten diğer ajanlar etkilenir ve ortam yeniden değişir. Bal arıları kovanını ve kovanın sürekli bakım (yavru bakımı, nem oranı, yiyecek toplama vb.) altında bulundurulmasını bir ortam olarak kabul edersek kovandaki nem oranı değişirse bazı arılar bu nem oranını ayarlamak için görevlerini değiştirir. Bu kavram, ABC algoritmasında ajanların karar değişkenlerini değiştirmesi ve değişiklikten diğer arıların da etkilenmesini sağlamak amacıyla (Kiran, 2021) çalışmada kullanılmıştır. Stigmerjiye dayalı davranış XABC algoritmasına, bu tezde SXABC olarak bahsedilecektir, eklenmiştir. SXABC algoritmasında XABC'deki aday çözüm oluşturma denklemiyle işçi arılar uzayı araştırırken aynı zamanda değişiklikler kaydedilmektedir. Kaydedilen bilgi gözcü arı aşamasında aşağıdaki denklemler ile aday çözüm oluşturmak amacıyla kullanılmıştır.

$$P_{01}^j = C_{01}^j / (C_{01}^j + C_{10}^j) \quad (3.22)$$

$$P_{10}^j = C_{10}^j / (C_{10}^j + C_{01}^j) \quad (3.23)$$

$$V_{i,j} = \begin{cases} 0, & \text{eğer } (r_{i,j} < P_{10}^j) \\ 1, & \text{değilse} \end{cases} \quad (3.24)$$

Denklemdaki C_{01}^j j. karar değişkeninin 0'dan 1'e geçişlerinin sayısı, C_{10}^j ilgili karar değişkeninin 1'den 0'a geçişlerinin sayısı, P_{01}^j ilgili karar değişkeninin 0'dan 1'e geçme oranını ve P_{10}^j ise 1'de 0'a geçme oranını göstermektedir ve $1 - P_{01}^j$ şeklinde de hesaplanabilir. Denklemlerdeki C terimleri sayaçtır ve 1 ile başlatılır. Ajanlar tarafından her başarılı aday çözüm oluşturmada (mevcut çözümden daha uygun çözüm oluşturulduğunda) 1 artırılır.

3.10. Aşırı Öğrenme Makinesi

Aşırı Öğrenme Makineleri, tek gizli katmanlı ileri beslemeli yapay sinir ağlarının (YSA) eğimi için geliştirilmiş bir öğrenme yaklaşımıdır (Huang ve ark., 2004; 2006). ELM'nin yapay sinir ağı eğitimi için temel varsayımı YSA mimarisinin giriş, tek gizli ve çıkış katmanından oluşmasıdır. ELM algoritmasında gradyan bazlı öğrenme algoritmalarından (örneğin geri yayımlı öğrenme algoritması) farklı olarak giriş katmanındaki ağırlıklar ve bias rastgele atanır ve çıkış katmanındaki ağırlıklar analitik olarak hesaplanmaktadır. Geleneksel ileri beslemeli bir YSA da ağı eğitilmesi genellikle iteratif algoritmalar ile gerçekleştirilirken, bu süreç ELM'de analitik bir denkleme dönüştürülmüş ve bu sebeple çok hızlı bir öğrenme yeteneği ELM'ye kazandırılmıştır. Ayrıca, ELM'de çıkış ağırlıkları kare matris olmaması durumunda (eğitim örnek sayısının gizli katmandaki nöron sayısına eşit olmadığı durumlarda) Moore-Penrose matrisi ile hesapladığından, yerel minimuma takılma veya optimum öğrenme parametresi belirlenmesi gibi performansı doğrudan etkileyen sorunları yoktur ve öğrenme kabiliyeti belirlenen varsayımlar altında daha yüksektir. ELM'de gizli katman çıkışı H ve gizli katman ile çıkış verisi arasındaki ağırlıklar ise W ile temsil edildiğinde aşağıdaki denklem ile çıkış hesaplanabilir.

$$Y = HW \quad (3.25)$$

Eğitim örneği sayısı N, gizli katman nöron sayısı M olarak alındığında Y, H ve W matrisleri sırasıyla $N \times 1$, $N \times M$, $M \times 1$ boyutuna sahip olurlar. ELM'de giriş ağırlıkları rastgele atandığından dolayı H doğrudan hesaplanabilir ve Y de eğitim çıkışları olarak bilindiğinden dolayı W'nin hesaplanabilmesi için H'nin kare matris olması zorunludur. H'nin kare matris olabilmesi için $N=M$ olması yani eğitim örnek sayısı ise gizli katman

nöron sayısının eşit olması gerekir. Eşit olmadığı durumda H matrisi Denklem 3.26 ile Moore-Penrose kaba tersi hesaplanır ve H^+ kare matristir.

$$H^+ = (H^T H)^{-1} H^T \quad (3.26)$$

Denklem 3.25 ve Denklem 3.26 birlikte ele alındığında W, Denklem 3.27 ile hesaplanabilir.

$$W = (H^+)^{-1} Y \quad (3.27)$$

Yukarıdaki hesaplamalar ile ELM'nin iteratif yöntemlerden neden daha hızlı olduğunu açıklamaktadır. Bu nedenle tez kapsamında çalışılan ABC tabanlı özellik seçmede amaç fonksiyon olarak kullanılan YSA'nın eğitiminde ELM algoritması kullanılmıştır.

3.11. Deneysel Çalışmada Kullanılan Veri Setleri

Deneysel çalışmalarda Kaliforniya Üniversitesi, Irvine Makine Öğrenmesi Deposu'ndan (UCI, 2021) alınan 5 veri seti kullanılmıştır. Bu veri setleri Wisconsin göğüs kanseri (Wisconsin breast cancer), boşanma (divorce), tepe vadisi (hillvalley), iyonosfer (ionosphere) ve madelon veri setleridir. Bu veri setlerinin tümü iki sınıflıdır ve özellik seçiminde problem türü sınıflandırma problemi olarak ele alınmıştır. Göğüs kanseri veri setinde 30 özellik, 569 örnek, iyonosfer veri setinde 34 özellik, 351 örnek, boşanma veri setinde 54 özellik, 170 örnek, tepe vadisi veri setinde 100 özellik, 606 örnek ve madelon veri setinde 500 özellik, 4400 örnek bulunmaktadır. Bu veri setleri ile ilgili detaylı bilgi (UCI, 2021) kaynağında bulunabileceğinden bu bölümde daha detaylı bilgilere yer verilmemiştir.

4. DENEYSEL SONUÇLAR

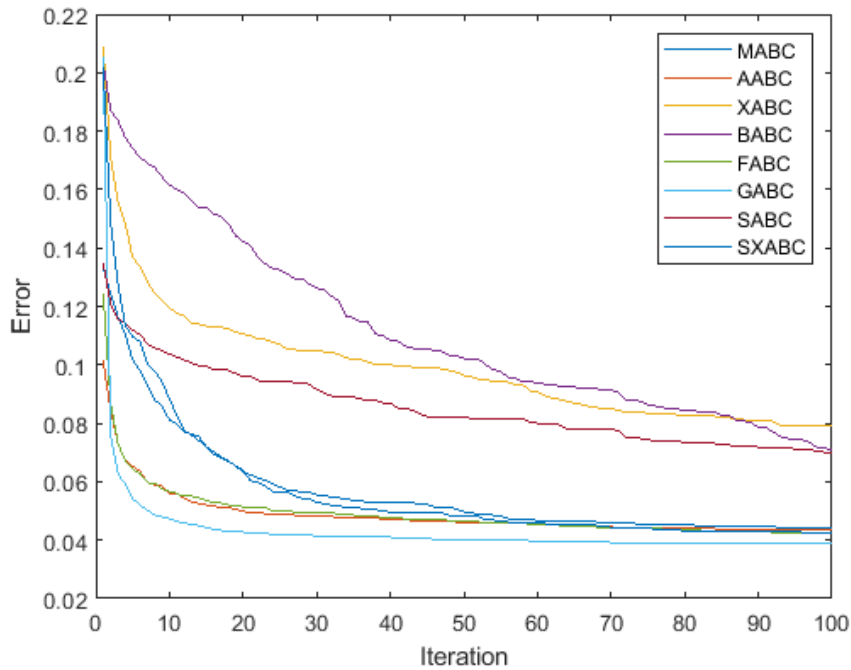
Tez çalışmasında kullanılan ikili ABC algoritmalarında durdurma kriteri olarak maksimum çevrim sayısı kullanılmıştır ve 100 olarak alınmıştır. Sadece GABC algoritmasındaki bir çevrimde bir yiyecek kaynağı etrafında birden fazla çözüm oluşturulduğu için çevrim sayısı yerine maksimum amaç fonksiyon değerlendirme sayısı durdurma kriteri olarak kullanılmıştır. Bu durumda popülasyon sayısının 40 olduğu kabul edildiğinde eşit bir karşılaştırma olması açısından durdurma kriteri $100 \times 40 = 4000$ olarak ayarlanmıştır. ABC algoritmasının kendine has parametresi *limit* değeridir. Limit değeri deneylerde 100 olarak alınmıştır. Yani bir kaynak etrafında 100 defa deneme sonunda daha kaliteli bir yiyecek kaynağı bulunamadıysa bu kaynağa atanmış olan işçi arı kâşif olarak değerlendirilmekte ve kâşif arı için yeni bir yiyecek kaynağı oluşturulmaktadır. Popülasyon sayısı literatürde genellikle 40 civarında seçilmiş ve ABC algoritması için arı popülasyon sayısının bu civarda olmasının makul sonuçlar elde etmek için yeterli olduğu görülmüştür (Karaboga ve Akay, 2009; Kiran ve Gündüz, 2013; Kiran ve ark., 2015). Bundan dolayı yiyecek kaynağı sayısı (ABC algoritmasında yiyecek kaynağı sayısı popülasyon boyutuna eşit seçilmektedir) 40 olarak seçilmiştir. Amaç fonksiyon olarak ELM algoritması ile eğitilmiş yapay sinir ağından elde edilen test hata oranı kullanılmıştır. ELM algoritması nöron aktivasyon fonksiyonu ve gizli katman nöron sayısı olmak üzere iki parametreye sahiptir. Gizli katman nöron sayısı 50 olarak alınmış ve aktivasyon fonksiyonu olarak tanjant sigmoid fonksiyonu kullanılmıştır. Ayrıca 10 katlı çapraz doğrulamanın ortalaması amaç fonksiyondan döndürülmektedir. Bu sayede tekil başarıdan ziyade bütün veri kümesi üzerindeki test başarısı aday çözümlerin değerlendirilmesinde kullanılmaktadır. Yukarıdaki şartlarda her bir yöntem 30 defa çalıştırılmıştır. Elde edilen sonuçlar test başarıları ve özellik sayıları olarak iki farklı tabloda ve algoritmaların yakınsama grafikleri aşağıda verilmiştir. Karşılaştırma tablolarının tümünde diğerlerine göre başarılı (üstün) olan algoritmayı kolaylıkla belirlemek amacıyla elde edilen değer koyu yazılmıştır.

