



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**Gri Kurt Meta-Sezgisel Algoritması Ve Rastgele Orman
Sınıflandırma Algoritmalarını Birleştirerek Koroner Kalp
Hastalığı Tanısının Doğruluğunun Artırılması**

Yüksek Lisans Tezi

Maysa Khodayverdian

Temmuz 2024



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**Gri Kurt Meta-Sezgisel Algoritması Ve Rastgele Orman
Sınıflandırma Algoritmalarını Birleştirerek Koroner Kalp
Hastalığı Tanısının Doğruluğunun Artırılması**

Yüksek Lisans Tezi

Maysa Khodayverdian

Danışman

Prof. Dr. Nihat Kabaoğlu

Temmuz 2024

TEZ JÜRİSİ ONAYI

Maysa KHodayverdian tarafından hazırlanan " Gri Kurt Meta-Sezgisel Algoritması Ve Rastgele Orman Sınıflandırma Algoritmalarını Birleştirerek Koroner Kalp Hastalığı Tanısının Doğruluğunun Artırılması" başlıklı bu Yüksek Lisans Tezi, Elektrik Elektronik Mühendisliği Anabilim Dalı'nda hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Prof. Dr. Nihat Kabaoğlu
İstanbul Medeniyet Üniversitesi

Üyeler:

Doç. Dr. Erdoğan Aydın
İstanbul Medeniyet Üniversitesi

Dr. Öğr. Üyesi Baran Tander
İstanbul Aydın Üniversitesi

Tez Savunma Tarihi: 05/ Temmuz / 2024

BEYANLAR

Yazım ve Kaynak Gösterme Kılavuzu Beyanı

Danışmanlığımda yazılan bu tezin APA yazım ve kaynak gösterme kılavuzunda belirtilen kurallara uygun olarak yapılandırıldığı ve bu kılavuzun metin içi kaynak gösterme standartlarının bu tezde tutarlı olarak uygulandığı tarafımdan incelenerek teyit edilmiştir.

İmza

Prof. Dr. Nihat Kabaoglu

Etik İlkeler Sadakat Beyanı

Hazırladığım bu tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi beyan ederim.

İmza

Maysa Khodayverdian

ÖZET

Gri Kurt Meta-Sezgisel Algoritması Ve Rastgele Orman Sınıflandırma Algoritmalarını Birleştirerek Koroner Kalp Hastalığı Tanısının Doğruluğunun Artırılması

KHodayverdian, Maysa

Yüksek Lisans Tezi, Elektrik Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Nihat Kabaoğlu

Temmuz 2024

Son yıllarda kalp-damar hastalıkları, dünya genelinde en önemli ölüm nedenlerinden biri haline gelmiştir. Günümüzde kalp hastalıklarının yüksek prevalansı nedeniyle, bu hastalıkların doğru ve kesin teşhisi için risk faktörlerinin belirlenmesi ve bu faktörlerin etkilerinin incelenmesi büyük önem taşımaktadır. Bu çalışmada, gri kurt optimizasyon algoritması (Gray Wolf Optimization (GWO)) ve rastgele orman (Random Forest (RF)) kombinasyonu kullanılarak koroner kalp hastalığı teşhis edilmektedir. Önerilen yöntem, özellik seçimi ve sınıflandırma olmak üzere iki ana aşamayı içermektedir. Sınıflandırıcı algoritma olarak rastgele orman algoritması tercih edilmiştir, zira elde edilen sonuçlar, rastgele orman algoritmasının büyük veri kümeleri için üstün performans sergilediğini göstermektedir. Ayrıca, veri kümesi üzerinde uygunluk fonksiyonunu (Fitness Function) kullanarak optimal özellik alt kümesini belirlemek için gri kurt algoritması uygulanmıştır. Özellik seçimi için gri kurt algoritmasının kullanımı, rastgele orman sınıflandırma algoritmasının performansını artırmıştır. Gri kurt algoritmasının uygunluk fonksiyonunun performansı, rastgele orman algoritması ile değerlendirilmiştir. Önerilen yöntem, Cleveland kalp hastalığı veri seti (Cleveland dataset) kullanılarak test edilmiştir. Deneysel sonuçlar, önerilen algoritmanın, tek başına rastgele orman algoritması ile

karşılaştırıldığında tahmin doğruluğunu yaklaşık %3 oranında artırdığını göstermiştir. Ayrıca, literatürdeki son çalışmalarla karşılaştırıldığında, gri kurt ve rastgele orman algoritmalarının birleşimi, Cleveland veri seti üzerinde diğer algoritmalarla karşılaştırıldığında %4,5'lik bir performans artışı sağlamıştır. Yöntemimiz, diğer algoritma kombinasyonları ile karşılaştırıldığında %97,93 kesinlik (Accuracy), %98,21 doğruluk (Precision) ve %97,34 kapsama (Recall) oranları elde etmiştir.

Anahtar Kelimeler: Koroner kalp hastalığı, rastgele orman algoritması, meta-sezgisel algoritmalar, gri kurt algoritması, değerlendirme kriterleri.

ABSTRACT

Improving The Accuracy Of Diagnosis Of Coronary Heart Disease By Combining Gray Wolf Meta-Heuristic Algorithm And Random Forest Classification Algorithms

Khodayverdian, Maysa

Master's Thesis, Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Nihat Kabaoğlu

July 2024

In recent years, cardiovascular diseases have become one of the most important causes of death in the world. Today, due to the high statistical population of heart diseases, identifying risk factors and paying attention to the effectiveness of these factors has led to accurate and correct diagnosis of this disease. In this thesis, the combination of gray wolf optimization algorithm and random forest is used to diagnose coronary heart disease. This proposed method includes two main stages of feature selection and classification. We use the random forest algorithm as the classifier algorithm because the results show that the random forest algorithm gives better results for large datasets. Also, the gray wolf algorithm has been used to select the subset of optimal features using the fitness function on the dataset. In fact, using the gray wolf algorithm for feature selection has improved the performance of the random forest classification algorithm. The performance of the fitness function of the gray wolf algorithm is evaluated using the random forest algorithm. The Cleveland heart disease dataset is used for the proposed method. The experimental results of the proposed algorithm are compared with the results of the random forest algorithm that is run alone on the data set, and the prediction accuracy is improved by nearly 3%. Also, in comparison with recent works, the combination of gray wolf and random forest algorithm

is compared with other algorithms on the Cleveland dataset. Our proposed algorithm had a performance improvement of 4.5% compared to the combination of neural network algorithm and random forest algorithm. Our method excels by achieving 97.93% in accuracy, 98.21% in correctness and 97.34% in coverage compared to other combinations of algorithms.

Keywords: Coronary heart disease, random forest algorithm, meta-heuristic algorithms, gray wolf algorithm, evaluation criteria.

ÖNSÖZ

Tez çalışmam sırasında kıymetli bilgi, birikim ve tecrübeleri ile bana yol gösterici ve destek olan saygıdeğer danışman hocam sayın Prof. Dr. Nihat KABAOĞLU ve her zaman yanımda olan sevgili aileme sonsuz teşekkür ederim.

Maysa KHODAYVERDIAN

Temmuz 2024

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	iii
ÖNSÖZ	v
İÇİNDEKİLER	vi
KISALTMALAR	x
ŞEKİL LİSTESİ.....	xi
TABLO LİSTESİ.....	xiii
GRAFİK LİSTESİ	xiv
DENKLEM LİSTESİ.....	xv
1. ARAŞTIRMAYLA İLGİLİ GENELLEMLER.....	1
1.1. Giriş	1
1.2. Problem Tanımı	1
1.3. Araştırmanın Motivasyonu Ve Önemi	2
1.4. Araştırmanın Hedefi	3
1.5. Yenilikçi Yönü.....	4
1.6. Tezin Yapısı.....	4
2. TEMEL TANIMLAR VE ARAŞTIRMA LİTERATÜRÜ.	6
2.1. Giriş	6
2.2. Kalp Hastalarının Teşhisinde Olan Bazı Risk Faktörlerinin Araştırılması ...	6
2.3. Koroner Kalp Hastalığı.....	8
2.4. Koroner Arter Hastalığının Belirtileri	9
2.4.1. Kardiyak Anjini.....	10
2.5. Kalp Hastalıklarını Teşhis Etmek İçin Yapılan Çalışmaların Özeti.....	11
2.6. Gri Kurt Algoritmasının Yaptığı İşin Arka Özeti_	13
2.7. Kalp Hastalıklarının Teşhisine Yönelik Bazı Uygulamaların Tanımı.....	16

2.8. Sonuç.....	18
3. ÖNERİLEN ALGORİTMA VE MOBİL SAĞLIĞA DAYALI PLATFORMLARA YÖNELİK AÇIKLAMALAR	19
3.1. Giriş.....	19
3.2. Öznitelik Seçimi	20
3.3. Meta-Sezgisel Algoritmalar.....	21
3.4. Meta-Sezgisel Algoritmelerde Keşif Ve Çıkarma Kavramı.....	25
3.5. Gri Kurt Optimizasyon Algoritması.....	27
3.6. Gri Kurtlarda Avlanma Yöntemi.....	27
3.7. Gri Kurt Algoritmasının Matematiksel Modellenmesi.....	28
3.8. Avın Kuşatılması İçin Matematiksel Denklemler.....	29
3.9. Avlanma Denklemleri.....	30
3.10. Av'a Erişim.....	31
3.11. Yerel Minimumlarda Tuzağa Düşen Meta-Sezgisel Algoritmaların Araştırması.....	32
3.12. Gri Kurt Algoritmasının Akış Şeması...../.....	33
3.13. Gri Kurt Algoritmasının Kullanımları.....	34
3.14. Gri Kurt Algoritmasının Teorik Yönleri.....	35
3.15. Birleşik Sınıflandırıcı Ve Boosting Ve Bagging Konuları.....	35
3.15.1. Bagging Yöntemi.....	36
3.15.2. Boosting Yöntemi.....	36
3.16. Karar Ağacı Algoritması.....	36
3.17. Rastgele Orman Algoritması.....	37
3.17.1. Rastgele Orman Algoritmasının Performansı.....	38
3.17.2. Rastgele Orman Algoritmasının Çalışma Yöntemi.....	38
3.17.3. Rastgele Orman Algoritmasının zellikleri.....	39
3.18. Sonuç.....	41
4. ÖNERİLEN ALGORİTMA VE UYGULANMASI İLE ÇIKTI SONUÇLARININ GÖRÜNTÜLENMESİ.	42
4.1. Giriş ...	42

4.2.	Önerilen Algoritma.....	42
4.3.	Gri Kurt Algoritmasının Özellik Seçme Mekanizması.....	44
4.4.	Gri Kurt Algoritmasında Özellik Seçiminin Değerlendirilmesi.....	45
4.5.	Önerilen Algoritmayı Uygulamak İçin Kullanılan Veriler.....	51
4.5.1.	Cleveland Verilerinin Özellikleri.....	51
4.5.2.	Farmingham Verilerinin Özellikleri.....	55
4.5.3.	Framingham Verilerinde Koroner Kalp Hastalığını Tanımaya Yönelik Özelliklerin Seçilmesindeki Öncelik.....	58
4.5.4.	Cleveland Verilerinde Koroner Kalp Hastalığını Tanımaya Yönelik Özelliklerin Seçilmesindeki Öncelik.....	58
4.6.	Önerilen Gri Kurt Ve Rastgele Orman Algoritmasını Uygulamak İçin Kullanılan Fonksiyonlar.....	58
4.6.1.	Gri Kurt Algoritmasının Uygulanması.....	61
4.7.	Çapraz Doğrulama Kriteri.....	61
4.8.	Değerlendirme kriterleri.....	63
4.8.1.	Performans Karakteristik Eğrisi.....	64
4.8.2.	Performans Karakteristik Diyagramının Altındaki Seviyenin Kriteri.....	66
4.8.3.	Karışıklık Matrisi.....	67
4.8.4.	Önerilen Gri Kurt Ve Rastgele Orman Algoritmasının Tekrarına Göre Maliyet Fonksiyonunun Diyagramı.....	71
4.8.5.	Önerilen Algoritmada Rastgele Orman Algoritmasını Oluşturan Karar Ağaçları.....	72
4.9.	Önerilen Gri Kurt Ve Rastgele o-Orman Algoritmasının Farklı Uygulamalardaki sonuçları.....	72
4.10.	Sonuç.....	74
5.	SONUÇ VE ÖNERİLER.....	75
5.1.	Giriş.....	75

5.2. Performans Karakteristik Diyagramının Analizi.....	75
5.3. Performans karakteristik diyagramının altındaki seviyenin kriter analizi.....	75
5.4. Karışıklık Matrisi Analizi Ve Değerlendirme Kriterleri.....	76
5.5. Tekrara Göre Maliyet Fonksiyon Grafiği Analizi.....	77
5.6. Karar Ağaçlarının Analizi.....	78
5.7. Önerilen Gri Kurt Ve Rastgele Orman Algoritması Tarafından Kaldırılan Özelliklerin Analizi.....	79
5.8. Öneriler Ve Gelecek Çalışmalar.....	80
KAYNAKLAR	82
TÜRKÇE - İNGİLİZCE SÖZLÜK.....	86

KISALTMALAR

ABC	Artificial Colony Algorithm
ACO	Ant Colony Bee Optimization
AUC	Area Under The Cover
BMI	Body mass index
EKG	Elektrokardiyografi
FNR	Miss rate or False Negative Rate
FPR	False Positive Rate
GA	Genetic Algorithm
GP	Genetic programming
GWO	Gray Wolf Optimization
HDL	High density lipoprotein
LDL	Low density lipoprotein
MATLAB	MATrix LABoratory
MCV	Mean Corpuscular Volume
NPV	Negative Predictive Value
PPV	Positive Predictive Value
PSO	Particle Swarm Optimization
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SA	Simulated Annealing
SVM	Support Vector Machine
TNR	True Negative Rate
TPR	True Positive Rate
UCI	University of California, Irvine

ŞEKİL LİSTESİ

Şekil 2.1 Damarlarda yağ ve kolesterol birikmesi.	10
Şekil 3.1 Özellik seçme yöntemleri	21
Şekil 3.2 İki filtre ve paketleme yönteminin karşılaştırılması.	21
Şekil 3.3 Meta-sezgisel algoritmaların sınıflandırılması	23
Şekil 3.4 Meta-sezgisel algoritmalarda çıkarma ve keşfetme kavramı.....	26
Şekil 3.5 Meta-sezgisel algoritmalarda madencilik ve keşif ilerlemesi.	26
Şekil 3.6 Gri kurt avlanma şekli.....	28
Şekil 3.7 Gri kurtların sosyal hiyerarşisi.....	28
Şekil 3.8 İki boyutlu (A) ve üç boyutlu (B) konum vektörleri ve olası konumlar.....	30
Şekil 3.9 Gri kurt algoritmasında kurtların konumunun güncellenmesi.....	30
Şekil 3.10 Kurtların konumu nasıl güncellenir ve ava doğru nasıl hareket edilir.....	31
Şekil 3.11 Arama ve çıkarmada A parametresinin etkisi.....	32
Şekil 3.12 Gri kurdun hareketinde A parametresinin rolü ve keşif ve çıkarma konuları.....	33
Şekil 3.13 Gri Kurt Algoritmasının akış şeması.....	34
Şekil 3.14 Gri kurt algoritmasının farklı alanlardaki kullanımı.....	34
Şekil 3.15 Gri Kurt Algoritmasının Teorik Yönleri.....	35
Şekil 3.16 Sınıflandırma yöntemi.....	36
Şekil 3.17 Rastgele ormanın performansı.....	39
Şekil 4.1 Önerilen gri kurt ve rastgele orman modeline dayalı sınıflandırma sistemi.....	43
Şekil 4.2 Önerilen gri kurt ve rastgele orman algoritması için veri madenciliği yapısı.....	44
Şekil 4.3 Çözümün özellik seçiminde gösterimi.....	44
Şekil 4.4 Özellik seçim ekranı. ...	45
Şekil 4.5 Önerilen algoritmanın akış şeması..	47
Şekil 4.6 Cleveland verilerin özellikleri.....	51
Şekil 4.7 Farmingham verilerin özellikleri	55
Şekil 4.8 Doğrulama yöntemi	62

Şekil 4.9 Önerilen gri kurt ve rastgele orman algoritmasının performans karakteristik diyagramı .	65
Şekil 4.10 Rastgele orman algoritması için performans karakteristik diyagramı.	65
Şekil 4.11 Önerilen gri kurt ve rastgele orman algoritmasının performans karakteristik diyagramı	66
Şekil 4.12 Rastgele orman algoritması için performans karakteristik diyagramı.....	66
Şekil 4.13 Karışıklık matrisi.....	68
Şekil 4.14 Önerilen gri kurt ve rastgele orman algoritması için karışıklık matrisi.....	69
Şekil 4.15 Rastgele orman algoritması için karışıklık matrisi.....	69
Şekil 4.16 Önerilen gri kurt ve rastgele orman algoritması için karışıklık matrisi.....	70
Şekil 4.17 Rastgele orman algoritması için karışıklık matrisi.....	70
Şekil 4.18 Rastgele orman algoritmasında karar ağacı.....	72

TABLO LİSTESİ

Tablo 4.1. Önerilen algoritmanın verimliliğinin önceki çalışmalarla karşılaştırılması.....	48
Tablo 4.2 Cleveland verilerinin özelliklerinin açıklaması.....	52
Tablo 4.3 Önerilen algoritmanın Cleveland verilerinin rastgele orman algoritması ile performans karşılaştırması.....	73
Tablo 4.4. Cleveland verilerinin önerilen algoritmasında kaldırılan özellikler.....	73
Tablo 4.5 Önerilen algoritmanın Farmingham verilerinin rastgele orman algoritması ile performans karşılaştırması.....	73
Tablo 4.6 Farmingham verilerinin önerilen algoritmasında kaldırılan özellikler.....	73

GRAFİK LİSTESİ

Grafik 4.1 Cleveland verilerinin tekrarına göre maliyet fonksiyonu grafiğı.....	71
Grafik 4.2 Farmingham verilerinin tekrarına göre maliyet fonksiyonu grafiğı.....	71

DENKLEM LİSTESİ

Denklem 3.1 Avın kuşatılmasıyla ilgili ilişki (kurtların farklı yinelemelerdeki konumu).....	29
Denklem 3.2 Kurtların önceki konumlarına göre avın kuşatılmasıyla ilgili ilişki.....	29
Denklem 3.3 Rastgele bir değerin r_1 ve r_2 cinsinden hesaplanması.....	29
Denklem 3.4 Avın kuşatılması ilişkisine ilişkin C'nin rastgele değerinin hesaplanması.....	29
Denklem 3.5 Avlanma denklemleriyle ilgili ilişki (alfa, en iyi avlanma konumu)	30
Denklem 3.6 Avlanma denklemleriyle ilgili ilişki (beta, ikinci en iyi avlanma konumu).....	30
Denklem 3.7 Avlanma denklemleriyle ilgili ilişki (Delta, üçüncü en iyi avlanma konumu).	30
Denklem 3.8 Ortalama avlanma konumlarının ilişkisi	30
Denklem 3.9 Gri kurt algoritmasının keşfini ve çıkarılmasını kontrol etme yöntemi arasındaki ilişki	31
Denklem 3.10 Gri kurt algoritmasının keşif ve çıkarma konularını kontrol etme şekli arasındaki ilişki ...	32
Denklem 3.11 Rastgele orman algoritmasının karar fonksiyonunun performans ilişkisi.....	38
Denklem 4.1 Sınıflandırma doğruluğunu en üst düzeye çıkarmak için uydurma fonksiyonunun ilişkisi.....	45
Denklem 4.2 Vücut kitle indeksi hesaplama ilişkisi.....	57
Denklem 4.3 Sınıflandırma doğruluğunu en üst düzeye çıkarmak için uydurma fonksiyonunun ağırlıklı bir fonksiyon olarak ilişkisi.....	59
Denklem 4.4 Doğruluk hesaplama ilişkisi	63
Denklem 4.5 Doğruluk hesaplama ilişkisi.....	63
Denklem 4.6 Duyarlılık hesaplama ilişkisi.....	63
Denklem 4.7 F1-score hesaplama ilişkisi.....	64
Denklem 4.8 Ayrıcılık olmanın hesaplanması.....	64

Denklem 4.9 Negatif tahmin değeri hesaplanması.....	64
Denklem 4.10 Kayıp oranı hesaplanması.....	64

1. ARAŞTIRMAYLA İLGİLİ GENELLEMLER

1.1. Giriş

Kalp veya kalp-damar hastalıkları ile ilgili hastalıklar, son birkaç on yılda dünyada birçok ölümün ana nedeni olmuş ve tüm dünyada en tehlikeli hastalık olarak ortaya çıkmıştır. Bu çalışmada koroner arter hastalığını teşhis etmek için bir makine öğrenme algoritması kullanıldı. Kullandığımız algoritma, rastgele orman algoritmasıdır. Rastgele orman algoritmasında çeşitli karar ağaçları kullanılır. Bu sınıflandırma algoritmasının performansını artırmak için gri kurt adı verilen meta-sezgisel algoritmalarından (Meta-heuristic) biri ile kombinasyonu kullanılır. Gri kurt algoritması ile özellik seçimi yapıldığında rastgele orman sınıflandırma algoritması uygulanmakta ve bu algoritmanın doğruluğunu ve verimliliğini %2 ila %3 oranında arttırmaktadır.