Çizelge 4.1. Göğüs kanseri veri seti üzerinde elde edilen test sonuçları

Yöntem	Ortalama	En İyi	En Kötü	Standart Sapma	Sıra
MABC	0.0440	0.0368	0.0649	0.0057	5
AABC	0.0435	0.0386	0.0474	0.0023	4
XABC	0.0790	0.0421	0.1105	0.0269	8
BABC	0.0709	0.0369	0.1616	0.0441	7
FABC	0.0424	0.0386	0.0457	0.0017	2
GABC	0.0389	0.0352	0.0422	0.0023	1
SABC	0.0701	0.0420	0.1090	0.0153	6
SXABC	0.0425	0.0352	0.0526	0.0038	3
ELM	0.1530	0.1230	0.1969	0.0177	9

Çizelge 4.2. Göğüs kanseri veri seti üzerinde seçilen özellik sayılarına ait sonuçlar

Yöntem	Ortalama	En Az	En Çok	Standart Sapma
MABC	10.00	7.00	13.00	1.86
AABC	7.37	2.00	15.00	3.37
XABC	15.67	10.00	21.00	3.36
BABC	11.87	8.00	17.00	2.52
FABC	4.77	2.00	16.00	3.08
GABC	6.30	2.00	12.00	3.14
SABC	14.57	11.00	22.00	2.56
SXABC	11.23	8.00	15.00	1.79



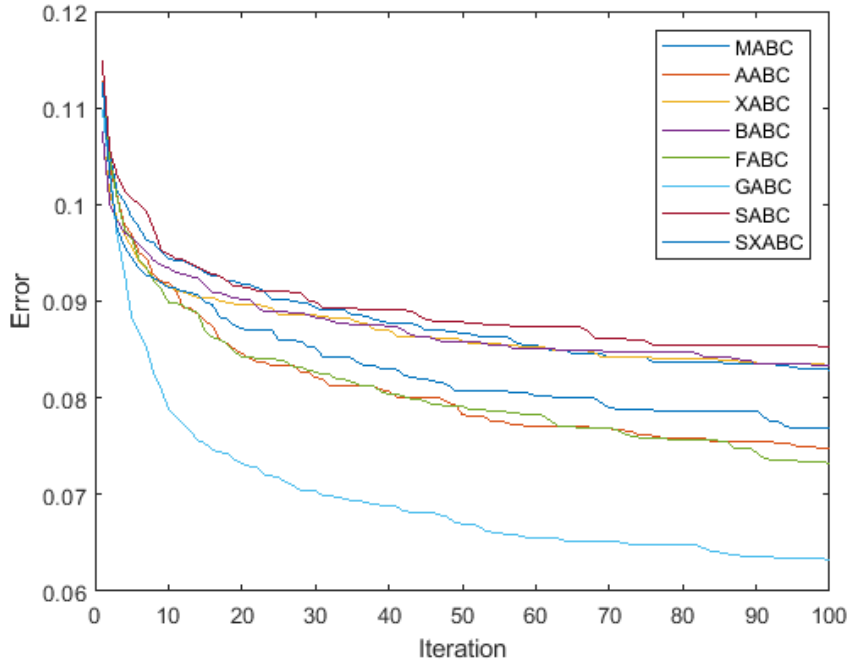
Şekil 4.1. Göğüs kanseri veri seti özellik seçiminde algoritmaların yakınsama grafiği

Çizelge 4.3. İyonosfer veri seti üzerinde elde edilen test sonuçları

Yöntem	Ortalama	En İyi	En Kötü	Standart Sapma	Sıra
MABC	0.0825	0.0737	0.0904	0.0049	5
AABC	0.0748	0.0626	0.0858	0.0055	3
XABC	0.0833	0.0738	0.0886	0.0044	6
BABC	0.0834	0.0743	0.0884	0.0032	7
FABC	0.0732	0.0596	0.0799	0.0052	2
GABC	0.0627	0.0488	0.0740	0.0057	1
SABC	0.0853	0.0677	0.0935	0.0053	8
SXABC	0.0768	0.0682	0.0856	0.0050	4
ELM	0.1421	0.1142	0.1623	0.0135	9

Çizelge 4.4. İyonosfer veri seti üzerinde seçilen özellik sayılarına ait sonuçlar

Yöntem	Ortalama	En Az	En Çok	Standart Sapma
MABC	15.10	6.00	24.00	3.72
AABC	7.33	3.00	17.00	2.78
XABC	18.63	15.00	21.00	1.75
BABC	18.00	14.00	20.00	1.62
FABC	7.07	3.00	12.00	2.21
GABC	5.60	3.00	13.00	2.36
SABC	20.70	15.00	28.00	3.13
SXABC	14.27	10.00	20.00	3.08



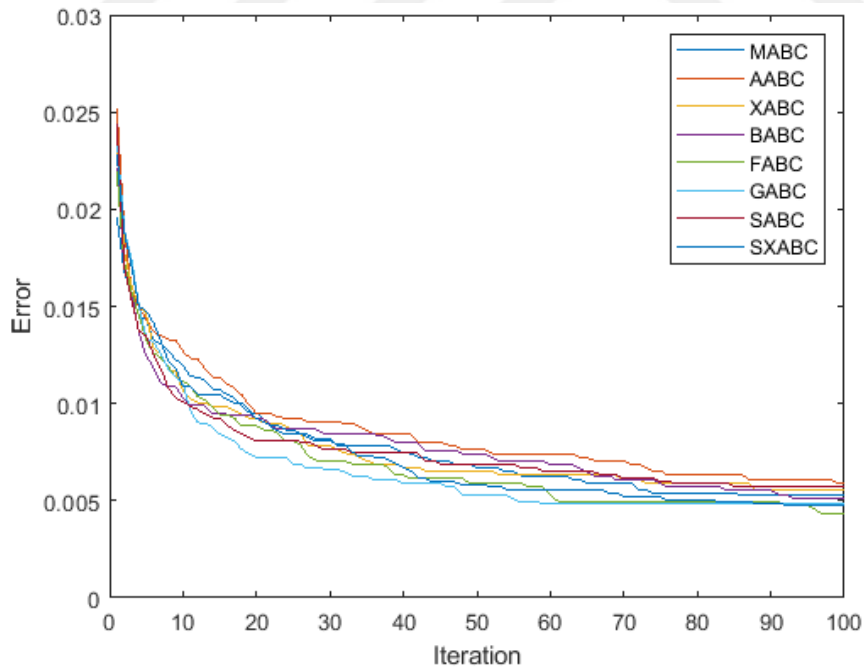
Şekil 4.2. İyonosfer veri seti özellik seçiminde algoritmaların yakınsama grafiği

Çizelge 4.5. Boşanma veri seti üzerinde elde edilen test sonuçları

Yöntem	Ortalama	En İyi	En Kötü	Standart Sapma	Sıra
MABC	0.0053	0.0000	0.0111	0.0021	5
AABC	0.0059	0.0000	0.0114	0.0027	8
XABC	0.0055	0.0000	0.0111	0.0023	6
BABC	0.0051	0.0000	0.0114	0.0034	4
FABC	0.0043	0.0000	0.0059	0.0024	1
GABC	0.0047	0.0000	0.0111	0.0030	2
SABC	0.0057	0.0000	0.0111	0.0021	7
SXABC	0.0048	0.0000	0.0062	0.0022	3
ELM	0.0509	0.0232	0.0829	0.0140	9

Çizelge 4.6. Boşanma veri seti üzerinde seçilen özellik sayılarına ait sonuçlar

Yöntem	Ortalama	En Az	En Çok	Standart Sapma
MABC	25.93	19.00	33.00	3.71
AABC	24.83	3.00	54.00	15.92
XABC	34.57	32.00	37.00	1.28
BABC	33.33	29.00	35.00	1.27
FABC	19.83	2.00	46.00	12.35
GABC	11.60	2.00	38.00	11.25
SABC	34.57	27.00	44.00	3.85
SXABC	30.53	24.00	37.00	3.18