1.2. Problem Tanımı

Günümüzde bilgisayar, ağ ve tıp bilimi alanındaki araştırmacılar akıllı sağlık alanında geniş bir bakış açısı oluşturmaya çalışmaktadırlar. İnternetin giderek yaygınlaşması ve mevcut sensörlerin ve akıllı araçların artmasıyla birlikte, hastalık teşhisi alanında platformların oluşturulduğuna tanık oluyoruz (Dehghan & Roostaei, 2013). Bu yeni alanda, insanlardan bilgi alma özelliğine sahip giyilebilir cihazlar, kullanıcılar arasında gün geçtikçe daha popüler hale gelmiş olup, insanların hareket kabiliyetini, nabzını, kan basıncını ve daha birçok aktiviteyi ölçerek iletişim kurabilecek merkezler için bunları belirleyebilmekte ve muhabere etmektedirler (Vahdat & Kaysari, 2017). Günümüzde insan vücudunun farklı bölgelerine bağlanan ve hastanın sağlık durumu hakkında çeşitli bilgiler sağlayan yaklaşık 3,7 milyon tıbbi cihaz bulunmaktadır (Gharib & Sabzian, 2017). Bu hastalık kan ve kalp damarlarındaki bozukluklarla ilişkilidir ve sıklıkla kalpte aritmi, felç, yüksek tansiyon ve kalp yetmezliğine yol açar. Kardiyovasküler hastalıklar, kronik ve bulaşıcı olmayan hastalıklar kategorisinde sınıflandırılır. Kalp hastalıkları yavaş, uzun süreli, hatta yaşam boyu seyrederek ve genellikle sessiz belirtiler göstererek yaşam kalitesini etkiler. Bu çalışmada kalp hastalıklarını teşhis etmek için iki farklı veri seti kullanılmış olup, bu veri setlerinden birinin verileri, kalp atış hızı ve kan basıncını kaydeden gömülü sensörlerden elde edilen bilgilerle, diğer veri seti ise kalp atış hızı ve kan basıncını kaydeden sonuçlar

laboratuvarında elde edilir. Daha sonra veri madenciliği teknikleri ve rastgele orman sınıflandırma algoritması kullanılarak kalp hastalıklarının yüksek doğruluk ve verimlilikle tahmin edilebilmesi için veriler analiz edilmiştir. Bu çalışmada rastgele orman algoritması ile meta-sezgisel gri kurt algoritmasının birleştirilmesi alanında yeni bir fikir araştırılmaktadır.

1.3. Araştırmanın Motivasyonu Ve Önemi

Kalp hastalıklarının bir önceki bölümde bahsedilen önemi ile teknolojinin giderek ilerlemesi ve cep telefonlarının akıllı hale gelmesi, mevcut sensörlerin ve onlardan elde edilen bilgilerin kullanılması birçok kalp hastalığını tanımlanabilir hale getirmiştir.

Sağlıksız yaşam tarzı ve kötü beslenme alışkanlıkları kalp hastalıklarının ana nedenlerinden biridir. Yaşam tarzı değişikliklerinin yanı sıra sigara, alkol, kolesterol düzeyleri, obezite, yüksek tansiyon, diyabet vb. gibi birçok faktör vardır ki kalp sorunlarına neden olan faktörler arasında yer almaktadırlar. Sağlıklı bir yaşam tarzını seçerek birçok kalp hastalığı türü ortadan kaldırılabilir. Ancak son araştırmalar, yapay zeka ve tıp biliminin devreye girmesiyle, aslında her türlü hastalığın önlenmesinde bu ikisinden büyük yardım alınabileceğini gösteriyor.

Bu nedenle kalp hastalıklarının erken ve zamanında teşhisi, bu hastalıkların kontrolünde çok önemlidir. Hastalık belirtilerinin sürekli takibi, kalp hastalıklarının kontrolünde önemli rol oynuyor. Günümüzde akıllı cep telefonları, akıllı bileklikler ve diğer uygun cihazlar gibi çeşitli cihazlara yerleştirilmiş sensörler kullanılarak istenilen belirtilerin sürekli olarak değerlendirilmesi mümkün olmaktadır. Bu konuda pek çok platform oluşturulmuş ve yaşam tarzının değiştirilmesinde, tansiyon, kalp atış hızı, kilo gibi hayati parametrelerin kontrol edilmesinde önemli ve hayati bir rol oynamıştır. Örneğin Google Fit platformu, kalp hastalarının yaşam tarzlarını iyileştirmelerine yardımcı olabilecek platformlardan biridir. Bu tez çalışmasında bir Xiaomi bilekliği Google Fit platformuna bağlanıyor ve bileklik verileri bu platformda görüntülenebiliyor. Makine öğrenmesi algoritmalarını kullanarak ve bunları meta-sezgisel algoritmalarla birleştirerek kalp hastalıklarını tanımlamak da mümkündür. Bu konuda daha önce yapılan araştırmalarda meta-sezgisel algoritmalar ve makine öğrenmesi algoritmalarının kullanılmasının önemi fark edilmiş, bu nedenle meta-sezgisel algoritmalarından biri olan gri kurt algoritması Rastgele orman algoritması ile

beraber kullanılarak, rastgele orman algoritmasının doğruluğu ve verimliliği artırılmıştır.

1.4 Araştırmanın Hedefi

Kalp hastalıklarının teşhisinin ve sürekli takibinin, kalp hızı, tansiyon gibi önemli ve ana faktörlerin kontrol edilmesinin önemi nedeniyle bu alanda giyilebilir cihazlar vasıtasıyla, sürekli takip edebilecek çeşitli sistem ve platformlar oluşturulur ve bunlardan birkaç tanesi kalp hastalıklarını izlemek için kullanılır. Örneğin; M Serebrum (M Cerebrum), Sağlık Kiti (Health Kit), Oyer (Aware) ve benzeri platformlar.

Ayrıca bu tezin amaçlarından biri de kalp hastalıkları alanında iyi bilinen ve güvenilir birçok platformu incelemektir. Hastaların bu giyilebilir cihazları kullanmaya teşvik edilmesi ve kan basıncı, kalp atış hızı, kilo ve diğer faktörlerin sürekli kontrol ve takip edilmesi birçok kalp krizini önleyebilir.

Özellikle sağlık tesislerine sınırlı erişimi olan kişiler için üstün sağlık ve gözetim hizmetlerinin uygun fiyat ve kalitede sağlanması esastır. Ayrıca, kalp hastalarının önemli fizyolojik parametrelerinin ve aktivitelerinin sürekli olarak uzun süreli izlenmesi yoluyla birçok tıbbi komplikasyon önlenabilir veya uygun şekilde yönetilebilir. Sağlığın uzun süreli izlenmesi, gelişen hastalıkların erken teşhis edilmesini sağlar. Öte yandan akıllı telefon tabanlı sağlık sistemleri, uzun vadeli sağlık takibi için uygun maliyetli bir alternatif sunabilir ve sağlık personelinin hastalarını uzaktan izlemesine ve değerlendirmesine olanak sağlayabilir.

Bu araştırmadaki temel ve önemli hedeflerden biri makine öğrenmesi algoritmalarını kullanmak ve minimum hatayla en iyi doğruluk seviyesine sahip bir model sağlamaktır. Yani bu araştırmanın temel ve önemli amacının koroner kalp hastalığı teşhisinin doğruluğunu arttırmak olduğu ifade edilebilir. Bu nedenle burada sunulan model, daha önce sunulan çalışmalara kıyasla kalp hastalığı teşhisinde minimum hatayla en iyi doğruluk seviyesine sahiptir.

Aslında bu çalışmada, özelliğin gri kurt optimizasyon algoritması kullanılarak seçilmesi ve ardından rastgele orman sınıflandırma algoritmasının uygulanmasıyla, koroner kalp hastalığı teşhisinin doğruluğunun, yalnızca rastgele orman algoritmasının kullanıldığı duruma kıyasla artırılabilceği gösterilmiştir.

1.5. Yenilikçi Yönü

Önceki bölümlerde de belirtildiği gibi, kalp hastalıklarını teşhis etmek, tahmin etmek ve verileri sınıflandırmak için makine öğrenmesi algoritmalarından, rastgele orman algoritması kullanılmıştır. Ayrıca, rastgele orman algoritmasının performansını ve verimliliğini artırmak için bu algoritma gri kurt meta-sezgisel algoritması ile birleştirilmiştir. Bu sayede, önemli özelliklerin seçilmesi için gri kurt algoritması kullanılırken, sonrasında üzerine rastgele orman algoritması uygulanarak rastgele orman algoritmasının doğruluğu ve verimliliği artırılmıştır.

Burada incelenen kalp hastalığı koroner kalp hastalığıdır. İkinci bölümde tanıtılan giyilebilir sensörler kullanılarak kalp hastalığının tanısında önemli faktörler olan kan basıncı, kalp atış hızı ve vücut ısısı inceleniyor. Belirtildiği gibi burada sınıflandırma için kullanılan makine öğrenimi algoritması rastgele orman algoritmasıdır. Bu algoritmanın MATLAB yazılımında uygulanmasıyla, mevcut iki veriye göre, algoritmanın kalp hastalığı tanısına yönelik doğruluğu %92 civarındadır.

Üçüncü bölümde ayrıntılı olarak anlatılan ve matematiksel modeli de sunulan gri kurt algoritması meta-sezgisel algoritmalarından biridir. Bu algoritmanın uygun özelliklerin seçilmesinde kullanılması ve sınıflandırma için ise makine öğrenmesi algoritmalarından birinin kullanılması (ki bu araştırmada rastgele orman algoritması kullanılmıştır), önerilen birleşik algoritmanın kalp hastalığı tanısının doğruluğunu yaklaşık %3 oranında artırmıştır.

1.6. Tezin Yapısı

Bu tezde, giyilebilir sensörlerden elde edilen veriler kullanılarak kalp hastalıklarının (koroner damarlar) tahmin edilmesine yönelik detaylı bir inceleme sunulmaktadır. Birinci bölümde bu çalışmanın genel yönleri, amaçları, önemi ve yenilikçi yönleri tartışılmaktadır. İkinci bölümde, kalbin yapısı, kalp-damar hastalıkları ve buna bağlı risk faktörlerinin¹⁸ incelenmesinin yanı sıra kalp hastalıklarının tanısına yönelik çeşitli platformların incelenmesini de içeren temel tanımlar ve araştırma literatürü gözden geçirilmiştir. Sonununcu bölümde ise, makine öğrenimi algoritmaları ve meta-sezgisel algoritmaların kombinasyonu kullanılarak kalp hastalıklarının teşhisi ve tahmini alanında daha önce yapılan bir dizi çalışma sunulmuştur. Üçüncü bölümde temel kavram ve yöntemler sunulmuştur. Gri kurt meta-sezgisel algoritmasının matematiksel modeli ve bu algoritmanın özellikleri ayrıntılı olarak anlatılmıştır. Ayrıca, rastgele

orman algoritmasının performansı da tanıtılmıştır. Dördüncü bölümde önerilen algoritma ve bu algoritma ile öznelilik seçme yöntemi, uydurma fonksiyonu kullanılarak ayrıntılı olarak anlatıldıktan sonra, önerilen algoritmanın uygulaması yapılarak tüm değerlendirme kriterleri ve bunlara ilişkin çıktılar görüntülenmiştir. Son olarak, beşinci bölümde tüm değerlendirme kriterlerinin analizi ve bu kriterlerden elde edilen diyagramlar ve çıktılar sunulmuştur.