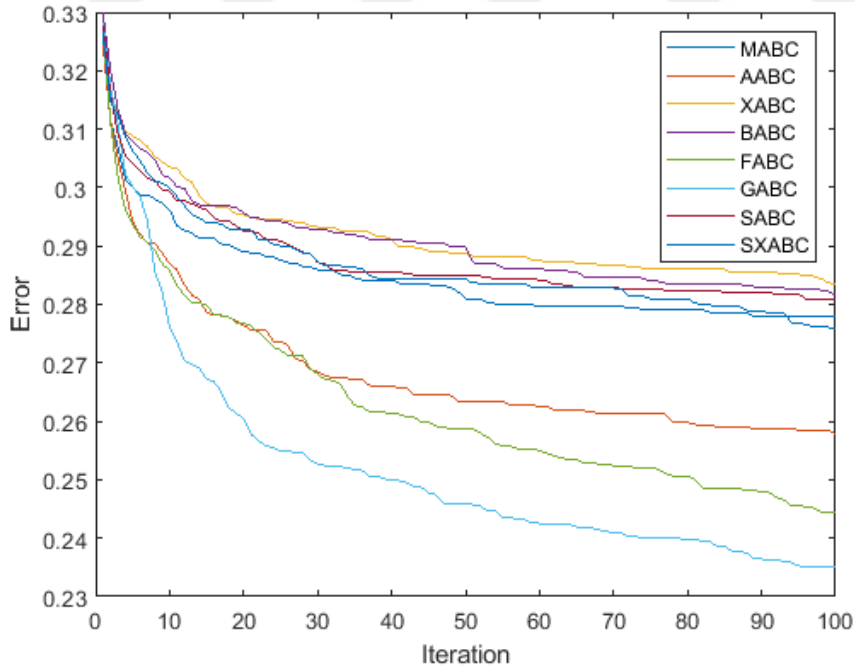
Şekil 4.3. Boşanma veri seti özellik seçiminde algoritmaların yakınsama grafiği

Çizelge 4.7. Tepe vadisi veri seti üzerinde elde edilen test sonuçları

Yöntem	Ortalama	En İyi	En Kötü	Standart Sapma	Sıra
MABC	0.2779	0.2574	0.2874	0.0080	5
AABC	0.2577	0.2328	0.2854	0.0129	3
XABC	0.2835	0.2674	0.2987	0.0071	8
BABC	0.2816	0.2598	0.2985	0.0091	7
FABC	0.2444	0.1896	0.2643	0.0153	2
GABC	0.2336	0.1962	0.2626	0.0152	1
SABC	0.2809	0.2671	0.2920	0.0066	6
SXABC	0.2761	0.2478	0.2923	0.0107	4
ELM	0.3694	0.3331	0.4323	0.0243	9

Çizelge 4.8. Tepe vadisi veri seti üzerinde seçilen özellik sayılarına ait sonuçlar

Yöntem	Ortalama	En Az	En Çok	Standart Sapma
MABC	47.93	36.00	60.00	5.60
AABC	15.80	4.00	34.00	8.36
XABC	49.90	48.00	52.00	1.12
BABC	49.20	46.00	52.00	1.42
FABC	9.33	4.00	25.00	4.77
GABC	18.97	3.00	37.00	10.65
SABC	64.87	54.00	77.00	5.00
SXABC	49.37	40.00	59.00	4.29



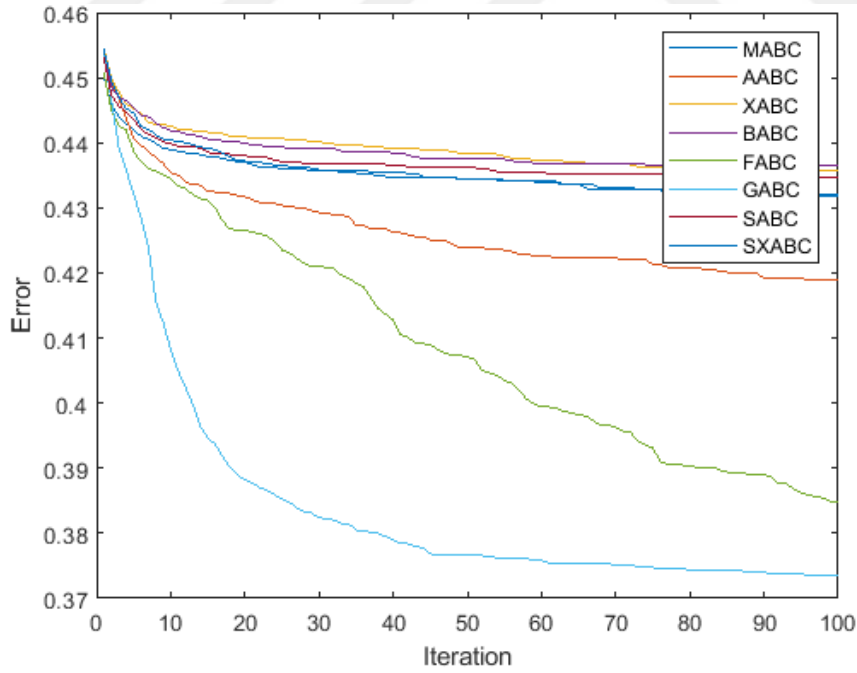
Şekil 4.4. Tepe vadisi veri seti özellik seçiminde algoritmaların yakınsama grafiği

Çizelge 4.9. Madelon veri seti üzerinde elde edilen test sonuçları

Yöntem	Ortalama	En İyi	En Kötü	Standart Sapma	Sıra
MABC	0.4320	0.4227	0.4385	0.0034	5
AABC	0.4189	0.3965	0.4346	0.0105	3
XABC	0.4358	0.4277	0.4419	0.0035	7
BABC	0.4364	0.4300	0.4408	0.0027	8
FABC	0.3847	0.3669	0.4069	0.0098	2
GABC	0.3727	0.3481	0.3835	0.0086	1
SABC	0.4346	0.4285	0.4396	0.0029	6
SXABC	0.4318	0.4196	0.4381	0.0039	4
ELM	0.4722	0.4458	0.4881	0.0090	9

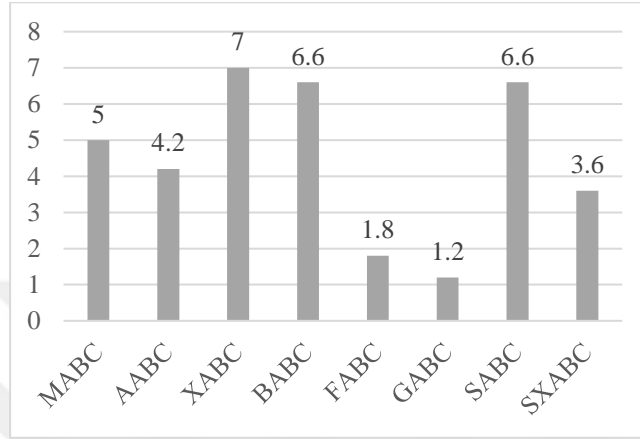
Çizelge 4.10. Madelon veri seti üzerinde seçilen özellik sayılarına ait sonuçlar

Yöntem	Ortalama	En Az	En Çok	Standart Sapma
MABC	247.17	230.00	280.00	11.08
AABC	51.73	7.00	346.00	69.85
XABC	250.30	247.00	253.00	1.21
BABC	249.70	247.00	252.00	1.34
FABC	3.77	2.00	15.00	2.74
GABC	2.87	2.00	7.00	1.48
SABC	324.87	295.00	351.00	12.38
SXABC	248.23	229.00	263.00	7.56



Şekil 4.5. Madelon veri seti özellik seçiminde algoritmaların yakınsama grafiği

Çizelgelerden görüldüğü üzere özellik seçildiği taktirde sınıflandırma doğruluğu bütün veri setlerinde özellik seçilmeden yapılan sınıflandırmaya göre artmıştır. Bununla birlikte ikili ABC algoritmaları kendi içlerinde karşılaştırıldığında GBABC algoritmasının performansının diğer ABC varyantlarından daha iyi olduğu görülmektedir. Bu durum ortalama sıra tabanlı bir metrikle Şekil 4.6'dan da kolaylıkla görülebilmektedir.



Şekil 4.6. İkili ABC varyantlarının ortalama sıra tabanlı karşılaştırması

Sınıflandırma algoritmalarında bir diğer karşılaştırma kriteri eğitim başarılarıdır. Her ne kadar özellik seçiminde bu tez kapsamında amacımız ikili ABC algoritmaları kullanarak daha az özellikli test hatasının minimizasyonu olsa da eğitim hatasının da karşılaştırılması sıklıkla literatürde görülmektedir. Bundan dolayı yöntemlerin eğitim hataları da Çizelge 4.11'de verilmiştir.

Çizelge 4.11. Yöntemlerin veri setleri üzerindeki ortalama eğitim hataları

Yöntem	Göğüs Kanseri	İyonosfer	Boşanma	Tepe Vadisi	Madelon
MABC	0.0389	0.0663	0.0019	0.3182	0.4574
AABC	0.0392	0.0556	0.0024	0.2991	0.4358
XABC	0.1092	0.0685	0.0029	0.3297	0.4620
BABC	0.0843	0.0687	0.0030	0.3274	0.4641
FABC	0.0396	0.0554	0.0019	0.2872	0.3955
GABC	0.0384	0.0533	0.0042	0.2988	0.3901
SABC	0.0568	0.0681	0.0028	0.3325	0.4612
SXABC	0.0373	0.0616	0.0026	0.3309	0.4587
ELM	0.1636	0.0765	0.0035	0.3287	0.4624

Yöntemler ile elde edilen özellik kümeleri ile yapılan eğitimde eğitim hatası da test sonuçlarına benzer şekilde karşımıza çıkmıştır. En yüksek eğitim başarısı göğüs kanseri veri setinde SXABC %96,27, iyonosfer veri setinde GABC yöntemi %94,67, boşanma veri setinde MABC ve FABC %99,81, tepe vadisi veri setinde FABC %71,28

ve madelon veri setinde GABC yöntemi %60,99 olarak elde etmiştir. Buradan da görüleceği üzere özellik sayısının artması çözüm uzayını üstel şekilde artırdığı için yöntemlerin başarısı görece düşmektedir. Bu durum test başarılarında da karşımıza çıkmaktadır.

Eğitim hatalarının tümü tek bir tabloda verilmiş ve kolaylıkla karşılaştırma yapılabilmiştir. Benzer bir karşılaştırma yapılabilmesi amacıyla test hataları da Çizelge 4.12’de sunulmuştur.

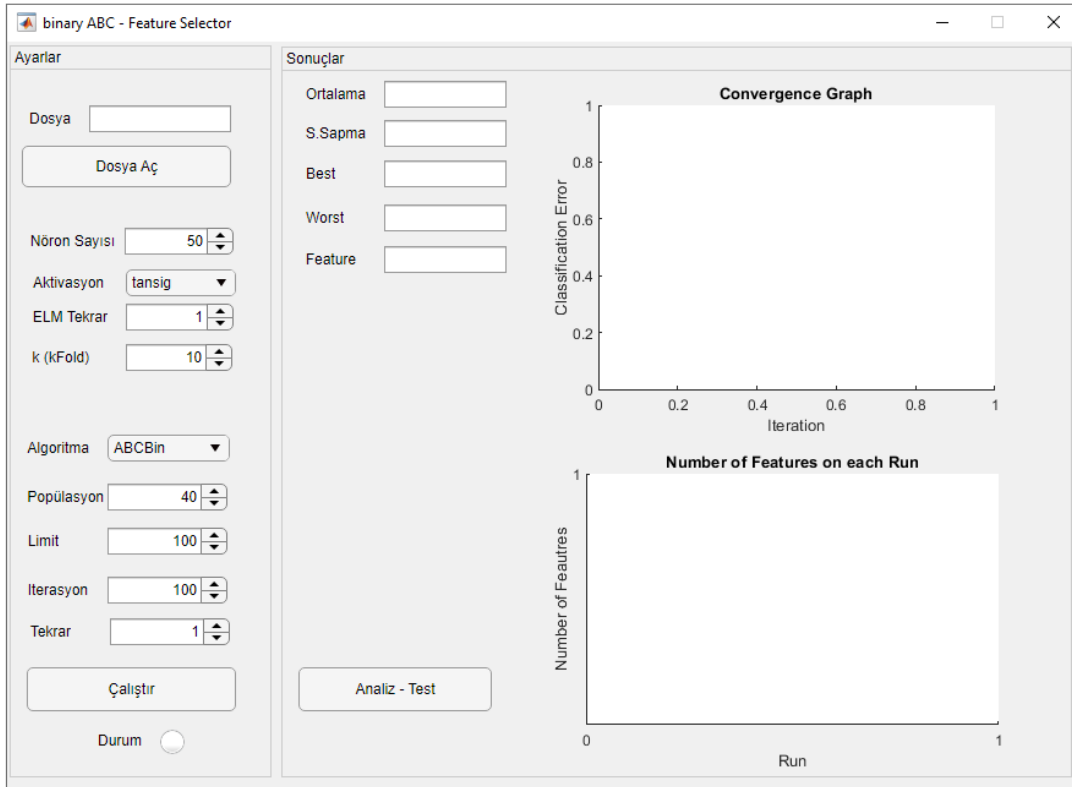
Çizelge 4.12. Yöntemlerin veri setleri üzerindeki ortalama test hataları

Yöntem	Göğüs Kanseri	İyonosfer	Boşanma	Tepe Vadisi	Madelon
MABC	0.0440	0.0825	0.0053	0.2779	0.4320
AABC	0.0435	0.0748	0.0059	0.2577	0.4189
XABC	0.0790	0.0833	0.0055	0.2835	0.4358
BABC	0.0709	0.0834	0.0051	0.2816	0.4364
FABC	0.0424	0.0732	0.0043	0.2444	0.3847
GABC	0.0389	0.0627	0.0047	0.2336	0.3727
SABC	0.0701	0.0853	0.0057	0.2809	0.4346
SXABC	0.0425	0.0768	0.0048	0.2761	0.4318
ELM	0.1530	0.1421	0.0509	0.3694	0.4722

Çizelge 4.12’den de görüldüğü üzere GABC algoritmasının test performansı diğer ikili ABC algoritmalarına göre daha iyidir. Sadece Boşanma veri setinde FABC algoritması diğer algoritmalara göre biraz daha iyi sonuçlar üretmiştir. Bununla birlikte özellik sayısının artması yöntemlerin tümü için çözüm uzayının üstel şekilde artmasından dolayı önemli bir sorun olarak görünmektedir.

5. GELİŞTİRİLEN ÖRNEK UYGULAMA

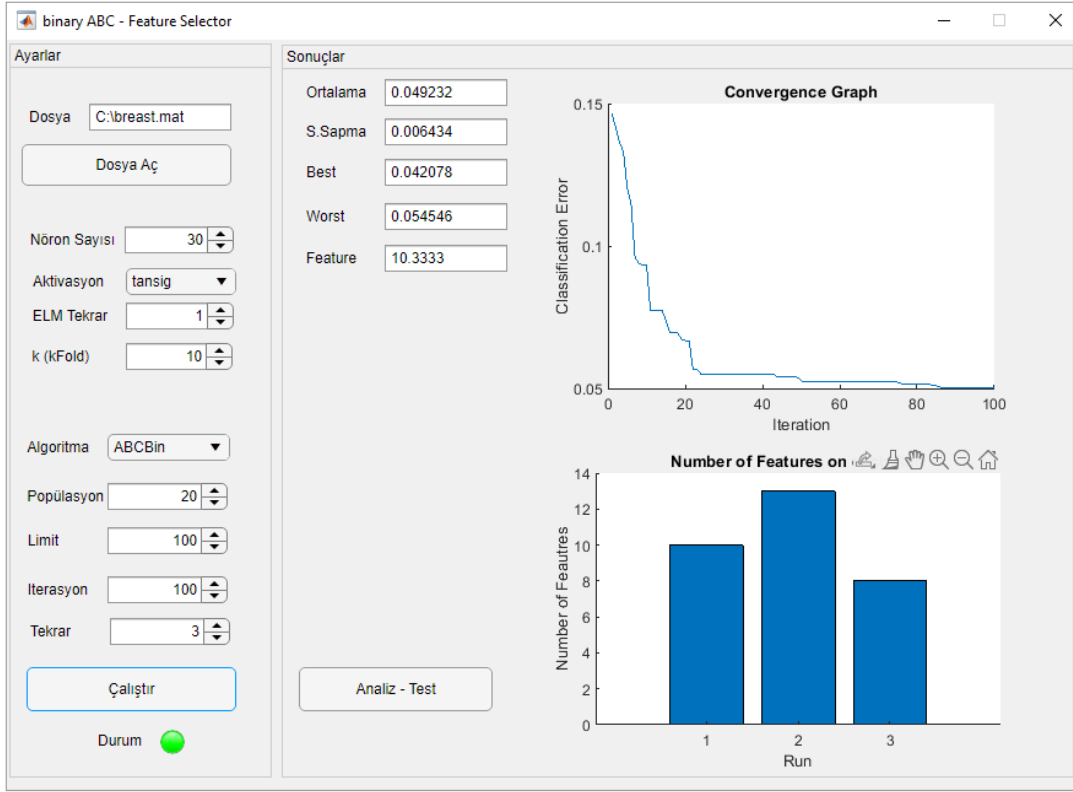
Önerilen yöntemlerin farklı arařtırmacılar ve uygulayıcılar tarafından kullanılabilmesi için örnek bir uygulama gerçekleştirilmiştir. Uygulamanın arayüzleri Matlab 2021a appDesigner kullanılarak, metotlar ise Matlab sınıfları olarak oluşturulmuştur. Verilerin yüklenebilmesi için mat uzantılı dosyalar kullanılmaktadır ve ilk sütun satır etiketlerini içermektedir. Uygulama çalıştırıldığında Şekil 5.1 ile verilen ekran karşımıza çıkmaktadır. Bu ekranda öncelikle eğitim amacıyla kullanılacak veri seti “Dosya Aç” ile yüklenmelidir. Eğitim verisi yüklendikten sonra ilk bölümde görüldüğü üzere ELM parametreleri ayarlanmalıdır. Varsayılan olarak gizli katman nöron sayısı 50, aktivasyon fonksiyonu tanjant sigmoid, ELM tekrar sayısı 1 ve 10 katlı çapraz doğrulama ayarlanmıştır. Uygulama ayarları kullanıcı tarafından özelleştirildikten sonra özellik seçimi için kullanılacak ikili ABC yöntemi seçilmelidir. Buradaki seçenekler de varsayılan olarak tezde verilen değerler olarak atanmıştır.