2. TEMEL TANIMLAR VE ARAŞTIRMA LİTERATÜRÜ

2.1. Giriş

Bu bölümde öncelikle kalp hastalıklarının teşhisine ilişkin risk faktörleri incelenerek, bu hastalığa neden olan çeşitli faktörler ve teşhis alanı incelenir. Örneğin kalp atış hızı, tansiyon, kan kolesterolü gibi kalp hastalıklarının bazı faktörleri inceleniyor. İleride koroner arter hastalığı kavramı ve semptomları tartışılmaktadır. Daha sonra gri kurt algoritması kullanılarak kalp hastalığı teşhisleri ve koroner arter hastalığı tanısı alanında yapılmış olan çalışmalar da incelenmiştir. Son olarak kalp hastalıklarını teşhis etmek için kullanılan bazı uygulamalar ve sensörler tanıtılmıştır.

2.2. Kalp Hastalarının Teşhisinde Olan Bazı Risk Faktörlerinin Araştırılması

Dünya Sağlık Örgütü'ne göre kalp hastalıkları, yüksek ve düşük gelirli ülkelerde birinci ölüm nedeni olup, kadın ve erkeklerde neredeyse eşit oranda başvermektedir. 2030 yılına kadar dünyadaki ölümlerin yaklaşık %76'sı bulaşıcı olmayan hastalıklardan (Non-Communicable Diseases (NCDs)) kaynaklanacaktır (Mann et al., 2014).

Kalp atış hızı, çeşitli aritmi türleri gibi kalple ilgili hastalıkları teşhis etmek için doktorlar tarafından günlük olarak izlenen dört "yaşam belirtisinden" biridir. Kalp hastalıklarının tanısı için önemli ve temel risk faktörlerinin bir kısmı kontrol edilebilir, bir kısmı ise kontrol edilememektedir. Kontrol edilebilir risk faktörleri arasında yüksek kan kolesterol düzeyleri, sigara kullanımı, ikinci tip diyabet, aşırı kilo veya obezite, fiziksel hareketsizlik, sağlıksız beslenme ve stres yer almaktadır.

Kontrol edilemeyen şeyler arasında yaş (yaşlanma riski artırır), cinsiyet (erkeklerde genellikle koroner arter hastalığı riski daha fazladır), aile öyküsü ve ırk yer almaktadır. Framingham çalışması koroner kalp hastalığının anlaşılmasına ve bu alandaki temel risk faktörlerinin belirlenmesine büyük katkı sağlamıştır. Bu veriler, yüksek tansiyon, yüksek kan kolesterolü, sigara içme, obezite, ikinci tip diyabet ve fiziksel hareketsizlik gibi bilgilerin yanı sıra kan trigliseritleri, yaş, cinsiyet ve psikososyal sorunlar gibi ilgili faktörlerin etkileri hakkında değerli bilgileri içerir. Bugüne kadar kalp hastalığı için risk faktörü olarak tek bir risk faktörü tanımlanmamıştır. Aksine, gelişiminden birbiriyle bağlantılı birkaç faktörün sorumlu olduğu görülmektedir. İki tür diyabet vardır. Birinci tip, kontrol edilmesi için insülin enjeksiyonu gerektiren insüline bağımlı

diyabet olup, ikinci tip yani yaygın tip ise orta yaşlı kişilerde ve vücudu insüline direnç gösteren kişilerde ortaya çıkar ve uygun beslenme ve ağızdan alınan ilaçlarla kontrol altına alınır.

Ayrıca pek çok beslenme faktörü kalp-damar hastalıklarının risk faktörleriyle ilişkilidir. Kilo alımı ile vücut kitle indeksi (BMI) ile yüksek tansiyon arasında doğrudan bir ilişki vardır. Aslında kilo alımı, kan kolesterolü ve LDL'in artmasının iyi bir göstergesi olabilir. Ayrıca artan yaş, kan kolesterol seviyelerinin artması için iyi bir öngörü faktörüdür.

- Bu bölümde toplam kolesterol düzeyi (LDL ve HDL kolesterolün toplamı) ve hangi seviyede sağlıklı olduğu incelenerek kalp ve damar sağlığına olumlu etkisi incelenecektir. Kan kolesterolünü teşhis etmenin tek yolu, toplam kolesterol, LDL (kötü yağ veya kalp ve kan damarları üzerinde kötü etkileri olan) veya HDL (iyi yağ veya kalp üzerinde olumlu etkileri olan) hakkında bilgi sağlayan bir kan testidir (kan yağı).

Toplam kan kolesterol düzeyi :

- 200 mg/dl'nin altında olması optimum seviye (düşük risk).
- Orta düzey 200-239 mg/dl (risk artışı)
- mg/dl 240 veya daha fazla, yüksek (optimum seviyenin iki katından fazla risk artışı)
- Normal kan basıncı 120/80 mmHg'dır. Yüksek tansiyon, sistolik kan basıncının 140 veya daha fazla olması veya diyastolik kan basıncının 90 veya daha fazla olması anlamına gelir ve bunun elbette uzun süre yüksek olması gerekir.
- Yetişkinlerde aşırı kilo ve obezitenin sınıflandırılması, kilonun boya göre basit bir indeksi olan vücut kitle indeksi (BMI) ile yapılır. BMI'nin 25-29'a kadar olması aşırı kilonun işaretidir. 30'dan büyük veya ona eşit bir BMI obeziteyi gösterir. BMI'nin bazı sınırlamaları vardır. Örneğin, BMI mutlak bir şekilde vücut yağ dağılımını yansıtmaz. Bu nedenle kalp-damar hastalığı gibi kronik bir hastalığa yakalanma riskini ölçmenin en kolay yolu bel çevresini ölçmektir. Bel ölçümü yalnızca yetişkinlerin kardiyovasküler hastalıklar da dahil olmak üzere kronik hastalıklara yakalanma riskinin ne kadar yüksek olduğunu belirtmek için kullanılmalıdır. Ancak dikkat edilmelidir ki çocuklar ve gençler için risk artımı gözükmemiştir. Erkeklerde bel çevresinin 90 cm'den,

kadınlarda ise 80 cm'den fazla olması kalp-damar hastalıkları riskinin artmasıyla ilişkilidir. Ancak erkeklerde bel çevresinin 102 cm'den, kadınlarda ise 88 cm'den fazla olması, kalp-damar hastalıkları riskinde artışa ve bazı kanserler, ikinci tip diyabet, kalp hastalıkları gibi kronik hastalıklar da dahil olmak üzere ciddi sorunların ortaya çıkmasına güçlü bir şekilde işaret etmektedir.

- Stres, kişinin hayatındaki değişiklikler ve ihtiyaçlar sonucunda yaşadığı fiziksel, zihinsel ve duygusal tepkilerdir. Stres yönetilmezse duygusal, psikolojik ve hatta kalp hastalığı, yüksek tansiyon, göğüs ağrısı veya düzensiz kalp atışı gibi fiziksel sorunlara yol açabilir. Sürekli stres, adrenalin ve kortizol gibi stres hormonlarını artırır. Araştırmalar stres ile kan pıhtıları arasındaki ilişkiyi ve sonuçta kalp krizi riskini artırdığını gösteriyor. Stresin kendisi bir risk faktörü olabilir veya diğer risk faktörlerinin (yüksek kolesterol veya yüksek tansiyon gibi) gelişmesine katkıda bulunabilir.

2.3. Koroner Kalp Hastalığı

Koroner arter tıkanıklığı veya kalp hastalığı olarak da adlandırılan koroner arter hastalığı, Amerika'da başlıca ölüm nedenidir ve Amerika'da 13 milyondan fazla insan bu hastalıktan ötürü acı çekmektedir. Kalp hastalığı, trombositlerin kan damarlarında birikmesi sonucu oluşur ve bunun sonucunda kan akışı engellenir; hastada kalp krizi ve Miyokardiyal enfarktüs riski artar. Kalp hastalığına, koroner arterlerde trombosit birikmesi neden olur. Kan akışının engellenmesine neden olan bu soruna arterioskleroz denir. Başlangıçta yumuşak ve elastik olan kan damarları daralır sertleşir ve kalbe kan akışı engellenir. Bunun sonucunda oksijen ve besinler kalbe ulaşamaz ve kalp kanı iyi bir şekilde pompalayamaz hale gelir. Genç yaşlardan itibaren yavaş yavaş kan damarlarının duvarlarında kolesterol plakları birikmeye başlar ve yaş ilerledikçe bu yağlı plaklar atardamarlarda daha fazla birikerek kan damarı duvarlarında iltihaplanmaya neden olur ve risk artar, kan pıhtıları ve kalp krizi oluştururlar. Bu plaklar, bir yandan damar iltihabının iyileşme sürecini hızlandıran kimyasallar salgılar, ancak bir yandan da kan damarlarının yapışmasına neden olur. Daha sonra damarlarda hareket eden hasarlı hücreler ve kalsiyum gibi diğer maddeler yavaş yavaş damar duvarlarına yapışır. Son olarak daralmış koroner arterler, kalbe kan dağıtan yeni kan damarlarının

oluşmasına yol açar. Ancak, kişi şiddetli stres altında olduğunda bu yeni damarlar oksijenden zengin kanı kalp kasına ulaştıramayabilir. Arteriyel tıkanıklık çocukluk çağında da başlayabilen yavaş ve ilerleyici bir hastalıktır. Kesin nedeni bilinmemekle birlikte ateroskleroz, atardamarın iç zarının hasar görmesi ile başlayabilir. Hasar aşağıdaki nedenlerden kaynaklanabilir;

- Yüksek tansiyon
- Yüksek kolesterol
- Yüksek trigliseritler
- Sigara kullanımı
- Obezite ve diyabet

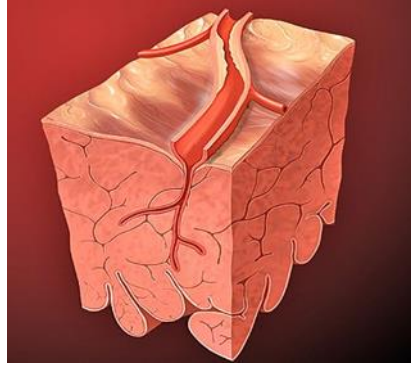
Bir arterin iç duvarı hasar gördüğünde, kan hücreleri ve diğer maddeler sıklıkla hasarlı bölgede sıkışıp kalır ve arterin içine yerleşir. Zamanla, hasar bölgesinde kolesterol ve diğer hücresel ürünlerden oluşan yağ birikintileri (plak) de oluşur ve sertleşerek arterleri daraltır. Tıkalı arterlere bağlı organ ve dokular düzgün çalışmaları için yeterli miktarda kan alamazlar. Ayrıca plak, vücudunuzun belirli bölgelerine kan akışını durdurabilecek kan pıhtılarına da neden olabilir (Dr. Najafi, 2015; Mann et al., 2014).

2.4. Koroner Arter Hastalığının Belirtileri

Koroner arter hastalığının en sık görülen semptomu anjina veya göğüs ağrısıdır. Anjina, ağırlık, basınç, ağrı, uyuşukluk ve yanma şeklinde hissedilebilir ve hasta bunu hazımsızlık veya mide yanması ile karıştırabilir. Anjina genellikle göğüs bölgesinde hissedilir ancak sol omuz, kollar, boyun, bel veya çenede de hissedilebilir (Dr. Najafi, 2015).

Koroner arter hastalığının test edilmesi ve ölçülmesi gereken ve genellikle sessiz kalan diğer semptomları arasında yüksek tansiyon, yüksek kan kolesterol düzeyleri ve BMI'daki artışı sayabiliriz.

Şekil 2.1 arterlerde yağ ve kolesterol birikimini göstermektedir (Dr. Najafi & M.H, 2015).



Şekil 2.1 Damarlarda yağ ve kolesterol birikmesi.

2.4.1. Kardiyak Anjini

Anjina göğüste ağrı veya rahatsızlık hissidir ve kalbinizin ihtiyaç duyduğu oksijeni almaması durumunda ortaya çıkar ve *iskemi* (düşük oksijen) adı verilen bir durum ortaya çıkar. Göğüs ağrısı olarak da bilinen anjina, genellikle göğüste ağrı, sıkışma, ağırlık veya kramp hissi olarak ifade edilen, koroner arterleri ve kalbi etkileyen hastalıklardan biridir. Angina genellikle sadece birkaç dakika sürer. Bu bölümde anjinası olan hastaların ifade ettiği duygu örnekleri işare edilmiştir:

- Göğüste sertlik ve ağırlık hissi.
- Nefes darlığı hissi (nefes almada zorluk).
- Göğüste yanma veya basınç hissi.
- Kollara, sırtta, boyna, çeneye veya mideye vurabilecek rahatsızlık hissi.

Zamanla kalbinize kan sağlayan arterler yağ, kolesterol ve diğer faktörlerin birikmesi nedeniyle tıkanabilir. Bu birikintilere plak denir. Bir veya daha fazla atardamar kısmen tıkalıysa kan yeterli miktarda akamaz ve göğüste ağrı veya rahatsızlık hissedebilirsiniz. Anjina yaygındır. Göğüste anjina oluşur ve düzelirse, bu kalp hastalığı olasılığını gösterir ve tedavi gerektirir. Anjina, kalbin size daha fazla oksijene ihtiyacı olduğunu söyleme şeklidir. Anjina türleri aşağıdakileri içerir;

- **Stabil (Kararlı) Anjina**

Stabil anjina sıklıkla fiziksel aktivite sırasında ortaya çıkar. Merdivenden çıktığınızda veya ağır fiziksel aktivite yaptığınızda kalbin oksijen ihtiyacı artar ancak kalbi besleyen damarların incilmesi nedeniyle kalbe yeterli kan sağlanamaz. Fiziksel aktivitenin yanı sıra, oksijen ihtiyacını ve talebini artıran diğer bazı

durumlar şunlardır: sinirsel stres, soğuk hava, çok miktarda yiyecek tüketmek ve sigara içmek. (Bunların sonucunda göğüs ağrısı yaşanabilir.)

- **Anstabil (Kararsız) Anjina**

Koroner arterlerin içindeki yağ içeren plaklar yırtıldığında ve üzerinde kan pıhtısı oluştuğunda, kalp kasına giden kan akışı hızla durur. Kararsız anjinanın bir başka nedeni de, kalbi besleyen kan damarının bir kısmını tıkayan pıhtılardır. Bu tam bir tıkanıklık değildir.

Kan uzun süre kalp kasına ulaşmazsa kalp hücreleri ölür ve kalp krizi adı verilen olay meydana gelir. Kararsız anjina tehlikeli bir durumdur ve acil tedavi gerektirir.

- **Prinzmetal (Değişken) Anjina**

Prinzmetal anjina olarak da adlandırılan bu anjina, sıklıkla kalp kasını besleyen kan damarlarının spazmı sonucu ortaya çıkar ve geçici daralmaya neden olur. Bu geçici daralma, kalp kasına giden kan akışında geçici bir azalmaya neden olur ve hatta istirahat halinde bile ortaya çıkabilir ve sıklıkla şiddetlidir. Prinzmetal anjina tedaviyle düzelebilir (Dr. Najafi, 2015).

2.5. Kalp Hastalıklarını Teşhis Etmek İçin Yapılan Çalışmaların Özeti

S. Kumar ve diğer meslektaşlarının 2020'deki makalesi, kalp hastalıklarının rastgele orman algoritması kullanılarak tahmin edilmesi ve bir özellik alt kümesinin seçilmesi başlığını taşıyor. Kalp hastalıkları dünya çapında gerçekleşen ölümlerin önde gelen nedenidir. Kalp hastalığı kentsel ve kırsal alanlarda bir numaralı katildir. Bir hastalığın sonucunu tahmin etmek zorlu bir iştir. Veri madenciliği, teşhis kurallarını otomatik olarak türetmek ve uzmanların daha ileri teşhis sürecini sağlamasına yardımcı olmak için kullanılabilir. Sağlık çalışanlarının kalp hastalıklarını tahmin etmelerine yardımcı olmak için araştırmacılar tarafından çeşitli veri madenciliği yöntemleri kullanılmaktadır. Rastgele orman algoritması, tıbbi uygulamalar için çok uygun olan en doğru makine öğrenme algoritmalarından biridir. İkinci dereceden özellik seçim kriteri, değişkenler arasında değerlendirme yapmak ve bunların ilişkili olup olmadığını belirlemek için kullanılır. Bu yazıda, kalp hastalığını tahmin etmek için rastgele orman ve ikinci dereceden özellik seçimini kullanan bir sınıflandırma modeli öneriyoruz. Yaklaşımımızı kalp hastalığı veri seti üzerinde değerlendiriyoruz. Deneysel sonuçlar, yöntemimizin diğer sınıflandırma yaklaşımlarına kıyasla sınıflandırma doğruluğunu

arttırdığını ve sunulan modelin sağlık profesyonellerinin kalp hastalıklarını tahmin etmelerine yardımcı olabileceğini göstermektedir (Kumar & Ann Romalt, 2020).

Sujania Chatterjee, Santhosh Kumar ve diğer meslektaşları tarafından 2016 yılında yapılan araştırmada, sigarayı bırakırken sürekli kullanım isteğinin olduğu tahmin edilmiştir. Bu makale doğal ortamdaki sensörlerden elde edilen verilere dayanmaktadır. Burada, cep telefonu sensör verilerini kullanarak sigara içme dürtüsünü (sigarayı bırakma sırasında) dakika düzeyinde tahmin etmek için hesaplamalı bir model oluşturma yolunda ilk adım atılıyor. Bu araştırmada, 2012 yılından bu yana, sigarayı bırakma araştırmasında yer alan 61 katılımcının saatlerce süren sensör verilerini ve sigara içme isteğine ilişkin 1812 raporu kullanılmıştır. Tutkuyu tahmin etmek için öncelikle sürekli bir stres ölçeği elde edilir ve bunun için sensör verileri kullanılır. Aslında bu stres, tüketme isteğini gösterir. Bu araştırma, tutkuyu tahmin etmek için sürekli olarak cep telefonlarının fizyolojik sensörlerini kullanan hesaplamalı bir model sunmaktadır. M-Crio'da açıklanabilir bir model kullanmak, olaylarla ilgili içgörülerin korunmasına ve alan uzmanlarından onay alınmasına yardımcı olur. Tutkunun tahmini için otomatik bir fizibilite çalışması oluşturmak, sağlık ve sağlıklı yaşamı iyileştirme konusunda önemli potansiyele sahip birçok heyecan verici araştırma fırsatının önünü açar (S.chatterjee et.al. 2016).

Amin Ahsan Ali, Santosh Kumar ve diğer meslektaşlarının 2012'deki makalesi, sigara içmenin solunum ölçümleri kullanılarak sigara dumanının neden olduğu Nefes kesintilerinin otomatik tespiti başlığını taşıyor. Bu makalede, Amerika Birleşik Devletleri'nde son beş yılda en önemli ölüm nedenlerinden birinin araştırmadaki neden olduğu kesin olarak kanıtlanmıştır. Sigarayı bırakma programlarının geliştirilmesi alanında kapsamlı ve etkili araştırmalar yapılmaktadır.

Çoğu sigarayı bırakma programının başarısı çok azdır. Çünkü, doğru zamanda müdahale edilememektedir. Sigara içen kişiyi sigara içme isteğine maruz bırakabilecek tehlikeli durumların belirlenmesi, sigarayı bırakmadan önce gelen farklı koşullar arasındaki bağlantıların keşfedilmesini içerir. Bu çalışmada, nefes ölçümleri kullanarak sigara içmeyi tespit etmeye yönelik yeni ve otomatik bir sistem olan M-Puff'ı ve bu bağlamda sigara içmeyi tespit etmek için geliştirilen model sunulmaktadır. Modeller, her gün sigara içen 10 kişiden toplanan nefes ölçüm verilerine dayanmaktadır ve tek bir sigara içme seansında sigara dumanının %91 doğrulukla

tespit edilebildiğini göstermektedir. Daha sonra stres, konuşma ve yürüme gibi diğer olaylardaki nefes ölçümleri dikkate alınmış ve önerilen modelin sigara içmenin etkilerini hala %86,7 doğrulukla tespit edebildiği gösterilmiştir. Sigara tespit yöntemlerini kullanarak, sigara içme ve cep telefonu kullanımı sırasında meydana gelen etkili müdahaleleri tespit etme şansına sahip olunabilir ve bu sayede ideal sonuçlara ulaşabilir (A. A. Ali et al., 2012).