Şekil 5.1. Örnek uygulama açılış ekranı

Şekil 5.1 ile gösterilen arayüz üzerinde hem amaç fonksiyon hem de özellik seçici algoritmanın parametreleri ayarlandıktan sonra uygulamanın çalıştırılması ve sonuçların raporlanması Şekil 5.2’de gösterilmiştir. Bu ekranda özellik seçici yöntemin

koşuturulma sayısına bağılı olarak ortalama, en iyi ve en kötü sınıflandırma performansı ile standart sapma raporlanmaktadır. Ayrıca her çalıştırmada seçilen özellik kümesindeki özellik sayısı ile özellik seçici algoritmanın yakınsama grafikleri de ekranın sağ tarafında gösterilmektedir.



Şekil 5.2. Örnek uygulama ile özellik seçimi sonuçlarının gösterimi

Şekil 5.2 ile örnek uygulama çalıştırıldıktan sonra özellik seçimi tamamlanmış olmaktadır. Kullanıcıların özellik seçimi yaptıktan sonra verilerini ELM ile sınıflandırabilmeleri amacıyla ikinci bir arayüz daha bulunmaktadır. Bu arayüze Analiz-Test butonuna tıklanarak ulaşılabilir ve arayüzün ekran görüntüsü Şekil 5.3 ile verilmiştir. Bu arayüze geçilirken arka plan değişkenleri ile özellik kümesi, en yüksek performansın elde edildiği seçili özellik alt kümesi, veri kümesi ve etiket kümesi de aktarılmaktadır. Görselleştirme amacıyla Özellikleri Getir butonu tıklanarak özellik kümesinin alt kümesi olan seçilen özellikler yeşil arka planda 1, seçilmeyen özellikler kırmızı arka planda 0 olarak gösterilmektedir.

The image shows a software application window titled "Feature Selector - Analiz". The window contains a sidebar with several buttons and input fields. The buttons are "Özellikleri Getir", "Test Verisi Yükle", "Test", and "Eğitim-Test". The input fields are "Nöron Sayısı" (Neuron Count) with a value of 50, "Aktivasyon" (Activation) with a dropdown menu showing "tansig", "ELM Tekrar" (ELM Repeat) with a value of 1, and "k (kFold)" with a value of 10. Below these are four performance metrics: "Eğitim Başarısı" (Training Success) 0, "Test Başarısı" (Test Success) 0, "Eğitim Zamanı" (Training Time) 0, and "Test Zamanı" (Test Time) 0. The main area of the window is currently empty.

Şekil 5.3. Özellik seçici uygulama ile analiz

Etiket bilgisi olmayan test verileri Test Verisi Yükle butonu ile test edilmeden önce yüklenmelidir. Test verisi özellik sayısı ile özellik kümesi eleman sayısı karşılaştırılmakta test verisinde fazladan etiket bilgisi varsa ilk sütunu silinmektedir. Özellikler gösterildikten ve test verisi yüklendikten sonraki görünüm Şekil 5.4 ile verilmiştir.

Feature Selector - Analiz										
	1	2	3	4	5	6	7	8	9	10
Ozellikleri Getir	0	0	0	0	0	0	1	0	0	0
Test Verisi Yükle	1	2	3	4	5	6	7	8	9	10
Nöron Sayısı: 50	17.9900	10.3800	122.8000	1001	0.1184	0.2776	0.3001	0.1471	0.2419	
Aktivasyon: tansig	20.5700	17.7700	132.9000	1326	0.0847	0.0786	0.0869	0.0702	0.1812	
ELM Tekrar: 1	19.6900	21.2500	130	1203	0.1096	0.1599	0.1974	0.1279	0.2069	
k (kFold): 10	11.4200	20.3800	77.5800	386.1000	0.1425	0.2839	0.2414	0.1052	0.2597	
Test	20.2900	14.3400	135.1000	1297	0.1003	0.1328	0.1980	0.1043	0.1809	
	12.4500	15.7000	82.5700	477.1000	0.1278	0.1700	0.1578	0.0809	0.2087	
Eğitim-Test	18.2500	19.9800	119.6000	1040	0.0946	0.1090	0.1127	0.0740	0.1794	
	13.7100	20.8300	90.2000	577.9000	0.1189	0.1645	0.0937	0.0599	0.2196	
Eğitim Başarısı: 0	13	21.8200	87.5000	519.8000	0.1273	0.1932	0.1859	0.0935	0.2350	
Test Başarısı: 0	12.4600	24.0400	83.9700	475.9000	0.1186	0.2396	0.2273	0.0854	0.2030	
Eğitim Zamanı: 0	16.0200	23.2400	102.7000	797.8000	0.0821	0.0667	0.0330	0.0332	0.1528	
Test Zamanı: 0	15.7800	17.8900	103.6000	781	0.0971	0.1292	0.0995	0.0661	0.1842	
	19.1700	24.8000	132.4000	1123	0.0974	0.2458	0.2065	0.1118	0.2397	
	15.8500	23.9500	103.7000	782.7000	0.0840	0.1002	0.0994	0.0536	0.1847	
	13.7300	22.6100	93.6000	578.3000	0.1131	0.2293	0.2128	0.0803	0.2069	
	14.5400	27.5400	96.7300	658.8000	0.1139	0.1595	0.1639	0.0736	0.2303	
	14.6800	20.1300	94.7400	684.5000	0.0987	0.0720	0.0740	0.0526	0.1586	
	16.1300	20.6800	108.1000	798.8000	0.1170	0.2022	0.1722	0.1028	0.2164	
	19.8100	22.1500	130	1260	0.0983	0.1027	0.1479	0.0950	0.1582	
	13.5400	14.3600	87.4600	566.3000	0.0978	0.0813	0.0666	0.0478	0.1885	

Şekil 5.4. Özellikler gösterildikten ve test verisi yüklendikten sonraki ekran görüntüsü

Şekil 5.4'te verilen ekrandaki matrizen görüldüğü üzere verilerde etiket bulunmamaktadır. ELM parametreleri ayarlandıktan sonra Test butonuna tıklandığında önceki ekrandan aktarılan veriler ve seçili özellik alt kümesi ile ELM eğitimi yapılmaktadır. Eğitim tamamlandıktan sonra özellik alt kümesi kullanılarak test kümesinin sütun boyutu indirgenmekte, ELM ile indirgenmiş veri kümesi sınıflandırılmakta ve sınıf sonuçları mevcut matris gösteriminde başa eklenmektedir. Bu durum Şekil 5.5'te görülmektedir.

Feature Selector - Analiz										
Özellikler Getir	1	2	3	4	5	6	7	8	9	10
	0	0	0	0	0	0	1	0	0	0
Test Verisi Yükle	1	2	3	4	5	6	7	8	9	10
	1	17.9900	10.3800	122.8000	1001	0.1184	0.2776	0.3001	0.1471	
	1	20.5700	17.7700	132.9000	1326	0.0847	0.0786	0.0869	0.0702	
	1	19.6900	21.2500	130	1203	0.1096	0.1599	0.1974	0.1279	
	1	11.4200	20.3800	77.5800	386.1000	0.1425	0.2839	0.2414	0.1052	
	1	20.2900	14.3400	135.1000	1297	0.1003	0.1328	0.1980	0.1043	
	1	12.4500	15.7000	82.5700	477.1000	0.1278	0.1700	0.1578	0.0809	
	1	18.2500	19.9800	119.6000	1040	0.0946	0.1090	0.1127	0.0740	
	1	13.7100	20.8300	90.2000	577.9000	0.1189	0.1645	0.0937	0.0599	
Test	1	13	21.8200	87.5000	519.8000	0.1273	0.1932	0.1859	0.0935	
	1	12.4600	24.0400	83.9700	475.9000	0.1186	0.2396	0.2273	0.0854	
	1	16.0200	23.2400	102.7000	797.8000	0.0821	0.0667	0.0330	0.0332	
	1	15.7800	17.8900	103.6000	781	0.0971	0.1292	0.0995	0.0661	
	1	19.1700	24.8000	132.4000	1123	0.0974	0.2458	0.2065	0.1118	
Eğitim-Test	0	15.8500	23.9500	103.7000	782.7000	0.0840	0.1002	0.0994	0.0536	
	1	13.7300	22.6100	93.6000	578.3000	0.1131	0.2293	0.2128	0.0803	
	1	14.5400	27.5400	96.7300	658.8000	0.1139	0.1595	0.1639	0.0736	
	1	14.6800	20.1300	94.7400	684.5000	0.0987	0.0720	0.0740	0.0526	
	1	16.1300	20.6800	108.1000	798.8000	0.1170	0.2022	0.1722	0.1028	
	1	19.8100	22.1500	130	1260	0.0983	0.1027	0.1479	0.0950	
	0	13.5400	14.3600	87.4600	566.3000	0.0978	0.0813	0.0666	0.0478	

Şekil 5.5. Etiketlenmiş test verileri

Ekranın sol alt bölümünde görülen Eğitim-Test butonu ise seçili özellik alt kümesi ile farklı ELM parametrelerinin denenmesi amacıyla eklenmiştir.

6. SONUÇLAR VE ÖNERİLER

Veri madenciliği ve makine öğrenmesi alanında verilerin hem hacmi hem de boyutu yöntemlerin performansını etkilemektedir. Az sayıda veri eğitimi etkilerken yüksek boyutlu veri gereksiz özellikler barındırabileceğinden veya sınıflandırıcılar yüksek boyutlu verilerin sınıflandırılmasında daha az başarılı sonuçlar üretebileceğinden dolayı sınıflandırma başarısını düşürmekte ve ayrıca işlem sürelerini uzatmaktadır. Hem yüksek sınıflandırma doğruluğu hem de daha düşük işlem sürelerini elde etmek amacıyla özellik seçimi bu alanlarda popüler bir konu haline gelmiştir. Özellik seçimi bu tez çalışmasında ikili optimizasyon problemi olarak ele alınmış ve özelliklerin seçilmesi ve seçilmemesini temsil etmek amacıyla $\{0,1\}$ kümesi kullanılmıştır: 0 bir özelliğin seçilmediğini ve 1 ise seçildiğini göstermektedir. Toplam bir veri kümesinde n adet özellik olduğu kabul edildiğinde boş olmayan özellik alt küme sayısı $2^n - 1$ ile kolaylıkla bulunabilir. Toplam küme sayısı kolaylıkla bulunabilmesine karşın özellik sayısına bağlı olarak alt küme sayısı üstel bir şekilde arttığından dolayı maksimum sınıflandırma doğruluğunu veren alt kümenin bulunması NP karmaşıklığına sahip bir problem olarak karşımıza çıkmaktadır. Bu nedenle problemin çözümü için kesin metotlardan ziyade günümüzde evrimsel hesaplama ve sürü zekâsı algoritmaları popüler olmuştur. Bu tez çalışmasında yukarıda sayılan sebeplerden dolayı sürü zekâsı alanının önemli bir üyesi olan yapay arı kolonisi – ABC algoritması ile özellik seçimi çalışılmıştır. Özellik seçimi amacıyla 8 farklı ikili ABC algoritması uygulanmış ve amaç fonksiyon olarak aşırı öğrenme makinesi – ELM ile eğitilen yapay sinir ağı kullanılmıştır. Bu bağlamda SXABC algoritması ilk defa özellik seçimine uygulanmakta, diğer ikili ABC algoritmaları ise ELM ile ilk defa beraber kullanılmaktadır.

6.1 Sonuçlar

Özellik seçimi problemi için kullanılan veri setleri üzerinde elde edilen deneysel sonuçlardan özellik seçimine ve algoritmaların performanslarına dair aşağıdaki bulgular çıkarılmıştır.

1. Özellik seçimi yapıldığında bütün veri kümelerinde tüm özelliklerin seçili olduğu duruma göre daha yüksek bir sınıflandırma doğruluğu elde edildiği deneylerde görülmüştür.
2. Genetik operatörler kullanılarak geliştirilmiş olan ikili ABC algoritması tez kapsamında ele alınan diğer ikili ABC varyantlarından daha başarılı sonuçlar

ürettiği görülmüştür. Farklılık tabanlı ABC algoritması sıralamada ikinci, SXABC algoritması bu bağlamda üçüncü sırada yer almaktadır.

3. Yakınsama grafiklerinden görüldüğü üzere algoritmaların performanslarının belirli noktalardan itibaren düzleştiği ve ikili ABC parametrelerinin makul olarak ayarlandığı sonucu çıkarılmıştır.
4. Veri setleri üzerindeki ortalama başarılar tabanlı karşılaştırmaya göre en iyi yöntemin seçtiği ortalama özellik sayıları göğüs kanseri veri setinde 6,3 (30), iyonosfer veri setinde 5,6 (34), boşanma veri setinde 19,83 (54), tepe vadisi veri setinde 18,97 (100) ve madelon veri setinde 2,87 (500) olduğu görülmüştür.
5. Daha az özellik ile test ve eğitim başarılarının yükseltilebildiği sonucuna varılmıştır.

6.2. Öneriler

Tez kapsamında yapılan deneysel çalışmaları, evrimsel hesaplama veya sürü zekâsı literatürünü ve sınıflandırma algoritmalarını birlikte ele aldığımızda gelecekte yapılabilecek çalışmalara dair öneriler aşağıda maddeler halinde verilmiştir.

1. Diğer evrimsel hesaplama ve sürü zekâsı yöntemlerinin de özellik seçimi amacıyla kullanılabilir ve kapsamlı bir analiz ve değerlendirme yapılabilir.
2. Tez kapsamındaki ikili ABC algoritmaları farklı ikili optimizasyon problemlerini çözmek için uyarlanabilir ve kullanılabilir.
3. Amaç fonksiyon hesaplanmasında kullanılan sınıflandırıcı algoritmalar değiştirilerek başarı değerlendirilmesi yapılabilir.
4. Kümeleme problemleri üzerinde ikili ABC algoritmalarının performansları incelenebilir.
5. İkili sınıflandırma problemlerinin yanı sıra çok sınıflı veri setleri üzerinde ikili ABC ve diğer evrimsel hesaplama algoritmalarının performansları incelenebilir.
6. İkili ABC algoritmaları ile sınıflandırma problemi çok amaçlı (daha az özellik sayısı, daha yüksek eğitim ve/veya test başarısı) olarak çözülebilir.
7. Aynı anda hem ELM parametrelerinin optimize edilmesi hem de özellik sayısının azaltılması konusunda araştırmalar yapılabilir.

KAYNAKLAR

- Abu-Mouti, F. ve El-Hawary, M., 2009, Modified artificial bee colony algorithm for optimal distributed generation sizing and allocation in distribution systems, *2009 IEEE Electrical Power & Energy Conference (EPEC)*, 1-9.
- Agrawal, V. ve Chandra, S., 2015, Feature selection using Artificial Bee Colony algorithm for medical image classification, *2015 Eighth International Conference on Contemporary Computing (IC3)*, 171-176.
- Akay, B. ve Karaboga, D., 2015, A survey on the applications of artificial bee colony in signal, image, and video processing, *Signal, Image and Video Processing*, 9 (4), 967-990.
- Akay, B., Karaboga, D., Gorkemli, B. ve Kaya, E., 2021, A survey on the Artificial Bee Colony algorithm variants for binary, integer and mixed integer programming problems, *Applied Soft Computing*, 106.
- Akila, M., Kumar, V. S., Anusheela, N. ve Sugumar, K., 2012, A novel feature subset selection algorithm using artificial bee colony in keystroke dynamics, *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*, 813-820.
- Arslan, S. ve Ozturk, C., 2019, Feature Selection for Classification with Artificial Bee Colony Programming, In: *Swarm Intelligence-Recent Advances, New Perspectives and Applications*, Eds: IntechOpen, p.
- Badem, H., Turkusagi, D., Caliskan, A. ve Çil, Z. A., 2019, Feature selection based on artificial bee colony for Parkinson disease diagnosis, *2019 Medical Technologies Congress (TIPTEKNO)*, 1-4.
- Benala, T. R., Villa, S. H., Jampala, S. D. ve Konathala, B., 2009, A novel approach to image edge enhancement using artificial bee colony optimization algorithm for hybridized smoothening filters, *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 1071-1076.
- Bindu, M. ve Sabu, M., 2020, A Hybrid Feature Selection Approach Using Artificial Bee Colony and Genetic Algorithm, *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, 211-216.
- Bindu, M. ve Sabu, M., 2021, GABC: A Hybrid Approach for Feature Selection Using Artificial Bee Colony and Genetic Operators, *International Journal of Swarm Intelligence Research (IJSIR)*, 12 (3), 78-95.
- Chahkandi, V., Yaghoobi, M. ve Veisi, G., 2013, Feature selection with chaotic hybrid artificial bee colony algorithm based on fuzzy (CHABCF), *Journal of Soft Computing and Applications*, 2013 (1), 1-8.
- Chen, B., Hong, J. ve Wang, Y., 1997, The minimum feature subset selection problem, *Journal of Computer Science and Technology*, 12 (2), 145-153.

- Chen, G., Zhang, X., Wang, Z. J. ve Li, F., 2015, An enhanced artificial bee colony-based support vector machine for image-based fault detection, *Mathematical Problems in Engineering*, 2015, 1-12.
- Chu, X., Li, S., Gao, D., Zhao, W., Cui, J. ve Huang, L., 2020a, A Binary Superior Tracking Artificial Bee Colony with Dynamic Cauchy Mutation for Feature Selection, *Complexity*, 2020.
- Chu, X., Li, S., Mao, W., Zhao, W. ve Huang, L., 2020b, A binary superior tracking artificial bee colony for feature selection, *International Conference on Neural Computing for Advanced Applications*, 296-305.
- Dadaneh, B. Z., Markid, H. Y. ve Zakerolhosseini, A., 2016, Unsupervised probabilistic feature selection using ant colony optimization, *Expert Systems with Applications*, 53, 27-42.
- Ding, S., Zhao, H., Zhang, Y., Xu, X. ve Nie, R., 2015, Extreme learning machine: algorithm, theory and applications, *Artificial Intelligence Review*, 44 (1), 103-115.
- Du, Z., Han, D. ve Li, K.-C., 2019, Improving the performance of feature selection and data clustering with novel global search and elite-guided artificial bee colony algorithm, *The Journal of Supercomputing*, 75 (8), 5189-5226.
- Durgut, R., Baydilli, Y. Y. ve Aydin, M. E., 2020, Feature Selection with Artificial Bee Colony Algorithms for Classifying Parkinson's Diseases, *International Conference on Engineering Applications of Neural Networks*, 338-351.
- Feng, L., Wang, Y. ve Zuo, W., 2017, Novel feature selection method based on random walk and artificial bee colony, *Journal of Intelligent & Fuzzy Systems*, 32 (1), 115-126.
- Ge, J., Zhang, X., Liu, G. ve Sun, Y., 2019, A Novel Feature Selection Algorithm Based on Artificial Bee Colony Algorithm and Genetic Algorithm, *2019 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 131-135.
- Guo, Y., Chung, F.-L., Li, G. ve Zhang, L., 2019, Multi-label bioinformatics data classification with ensemble embedded feature selection, *IEEE Access*, 7, 103863-103875.
- Hanbay, K., 2021, A new standard error based artificial bee colony algorithm and its applications in feature selection, *Journal of King Saud University-Computer and Information Sciences*.
- Hancer, E., Xue, B., Karaboga, D. ve Zhang, M., 2015a, A binary ABC algorithm based on advanced similarity scheme for feature selection, *Applied Soft Computing*, 36, 334-348.
- Hancer, E., Xue, B., Zhang, M., Karaboga, D. ve Akay, B., 2015b, A multi-objective artificial bee colony approach to feature selection using fuzzy mutual information, *2015 IEEE congress on evolutionary computation (CEC)*, 2420-2427.