Nazir Salehin ve Santosh Kumar'ın 2016'da yazdığı makale, *Mobil Sensör Verilerinin Zaman Serilerinde Gizlilik* başlığını taşıyor. Mobil sensör bilgilerinin, özellikle de fizyolojik verilerin paylaşılması gizlilik zorluklarını artırır; bu da, sensör verilerinin zaman serilerinden ortaya çıkarılabilecek özel davranışlarının korunmasına neden olur. Bu araştırmada davranışsal bir gizlilik kriteri tanımlanmış ve davranışsal gizliliği korumak için yeni bir veri değiştirme mekanizması önerilmiştir. Programın etkinliği 43 katılımcıdan toplanan 660 saatlik EKG, solunum ve aktivite verilerini kullanarak değerlendirilmiştir. İncelenen makalede, doğal ortamda toplanan kişisel fizyolojik verilerin paylaşımını kolaylaştırmak için davranışların mahremiyetini korumak amacıyla ikameyi bir mekanizma olarak kullanan bir sistem olan MCV sunulmaktadır (N. Saleheen et al., 2016).

2.6. Gri Kurt Algoritmasının Yaptığı İşin Arka Özeti

Gri kurt algoritması, yüzdeki duygu tanıma için özellik seçme yöntemi, hastalık teşhisi için sinir sinyali sınıflandırması veya elektromiyografi (Electromyography), gen seçimi ve sızma tespit sistemleri olarak birçok alanda kullanılmıştır.

Vinod Vaze ve Praveen S. Gameve diğer meslektaşlarının 2019 tarihli makalesi, kalp hastalıklarının sınıflandırılmasına yönelik gri kurt eniyileycisini temel alan karar ağaçlarına bir giriş niteliğindedir. Kalp hastalıklarının sınıflandırılması temel bir sorundur. Çünkü, bu hastalık dünya genelinde kanser kadar ölüme neden olmaktadır. Bu çalışmada önerilen yöntem, Long Beach VA veri setini sınıflandırmak için gri kurt algoritmasını ID3 karar ağacı algoritmasıyla birleştirmektedir. Sonuçlar, veri kümesi bölümlenmesi kullanıldığında lojistik model ağaçlarından ve rastgele ağaçlardan daha iyi bir performans sergilemektedir. Dolayısıyla, bu modelin performansını artırmak için geliştirilebileceğini göstermektedir. Gelecekte, sınıflandırma doğruluğunu artırmak amacıyla özelliklerin daha iyi seçilmesi için diğer ön işleme yöntemleri kullanılacaktır (P. S. Game et al., 2019).

Seyed Ali Mirjalili, Qasem Al-Tashi ve diğer meslektaşlarının 2019 tarihli *Gri Kurt Optimizasyonuna Genel Bakış* başlıklı makalesi, sınıflandırma için özellik seçim yöntemlerine dayanmaktadır. Bu makalede, gri kurt algoritmasına dayalı öznelik seçimiminin kapsamlı bir incelemesi sunulmaktadır. Özellik seçimi ve yaklaşımları anlatılmıştır. Gri kurt algoritmasının matematiksel modeli ve gri kurt algoritmasını temel alan özellik seçme mekanizması da burada sunulmuştur. Gri kurt algoritmasının özellik seçiminde gösterdiği mükemmel performans nedeniyle birçok araştırma bu yöntemi kullanılarak yapılabilir. Gri kurt algoritmasının kullanımını geliştirmek amacıyla özellik seçimi problemlerinin çözümünde ikili operatörlerin gri kurt algoritması ile kombinasyonunun kullanılması önerilmektedir. Önerilen diğer bir yöntem ise gri kurt algoritması ile birlikte kullanılan sarmalayıcı ve filtre özelliği seçim yöntemlerinin birleşimidir. Ancak özellik seçimi, özellik sayısını ve sınıflandırma hata oranını azaltmayı amaçlayan çok amaçlı bir işlemdir (Al-Tashi et al., 2019).

Qasem Al-Tashi ve arkadaşlarının, 2019 makalesi, *Koroner Kalp Hastalığının Sınıflandırılması İçin Gri Kurt Optimizasyonuna Dayalı Özellik Seçme Yöntemi* başlığını taşımaktadır. Bu çalışmada gri kurt optimizasyonu ve destek vektör makinesini temel alan sarmalayıcı özellik seçim yöntemi sunulmaktadır. Önerilen yöntem, Cleveland veri setinden ilgili özellikleri seçme ve gereksiz ve ilgisiz özellikleri eleme yeteneğine sahiptir. Deney sonuçları, önerilen yöntemin doğruluk, duyarlılık ve diğer değerlendirme kriterleri açısından çok iyi çalıştığını ve diğer yöntemlere göre daha iyi performans sergilediğini göstermiştir. Öte yandan, sonuçları daha da güçlendirmek için gri kurt algoritmasına sahip farklı sınıflandırıcılar kullanılabilir (Q. Al-Tashi et al., 2019).

Reza Kamel ve diğer meslektaşlarının 2019'daki makalesi, *Meme Kanseri Tespitinin Doğruluğunu Artırmak İçin Gri Kurt Algoritmasını Kullanarak En İyi Özellikleri Seçerek Destek Vektör Makinesi (Support Vector Machine (SVM)) Algoritmasının Performansını Artırma* başlıklı makaledir. Meme kanseri tanısının doğruluğunu arttırmak amacıyla özellik seçimi ve veri sınıflandırması için gri kurt ve destek vektör makinesi algoritmaları kullanılmaktadır. Deneysel sonuçlar MATLAB (MATrix LABoratory) veri kümeleri ve UCI (University of California, Irvine) deposunu kullanılarak elde edilmiştir. En iyi sonuçlar, etkili özellik alt kümesi seçimi

için destek vektör makinesi algoritmaları ve gri kurt algoritmasının birleşiminden elde edilmiştir. Gelecekteki çalışmalar, kalp hastalıkları, diyabet ve diğer vakaların tanısında bu bakış açısını benimseyebilir (Kamel et al., 2019).

Prana Sharma, Shires Sundaram ve arkadaşlarının 2019 tarihli makalesi, *Islah Edilmiş Gri Kurt Algoritması Kullanılarak Parkinson Hastalığının Teşhisi* başlıklı makaledir. Bu makalede, Parkinson hastalığının semptomlarının erken aşamada belirlenmesine yardımcı olan iyileştirilmiş gri kurt algoritması (Modified Grey Wolf Optimization (MGWO)) sunulmaktadır. Parkinson hastalığı, zamanında tedavi edilmezse ölümcül olabilen bir hastalıktır. Bu çalışmada, özellik seçimi için bir arama stratejisi görevi gören geleneksel gri kurt eniyileyicisi kullanılarak değiştirilmiş gri kurt adında yeni bir model önerilmektedir. Gri kurt algoritması, kurtların avlanma davranışlarını takip eden meta-sezgisel bir algoritmadır. Önerilen model farklı ses türleri, el yazısı (sarmal ve labirent) ve konuşma veri kümelerini kullanarak değerlendirilmiştir. Önerilen algoritma, Parkinson hastalığını %94,83 doğrulukla ve %98,28 tespit oranıyla tahmin etmek için kullanılıyor. Deneysel sonuçlar, önerilen algoritmanın doğruluğu en üst düzeye çıkarmaya ve seçilen özelliklerin sayısını en aza indirmeye yardımcı olduğunu göstermektedir (Sharma et al., 2018).

Sankara Babu ve diğer meslektaşlarının 2018 yılında yazdığı makale, *Geri Dönüşümlü Sinir Ağlarına Dayalı Gri Kurt Optimizasyon Algoritması Ve Otomatik Kodlayıcı Kullanılarak Tıbbi Hastalıkların Tahmini* başlığını taşıyor. Tıp alanında birçok araştırmacı, çeşitli tıbbi uygulamaları kullanarak çeşitli hastalıkları tahmin etmeye odaklanmıştır. Bu araştırmada tıbbi hastalıkların tahmin edilmesinde gri kurt algoritması ve tekrarlayan sinir ağı (GWO + RNN) algoritmasının birleştirildiği bir yöntem kullanılmıştır. Gereksiz ve ilgisiz özellikleri ortadan kaldıran özelliğin seçilmesi için gri kurt algoritması yöntemi kullanılmaktadır. Önerilen yöntem UCI verileri üzerinde kullanılmış ve önerilen algoritmanın etkinliğinin başlangıç durumuna göre arttığı gözlemlenmiştir (Sankara Babu et al., 2018).

2019 yılında özellik seçme yöntemi olarak gri kurt algoritması ile çeşitli çalışmalar yapılmıştır. Abdul Bast ve arkadaşları tarafından yazılan, *Özellik Seçimi İçin Gri Kurt Optimizasyon Algoritmasına Dayanan Yüz Tanıma* başlıklı makale 2019 yılında yayınlanmıştır. Bu makalede, özellik seçimi için gri kurt algoritmasını ve sınıflandırma için K en yakın komşuyu kullanan etkili bir yüz tanıma sistemi tanıtılmış ve önerilen

yöntem Yale Face veritabanı kullanılarak değerlendirilmiştir. Ayrıca önerilen yöntemin performansının arttığını göstermek için performansı başka bir yüz tanıma sistemiyle karşılaştırılmıştır ve elde edilen sonuçların doğruluk ve yürütme süresi açısından daha iyi olduğu gösterilmiştir (Abd AL et al.,2018).

Gri Kurt Optimize Edici Algoritmasının Çoklu Stratejisi ve Özellik Seçimi Uygulaması başlıklı bir diğer makale, Chen X ve diğer meslektaşlarının 2019'da Gri Kurt Algoritmasının çeşitli stratejilerinin önerildiği çalışmasıdır. Bu çalışmada, gri kurt algoritmasının yerel arama kabiliyetini genişletmek için ilk küresel iyileştirme ile birlikte çözümleri iyileştirmek amacıyla güncellenmiş üç arama yöntemi sunulmaktadır. İkinci ayarlanabilir yaklaşım, bir vektörün boyutlarının gri kurt algoritmasının çalışma çerçevesine yerleştirilmesidir. Son olarak, ayar parametresine dayalı arama dağılım stratejisi tanıtılmaktadır. Deneysel sonuçlar, önerilen yöntemin yakınsama hızı ve doğruluk açısından daha iyi performans sergilediğini göstermektedir (Tu Q et al., 2019).