- Hancer, E., Xue, B., Zhang, M., Karaboga, D. ve Akay, B., 2018, Pareto front feature selection based on artificial bee colony optimization, *Information Sciences*, 422, 462-479.
- Harudin, N., Ramlie, F., Muhamad, W., Azman, W. Z., Muhtazaruddin, M., Jamaludin, K. R., Abu, M. Y. ve Marlan, Z. M., 2021, Binary Bitwise Artificial Bee Colony as Feature Selection Optimization Approach within Taguchi's T-Method, *Mathematical Problems in Engineering*, 2021.
- Hijazi, M., Zeki, A. ve Ismail, A., 2021, Arabic Text Classification Using Hybrid Feature Selection Method Using Chi-Square Binary Artificial Bee Colony Algorithm, *Computer Science*, 16 (1), 213-228.
- Huang, G.-B., Zhu, Q.-Y. ve Siew, C.-K., 2004, Extreme learning machine: a new learning scheme of feedforward neural networks, *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)*, 985-990.
- Huang, G.-B., Zhu, Q.-Y. ve Siew, C.-K., 2006, Extreme learning machine: theory and applications, *Neurocomputing*, 70 (1-3), 489-501.
- Jain, I., Jain, V. K. ve Jain, R., 2018, Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification, *Applied Soft Computing*, 62, 203-215.
- Jia, D., Duan, X. ve Khan, M. K., 2014, Binary Artificial Bee Colony optimization using bitwise operation, *Computers & Industrial Engineering*, 76, 360-365.
- Karaboga, D., 2005, An idea based on honey bee swarm for numerical optimization, *Technical Report-TR06, Erciyes University, Engineering Faculty, Department of Computer Engineering*.
- Karaboga, D., Akay, B. ve Ozturk, C., 2007, Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks, *International conference on modeling decisions for artificial intelligence*, 318-329.
- Karaboga, D. ve Basturk, B., 2007, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, *Journal of global optimization*, 39 (3), 459-471.
- Karaboga, D. ve Basturk, B., 2008, On the performance of artificial bee colony (ABC) algorithm, *Applied Soft Computing*, 8 (1), 687-697.
- Karaboga, D. ve Akay, B., 2009, A comparative study of artificial bee colony algorithm, *Applied mathematics and computation*, 214 (1), 108-132.
- Karaboga, D. ve Ozturk, C., 2009, Neural networks training by artificial bee colony algorithm on pattern classification, *Neural Network World*, 19 (3), 279.
- Karaboga, D., Gorkemli, B., Ozturk, C. ve Karaboga, N., 2014, A comprehensive survey: artificial bee colony (ABC) algorithm and applications, *Artificial Intelligence Review*, 42 (1), 21-57.

- Karaboga, N., 2009, A new design method based on artificial bee colony algorithm for digital IIR filters, *Journal of the Franklin Institute*, 346 (4), 328-348.
- Kashan, M. H., Nahavandi, N. ve Kashan, A. H., 2012, DisABC: a new artificial bee colony algorithm for binary optimization, *Applied Soft Computing*, 12 (1), 342-352.
- Kaushik, R. ve Keswani, B., 2014, Hybrid Approach for Feature Subset Selection Hybrid Approach for Feature Subset Selection, *International Journal of Computer Science and Management Studies*, 14 (3), 10-14.
- Keleş, M. K., Keleş, A. E. ve Kılıç, Ü., 2018, Prediction of concrete strength with data mining methods using artificial bee colony as feature selector, *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1-4.
- Keleş, M. K. ve Kılıç, Ü., 2018, Artificial bee colony algorithm for feature selection on scadi dataset, *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, 463-466.
- Kharis, S., Hadi, I. ve Hasanah, K., 2019, Multiclass Classification of Brain Cancer with Multiple Multiclass Artificial Bee Colony Feature Selection and Support Vector Machine, *Journal of Physics: Conference Series*, 012015.
- Kiliç, Ü. ve Keleş, M. K., 2018, Feature selection with artificial bee colony algorithm on Z-Alizadeh Sani dataset, *2018 innovations in intelligent systems and applications conference (asyu)*, 1-3.
- Kiran, M. S. ve Gündüz, M., 2013, XOR-based artificial bee colony algorithm for binary optimization, *Turkish Journal of Electrical Engineering & Computer Sciences*, 21 (Sup. 2), 2307-2328.
- Kiran, M. S., 2015, The continuous artificial bee colony algorithm for binary optimization, *Applied Soft Computing*, 33, 15-23.
- Kiran, M. S., Hakli, H., Gunduz, M. ve Uguz, H., 2015, Artificial bee colony algorithm with variable search strategy for continuous optimization, *Information Sciences*, 300, 140-157.
- Kiran, M. S., 2021, A binary artificial bee colony algorithm and its performance assessment, *Expert Systems with Applications*, 175, 114817.
- Kumar, A., Kumar, D. ve Jarial, S., 2017, A review on artificial bee colony algorithms and their applications to data clustering, *Cybernetics and Information Technologies*, 17 (3), 3-28.
- Kumar, S. V. K. ve Nedunchezian, R., 2016, An Intelligent Feature Selection Approach for Gene Expression Data using Hybrid Biogeography based Optimization (BBO) with Artificial Bee Colony (ABC) Algorithm, *IIOAB JOURNAL*, 7 (9), 542-551.
- Kuo, R.-J., Huang, S. L., Zulvia, F. E. ve Liao, T. W., 2018, Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction, *Knowledge and Information Systems*, 55 (1), 253-274.

- Labani, M., Moradi, P., Ahmadizar, F. ve Jalili, M., 2018, A novel multivariate filter method for feature selection in text classification problems, *Engineering Applications of Artificial Intelligence*, 70, 25-37.
- Li, X., Li, M. ve Yin, M., 2016, Multiobjective ranking binary artificial bee colony for gene selection problems using microarray datasets, *IEEE/CAA Journal of Automatica Sinica*.
- Lin, K.-C. ve Hsieh, Y.-H., 2015, Classification of medical datasets using SVMs with hybrid evolutionary algorithms based on endocrine-based particle swarm optimization and artificial bee colony algorithms, *Journal of medical systems*, 39 (10), 1-9.
- Lin, Y., Wang, J., Li, X., Zhang, Y. ve Huang, S., 2021, An Improved Artificial Bee Colony for Feature Selection in QSAR, *Algorithms*, 14 (4), 120.
- Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y. ve Gao, Z., 2017, A hybrid feature selection algorithm for gene expression data classification, *Neurocomputing*, 256, 56-62.
- Mafarja, M. ve Mirjalili, S., 2018, Whale optimization approaches for wrapper feature selection, *Applied Soft Computing*, 62, 441-453.
- Mala, D. J. ve Mohan, V., 2009, ABC tester-artificial bee colony based software test suite optimization approach, *International Journal of Software Engineering*, 2 (2), 15-43.
- Maldonado, S. ve López, J., 2018, Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification, *Applied Soft Computing*, 67, 94-105.
- Marinakakis, Y., Marinaki, M. ve Matsatsinis, N., 2009, A hybrid discrete artificial bee colony-GRASP algorithm for clustering, *2009 International Conference on Computers & Industrial Engineering*, 548-553.
- Miao, J. ve Niu, L., 2016, A survey on feature selection, *Procedia Computer Science*, 91, 919-926.
- Mlakar, U., Fister, I., Brest, J. ve Potočnik, B., 2017, Multi-objective differential evolution for feature selection in facial expression recognition systems, *Expert Systems with Applications*, 89, 129-137.
- Mlakar, U. ve Fister, I., 2020, Impact of solution representation in nature-inspired algorithms for feature selection, *IEEE Access*, 8, 134728-134742.
- Mohammadi, F. G. ve Abadeh, M. S., 2013, An improved image steganalysis using a novel feature selection algorithm based on artificial bee colony, In: *Intelligent Systems and Decision Making for Risk Analysis and Crisis Response*, Eds: CRC Press, p. 609-614.
- Mohammadi, F. G. ve Abadeh, M. S., 2014, Image steganalysis using a bee colony based feature selection algorithm, *Engineering Applications of Artificial Intelligence*, 31, 35-43.