2.7. Kalp Hastalıklarının Teşhisine Yönelik Bazı Uygulamaların Tanımı

- **Google Fit platformu**

Google Fit, çoğu modern Android cihazında yerleşik sensörleri kullanarak aktivite ve egzersizi tek bir platformda izleyen bir Android uygulamasıdır. Bu uygulama, kullanıcıların fitness ve kilo verme hedeflerini belirlemelerine ve ilerlemeyi ve çeşitli egzersizleri ölçerek bu hedeflere ulaşmalarına olanak tanır. Google Fit, Google'ın Apple'ın mobil HealthKit'ine yanıt olarak yayınladığı bir uygulamadır. Bu uygulama, akıllı telefon sensörlerini kullanarak bulunduğu ortamda yapılan yürüyüş, koşu, bisiklete binme gibi spor aktivitelerinizi otomatik olarak kaydedip kayıt altına alır. Ayrıca Google Fit, gün, hafta ve ay boyunca kilo verme programını ve spor hedeflerini grafik üzerinde görüntüleyip takip edebilmektedir. Google Fit uygulaması aktivite ve sağlık takibi sloganıyla oluşturulmuştur. Sağlıklı kalmak için ne kadar veya ne tür bir aktiviteye ihtiyaç duyulduğunu kişinin kendisinin bilmesi zordur. Bu nedenle Google Fit, aktivitenin amacını, kişinin sağlığını iyileştirmeye yardımcı olacak kalp puanlarını sağlamak için Dünya Sağlık Örgütü ve Amerikan Kalp Derneği ile ortaklık kurmuştur. Kalbin yüksek devirde kan pompalamasını sağlayan aktivitelerin insan kalbi ve zihni için muazzam faydaları vardır. Adım hızını artırmak gibi orta dereceli aktivitelerin her

dakikası için bir kalp puanı kazanılırken koşma gibi daha yoğun aktiviteler için iki kat puan kazanılır. Haftada beş gün sadece 30 dakikalık tempolu yürüyüşle, Dünya Sağlık Örgütü ve Amerikan Kalp Derneği tarafından önerilen, fiziksel aktivite miktarına ulaşılır. Google Fit açık bir ekosistemdir ve geliştiricilerin sağlık verilerini, kullanıcıların çeşitli verilerine tek bir cihazdan erişebilecekleri merkezi bir veritabanına yüklemelerine olanak tanır. Bu program giyilebilir cihazların veya farklı sensörlerin verilerini kaydedebilir ve diğer programlara erişebilir.

- **M Cerbram Platformu**

M Cerbram, akıllı telefonlar ve giyilebilir sensörler için özelleştirilebilir bir yazılım platformudur. M Cerbram açık kaynaklı bir programdır ve güvenilir gerçek zamanlı veri toplama için destek sağlar. Bu yazılım ortamı tamamen özelleştirilebilir ve saniyede yüzlerce veri noktasında geniş bir veri yelpazesi toplamak için aynı anda birden fazla sensörle kullanılabilir.

Bu platform, yüksek frekansta bile gelişmiş analizleri kullanarak sensör verilerini sürekli olarak sağlık göstergelerine dönüştürür. Bu uygulama davranış ve risk faktörlerini (stres, sigara içme, yemek yeme ve aktivite gibi) inceler.

Çoklu duyuşal biyobelirteçlerin geliştirilmesi ve doğrulanması, ham sensör verilerinin ve ilgili etiketlerin bu ortamda toplanmasını gerektirir. Şu anda M Cerbram bu tür çalışmaların yürütülmesine yönelik genel amaçlı bir mobil yazılım platformudur. Bu platformun bilişsel araçları sınırlandırması en belirgin dezavantajıdır. Bileklik seçiminin kısıtlı olması ve bu platformun piyasadaki markalarla uyumlu olmaması bu platformun diğer dezavantajlarıdır. Bu platformun Microsoft markalı bilekliklerle uyumlu olması ve bu bilekliklerin üretiminin durdurulmuş olması da bu platformun önemli bir dezavantajıdır. Mevcut Microsoft marka bilekliklerin sınırlı sayıda olması ve bu bilekliklere erişim sağlanamaması, bu platformun kullanımının sınırlı kalmasına neden olmuştur. Yapısı değiştirilerek ve piyasadaki diğer marka bilekliklerle uyumu sağlanarak bu platformun kullanımı yaygınlaştırılabilir.

2.8. Sonuç

Bu bölümde kalp hastalıklarının tanısında kullanılan temel kavram ve faktörler ile koroner kalp hastalığının tanıtımı ve bu hastalığın tanısında kullanılan bazı belirtiler tartışılmıştır. Ardından, Google Fit ve M Cerbram gibi öne çıkan bazı platformlara değinilmiştir. Bir sonraki bölümde kalp hastalıklarını sınıflandırmak ve teşhis etmek için rastgele orman adı verilen makine öğrenmesi algoritmalarından biri kullanılacaktır. Bölümün sonunda kalp hastalıklarının tanısı ve gri kurt algoritmasının diğer makine öğrenmesi algoritmalarıyla kombinasyonu ile ilgili daha önce yapılan bazı çalışmaların incelemesi yapılacaktır.

3. ÖNERİLEN ALGORİTMA VE MOBİL SAĞLIĞA DAYALI PLATFORMLARA YÖNELİK AÇIKLAMALAR

3.1. Giriş

Son yıllarda veri madenciliği, bilgi teknolojisi araştırmalarının hızla büyüyen bir alanı olarak kabul edilmektedir. Bunun nedeni, günlük olarak büyük miktarda veri toplanması ve bu verilerin yararlı bilgilere dönüştürülmesi ihtiyacıdır. Veri madenciliği, kalıpların keşfedildiği ve büyük bir veri kümesinden bilginin çıkarıldığı bir süreçtir. Veri madenciliği görevleri arasında korelasyon analizi, anormallik tespiti, kümeleme, regresyon ve sınıflandırma yer alır. Özellik seçimi (Feature selection), bir veri setindeki gereksiz ve ilgisiz özelliklerin ortadan kaldırılması amacıyla ön işlemenin ana aşamalarından biridir. Yüksek boyutlu verilerin gereksiz ve ilgisiz özelliklere sahip olduğu durumlarda, makine öğrenimi ve veri madenciliğinde özellik seçimi gereklidir. Özellik seçimi, sınıflandırma performansını artırmak ve hesaplama maliyetini azaltmak için veri boyutlarının küçültülmesi ve ilgili en önemli özelliklerin seçilmesiyle kullanılır. Bu, üstel bir büyümeye sahiptir ve bu da onu özellikle çok sayıda özelliğe sahip veri kümeleri için zorlaştırır. Bu sorunu çözmek için çok çeşitli optimizasyon algoritmaları kullanılır ve gri kurt eniyileyicisi bu son örneklerden biridir. Bu algoritma bu bölümde ele alınacak meta-sezgisel algoritmalarından biridir. Bazı bilim adamları tarafından sınıflandırılan üç özellik seçme yöntemi kategorisi; paketleme, filtreleme ve gömülü yöntemleri (Embedded) içerir.

Özellikleri seçerek modelleri öğrenmek ve tahmin etmek için kullanılan k en yakın komşu algoritmaları, destek vektör makineleri ve yapay sinir ağları vb. gibi çeşitli sınıflandırma algoritmaları vardır.

Besin kaynaklarının bulunmasında doğadan veya hayvanlarla etkileşimden elde edilen optimizasyon yaklaşımlarının gelişmesi talebi artırmıştır. Biyoloji tabanlı kolektif kalabalık zekası kullanılarak çeşitli optimizasyon teknikleri geliştirilmiştir. 2014 yılında Mirjalili ve meslektaşları gri kurt adı verilen bir optimizasyon tekniğini sundular. Bu yöntem temel olarak gri kurdun avını avlamanın en iyi yolunu bulma mekanizması tarafından yönlendirilir. Gri kurt algoritması temel olarak kurtların farklı kişiliklerini organize etmek için "doğanın mekanizmasını" kullanır.. İleride gri kurt algoritması detaylı olarak incelenecektir.

3.2. Öznitelik Seçimi

Özellik seçimi, yüksek boyutlu veri kümelerinin gereksiz, marjinal ve alakasız özellikler içerdiği makine öğrenimi ve veri madenciliği alanında önemli bir rol oynar. Bu nedenle özellik seçimi, veri boyutluluğunu azaltmayı ve sınıflandırma performansını artırmak ve hesaplama masraflarını azaltmak için yalnızca en uygun özellikleri seçmeyi amaçlamaktadır. Özellik sayısı çok olduğunda arama alanı da çok büyük olduğundan, arzu edilen özelliği bulma sorunu karmaşık bir hal alır.. Özellik seçiminin temel amacı, belirli bir problem alanından yüksek sınıflandırma performansını yansıtan küçük bir özellik alt kümesini belirlemektir. Genel olarak özellik seçimi birkaç nedenden dolayı kullanılır. Birincisi, kullanıcıların, özellikle araştırmacıların, verilerin kullanımını ve yorumlanmasını kolaylaştırmak amacıyla veri basitleştirmesidir. İkincisi, optimum özellik seçimine ulaşmak için yalnızca en önemli işleme özelliğini seçerek daha az zaman harcamaktır.

Aslında yanıltıcı, ilgisiz ve gereksiz özellikler başarıyla kaldırılarak uygun özelliklerin üretilmesi sağlanır. Genellikle, özellik seçimi için filtreleme ve paketleme yöntemlerini içeren iki ortak algoritma kategorisi kullanılır. Temel farkları, özellik alt kümesinin değerlendirme aşamasında paketleme yaklaşımının aynı zamanda bir sınıflandırma algoritması içermesidir

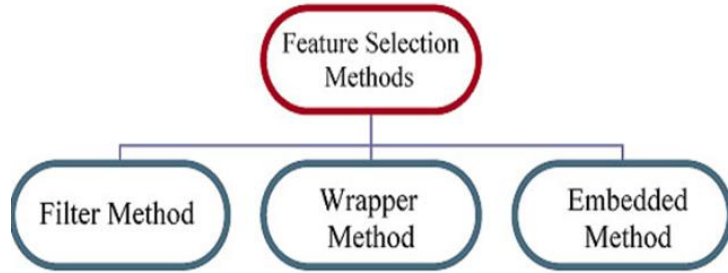
değerlendirmek için bir sınıflandırma algoritması kullanırken, filtreleme yöntemi için özellik seçim süreci, sınıflandırmaya yönelik herhangi bir algoritmadan bağımsızdır.

Paketleme yöntemlerinin hesaplama açısından filtreleme göre daha pahalı olduğu iddia edilebilir. Ancak filtre yönteminde, sınıflandırma algoritmasında seçilen özelliklerin iyiliğinin kontrol edilmesi ihmal edilirken, paketler, sınıflandırma kullanarak performanslarına göre bir özellik alt kümesini değerlendirir ve bu da genellikle paketleme yönteminin filtreleme göre daha iyi performans göstermesiyle sonuçlanır.