- Moradi, P. ve Gholampour, M., 2016, A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy, *Applied Soft Computing*, 43, 117-130.
- Nadira, T. ve Rustam, Z., 2018, Classification of cancer data using support vector machines with features selection method based on global artificial bee colony, *AIP Conference Proceedings*, 020205.
- Nagarajan, S. M., Muthukumar, V., Murugesan, R., Joseph, R. B. ve Munirathanam, M., 2021, Feature selection model for healthcare analysis and classification using classifier ensemble technique, *International Journal of System Assurance Engineering and Management*, 1-12.
- Naik, A. K., Kuppili, V. ve Edla, D. R., 2020, Efficient feature selection using one-pass generalized classifier neural network and binary bat algorithm with a novel fitness function, *Soft Computing*, 24 (6), 4575-4587.
- Nguyen, B. H., Xue, B. ve Zhang, M. J., 2020, A survey on swarm intelligence approaches to feature selection in data mining, *Swarm and Evolutionary Computation*, 54.
- Nouri-Moghaddam, B., Ghazanfari, M. ve Fathian, M., 2021, A novel multi-objective forest optimization algorithm for wrapper feature selection, *Expert Systems with Applications*, 175, 114737.
- Omkar, S. ve Senthilnath, J., 2009, Artificial bee colony for classification of acoustic emission signal source, *International Journal of Aerospace Innovations*, 1 (3), 129-143.
- Ozturk, C., Hancer, E. ve Karaboga, D., 2015, A novel binary artificial bee colony algorithm based on genetic operators, *Information Sciences*, 297, 154-170.
- Özger, Z. B., Bolat, B. ve Dırı, B., 2016, A comparative study on binary artificial bee colony optimization methods for feature selection, *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 1-4.
- Palanisamy, S. ve Kanmani, S., 2012a, Classifier ensemble design using artificial bee colony based feature selection.
- Palanisamy, S. ve Kanmani, S., 2012b, Artificial bee colony approach for optimizing feature selection, *International Journal of Computer Science Issues (IJCSI)*, 9 (3), 432.
- Pampará, G. ve Engelbrecht, A. P., 2011, Binary artificial bee colony optimization, *2011 IEEE Symposium on Swarm Intelligence*, 1-8.
- Prasartvit, T., Kaewkamnerdpong, B. ve Achalakul, T., 2011, Dimensional reduction based on artificial bee colony for classification problems, *International Conference on Intelligent Computing*, 168-175.

- Pulikanti, S. ve Singh, A., 2009, An artificial bee colony algorithm for the quadratic knapsack problem, *International Conference on Neural Information Processing*, 196-205.
- Punitha, S., Al-Turjman, F. ve Stephan, T., 2021, An automated breast cancer diagnosis using feature selection and parameter optimization in ANN, *Computers & Electrical Engineering*, 90, 106958.
- Rakshit, P., Bhattacharyya, S., Konar, A., Khasnobish, A., Tibarewala, D. ve Janarthanan, R., 2013, Artificial bee colony based feature selection for motor imagery EEG data, *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)*, 127-138.
- Rostami, M., Forouzandeh, S., Berahmand, K. ve Soltani, M., 2020, Integration of multi-objective PSO based feature selection and node centrality for medical datasets, *Genomics*, 112 (6), 4370-4384.
- Rostami, M., Berahmand, K., Nasiri, E. ve Forouzandeh, S., 2021, Review of swarm intelligence-based feature selection methods, *Engineering Applications of Artificial Intelligence*, 100, 104210.
- Rostami, O. ve Kaveh, M., 2021, Optimal feature selection for SAR image classification using biogeography-based optimization (BBO), artificial bee colony (ABC) and support vector machine (SVM): a combined approach of optimization and machine learning, *Computational Geosciences*, 25 (3), 911-930.
- Santana Jr, C. J., Macedo, M., Siqueira, H., Gokhale, A. ve Bastos-Filho, C. J., 2019, A novel binary artificial bee colony algorithm, *Future Generation Computer Systems*, 98, 180-196.
- Schiezaro, M. ve Pedrini, H., 2013, Data feature selection based on Artificial Bee Colony algorithm, *EURASIP Journal on Image and Video Processing*, 2013 (1), 1-8.
- Selvakumar, B. ve Muneeswaran, K., 2019, Firefly algorithm based feature selection for network intrusion detection, *Computers & Security*, 81, 148-155.
- Seth, J. K. ve Chandra, S., 2018, An effective DOS attack detection model in cloud using Artificial Bee Colony Optimization, *3D Research*, 9 (3), 1-13.
- Shi, Y., Pun, C.-M., Hu, H. ve Gao, H., 2016, An improved artificial bee colony and its application, *Knowledge-Based Systems*, 107, 14-31.
- Shokouhifar, M. ve Farokhi, F., 2010, An artificial bee colony optimization for feature subset selection using supervised fuzzy c-means algorithm, *3rd International conference on information security and artificial intelligent (ISAI)*, 427-432.
- Shokouhifar, M. ve Sabet, S., 2010, A hybrid approach for effective feature selection using neural networks and artificial bee colony optimization, *3rd international conference on machine vision (ICMV 2010)*, 502-506.
- Shunmugapriya, P., Kanmani, S., Supraja, R. ve Saranya, K., 2013, Feature selection optimization through enhanced Artificial Bee Colony algorithm, *2013*

- International Conference on Recent Trends in Information Technology (ICRTIT)*, 56-61.
- Shunmugapriya, P. ve Kanmani, S., 2017, A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid), *Swarm and Evolutionary Computation*, 36, 27-36.
- Singh, A., 2009, An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem, *Applied Soft Computing*, 9 (2), 625-631.
- Sivakumar, S. ve Chandrasekar, C., 2014, Feature selection using abc for the lung ct scan images, *International Journal of Scientific Engineering and Technology*, 3 (6), 781-784.
- Solorio-Fernández, S., Martínez-Trinidad, J. F. ve Carrasco-Ochoa, J. A., 2020, A Supervised Filter Feature Selection method for mixed data based on Spectral Feature Selection and Information-theory redundancy analysis, *Pattern Recognition Letters*, 138, 321-328.
- Subanya, B. ve Rajalaxmi, R., 2014, Feature selection using Artificial Bee Colony for cardiovascular disease classification, *2014 International Conference on Electronics and Communication Systems (ICECS)*, 1-6.
- Syarifahadilah, M., Abdullah, R. ve Venkat, I., 2012, ABC algorithm as feature selection for biomarker discovery in mass spectrometry analysis, *2012 4th Conference on Data Mining and Optimization (DMO)*, 67-72.
- UCI, 2021, Uci Machine Learning Datasets, <https://archive.ics.uci.edu/ml/datasets.php>: [12/12/2021].
- Uzer, M. S., Yilmaz, N. ve Inan, O., 2013, Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification, *The Scientific World Journal*, 2013.
- Uzer, M. S., Yilmaz, N. ve Inan, O., 2014, A Case Study: Effect of ABC-based Feature Selection Algorithm on Breast Cancer Diagnosis, *Global Journal on Technology*, 5.
- Wang, J., Li, T. ve Ren, R., 2010, A real time IDSs based on artificial bee colony-support vector machine algorithm, *Third International Workshop on Advanced Computational Intelligence*, 91-96.
- Wang, S., Pedrycz, W., Zhu, Q. ve Zhu, W., 2015a, Unsupervised feature selection via maximum projection and minimum redundancy, *Knowledge-Based Systems*, 75, 19-29.
- Wang, S., Gao, C., Zhang, Q., Dakulagi, V., Zeng, H., Zheng, G., Bai, J., Song, Y., Cai, J. ve Zong, B., 2020a, Research and experiment of radar signal support vector clustering sorting based on feature extraction and feature selection, *IEEE Access*, 8, 93322-93334.

- Wang, X.-h., Zhang, Y., Sun, X.-y., Wang, Y.-l. ve Du, C.-h., 2020b, Multi-objective feature selection based on artificial bee colony: An acceleration approach with variable sample size, *Applied Soft Computing*, 88, 106041.
- Wang, Z., Li, M. ve Li, J., 2015b, A multi-objective evolutionary algorithm for feature selection based on mutual information with a new redundancy measure, *Information Sciences*, 307, 73-88.
- Wikipedia, 2021, Stigmergy, https://en.wikipedia.org/wiki/Stigmergy#cite_note-mpira.ub.uni-muenchen.de-1: [12.12.2021].
- Yan, K., Ma, L., Dai, Y., Shen, W., Ji, Z. ve Xie, D., 2018, Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis, *International Journal of Refrigeration*, 86, 401-409.
- Yavuz, G. ve Aydin, D., 2016, Angle modulated artificial bee colony algorithms for feature selection, *Applied Computational Intelligence and Soft Computing*, 2016.
- Ye, D. ve Chen, Z., 2015, A new approach to minimum attribute reduction based on discrete artificial bee colony, *Soft Computing*, 19 (7), 1893-1903.
- Yusoff, S. A. M., Abdullah, R. ve Venkat, I., 2014, Adapted bio-inspired artificial bee colony and differential evolution for feature selection in biomarker discovery analysis, In: Recent advances on soft computing and data mining, Eds: Springer, p. 111-120.
- Zhang, T., Ding, B., Zhao, X. ve Yue, Q., 2018, A fast feature selection algorithm based on swarm intelligence in acoustic defect detection, *IEEE Access*, 6, 28848-28858.
- Zhang, X., Wu, G., Dong, Z. ve Crawford, C., 2015, Embedded feature-selection support vector machine for driving pattern recognition, *Journal of the Franklin Institute*, 352 (2), 669-685.
- Zhang, Y., Cheng, S., Shi, Y., Gong, D.-w. ve Zhao, X., 2019, Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm, *Expert Systems with Applications*, 137, 46-58.
- Zhang, Y., Wang, J., Li, X., Huang, S. ve Wang, X., 2021, Feature Selection for High-Dimensional Datasets through a Novel Artificial Bee Colony Framework, *Algorithms*, 14 (11), 324.
- Zorarpacı, E., Özel, S. A. ve Güngör, S., 2015, An Artificial Bee Colony Based Algorithm for Feature Selection, *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 30 (1), 25-32.
- Zorarpacı, E. ve Özel, S. A., 2016, A hybrid approach of differential evolution and artificial bee colony for feature selection, *Expert Systems with Applications*, 62, 91-103.