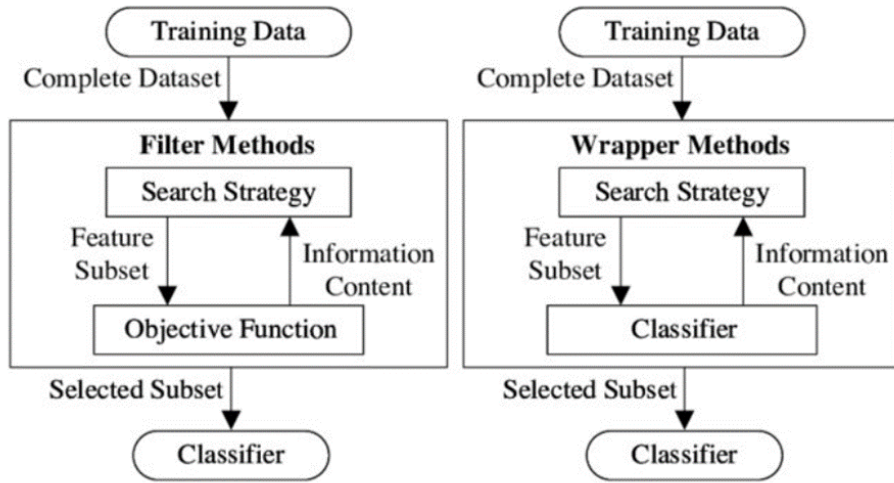
Ek olarak, bazı bilim adamları özellik seçme yöntemlerini üç sınıfa ayırmaktadırlar: filtreleme, paketleme ve gömülü yöntemler. Bu gömülü yöntemde hem seçilen özellikler hem de sınıflandırıcı tek bir süreçte birleştirilir. Şekil 3.1'de özellik seçim yöntemlerinin üç yönteme ayrıldığı görülmektedir.

Bunun yanı sıra Şekil 3.2 iki filtre ve paketleme yöntemi arasındaki karşılaştırmayı göstermektedir. Filtreleme yönteminde, arama stratejisi uygulandıktan sonra maliyet fonksiyonu bir özellik alt kümesine uygulanır. Ancak paketleme yönteminde, arama

stratejisi uygulandıktan sonra istenilen sınıflandırma algoritması bir özellik alt kümesine uygulanır (Barnali Sahu et al., 2018).



Şekil 3.1 Özellik seçme yöntemleri.



Şekil 3.2 İki filtre ve paketleme yönteminin karşılaştırılması.

3.3. Meta-Sezgisel Algoritmalar

Meta-sezgisel optimizasyon teknikleri son yirmi yılda çok popüler hale geldiler. Şaşırtıcı bir şekilde, genetik algoritma (Genetic Algorithm (GA)), karınca kolonisi optimizasyonu (Ant Colony Optimization (ACO)) ve parçacık sürüsü optimizasyonu (Particle Swarm Optimization (PSO)) gibi bazı yöntemler sadece bilgisayar bilimcileri arasında değil, farklı alanlardan bilim adamları tarafından da iyi bilinmektedir. Çok sayıda teorik çalışmanın yanı sıra bu tür optimizasyon teknikleri çeşitli akademik alanlarda da kullanılmaktadır.

Meta-sezgisel algoritmaların kullanımı önemli ölçüde arttı. Bu sorunun cevabını basitlik, esneklik, türetmeden mekanizma ve yerel optimizasyonun engellenmesi olmak üzere dört ana nedende özetleyebiliriz.

1. Meta-sezgisel algoritmalar nispeten basittir. Çoğunlukla çok basit konseptlerden ilham alırlar. Bu ilhamlar genellikle fiziksel olaylarla, hayvan davranışlarıyla veya evrimsel kavramlarla ilgilidirler. Bu basitlik, bilgisayar bilimcilerinin çeşitli doğal kavramları simüle etmesine, yeni meta-sezgisel algoritmalar önermesine, iki veya daha fazla meta-sezgisel algoritmayı birleştirmesine veya mevcut bir meta-sezgisel algoritmayı geliştirmesine olanak tanır. Bunun yanı sıra bu basitlik, diğer bilim adamlarının meta-sezgisel algoritmaları hızlı bir şekilde öğrenmelerine ve bunları problemlerine uygulamalarına yardımcı olur.
2. Esneklik, meta-sezgisel algoritmaların, algoritma yapısında belirli değişiklikler olmadan farklı problemlere uygulanmasını ifade eder. Meta-sezgisel algoritmaların çeşitli problemler için kullanımı kolaydır. Çünkü çoğu problemi kara kutu olarak kabul ederler. Başka bir deyişle, bir meta-sezgisel algoritma için bir sistemin yalnızca girdileri ve çıktıları önemlidir.
3. Çoğu meta-sezgisel algoritmanın türevsel olmayan mekanizmaları vardır. Gradyan tabanlı optimizasyon yaklaşımlarının aksine, meta-sezgisel algoritmalar sorunları rastgele optimize ederler. Optimizasyon süreci rastgele çözümlerle başlar ve en uygun çözümü bulmak için arama alanlarının türevini hesaplamaya gerek yoktur. Son olarak meta-sezgisel algoritmalar, geleneksel optimizasyon teknikleriyle karşılaştırıldığında yerel minimumlara takılmayı önleme konusunda üstün yeteneklere sahiptirler.

Gerçek problemlerin arama alanı genellikle bilinmemektedir. Üstelik çok karmaşık ve çok sayıda yerel minimuma sahiptir. Dolayısıyla, meta-sezgisel algoritmalar bu zorlu gerçek problemleri optimize etmek için iyi bir seçenektir (Barnali Sahu et al., 2018).

Popülasyon tabanlı meta-sezgisel algoritmaların en ilginç dallarından biri kalabalık zeka algoritmasıdır. Kalabalık zeka algoritması kavramları ilk kez 1993 yılında öne sürülmüştür (Beni G & Wang J, 1993). Kalabalık zeka (Swarm intelligence) algoritmalarının ilhamı çoğunlukla doğal kolonilerden, sürülerden ve popülasyonlardan gelmektedir. En popüler kalabalık zeka tekniklerinden bazıları karınca kolonisi algoritmalarını, parçacık sürüsü optimizasyonunu ve yapay arı kolonilerini içerir.

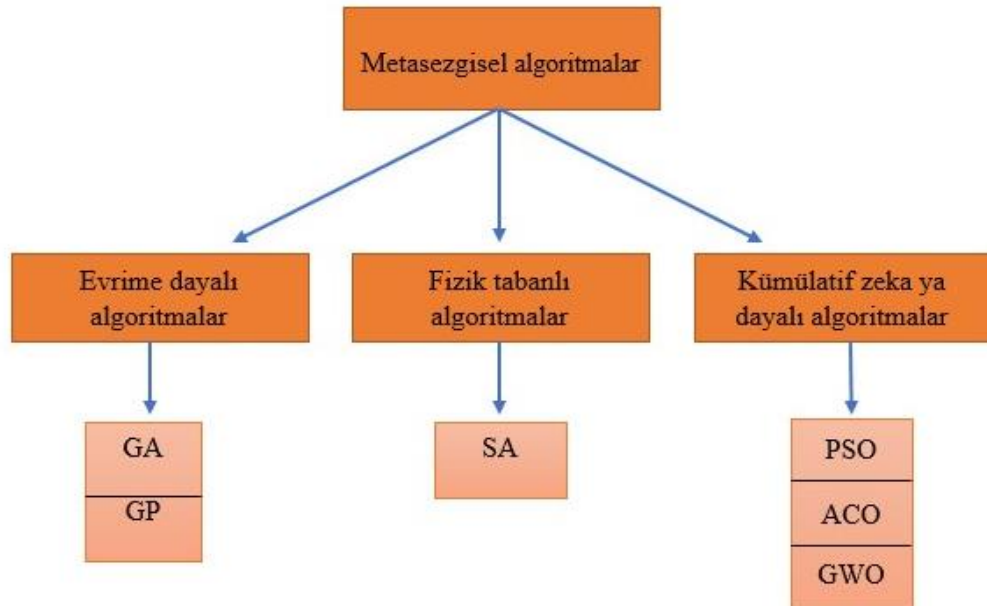
Kalabalık zeka algoritmalarının bazı avantajları şunlardan ibarettir:

- Tekrarlanma sırasında arama alanı hakkındaki bilgileri korurken, evrimsel algoritmalar önceki nesillerden gelen bilgileri atarlar.
- Elde edilen en iyi çözümü depolamak için sıklıkla belleği kullanmaktadır.
- Genellikle ayarlanması gereken daha az parametresi vardır.
- Evrimsel yaklaşımlara (mutasyon, elitizm vb.) göre daha az etkilidir.
- Uygulanması kolaydır.

Meta-sezgisel algoritmalar arasındaki farklara bakılmaksızın, arama süreci iki aşamaya ayrılır: keşif ve yararlanma.

Keşif aşaması, arama alanının gelecek vaat eden alanlarının araştırılması süreciyle ilgilidir ve mümkün olduğu kadar geniştir. Bu adımı desteklemek için, bir algoritmanın arama alanını rastgele ve global olarak aramak için rastgele operatörlere sahip olması gerekir.

Ancak kullanım, arama aşamasında elde edilen umut verici alanların çevresinde yerel olarak arama yapma yeteneğini ifade eder. Bu iki adım arasındaki doğru dengeyi bulmak, meta-sezgisel algoritmaların rastgele doğasından dolayı zorlu bir iştir. Meta-sezgisel algoritmalar genellikle üç ana sınıfa ayrılır ve evrimsel, fizik tabanlı ve kümülatif zeka algoritmalarını içerir, Şekil 3.3'te meta-sezgisel algoritmaların üç farklı grupta sınıflandırılmasını görebilirsiniz (Holland JH, 1992).



Şekil 3.3 Meta-sezgisel algoritmaların sınıflandırılması.

Evrimsel algoritmalar (Evolutionary Algorithm) genellikle doğadaki evrim kavramlarından ilham alırlar. Bu daldaki en popüler algoritma genetik algoritmadır. Bu algoritma 1992 yılında Holland tarafından önerilmiştir ve Darwinci evrim kavramlarını simüle etmektedir (Holland JH, 1992). Genetik algoritmanın mühendislik uygulamaları Goldberg tarafından kapsamlı bir şekilde incelenmiştir. Genel olarak optimizasyon, evrimsel algoritmalarda stokastik bir başlangıç çözümü geliştirilerek yapılır. Her yeni popülasyon, önceki nesildeki bireylerin birleşimi ve mutasyonuyla oluşturulur. Ayrıca, genetik algoritmaları ve genetik programlamayı içeren evrime dayalı algoritmalar da sürece dahil olurlar.

Meta-sezgisel algoritmanın ikinci ana dalı, benzetimli tavlama (Simulated Annealing (SA)) algoritmasını içeren fizik tabanlı tekniklerdir. Bu tür optimizasyon algoritmaları genellikle fiziksel yasaları takip etmektedirler. Bu algoritmaların mekanizması, rastgele bir arama kümesinin birbiriyle fiziksel yasalara göre iletişim kurması ve arama alanında hareket etmesi açısından evrimsel algoritmalarından farklıdır. Benzetimli tavlama algoritması, geniş arama alanlarındaki optimizasyon problemlerini çözmek için basit ve etkili bir meta-sezgisel optimizasyon algoritmasıdır. Benzetimli tavlama algoritması aslında çok sayıda değişkene sahip bir fonksiyonun minimize edilmesi problemini istatistiksel mekanik problemine indirgeyen bir yaklaşımdır. Algoritmanın kurucuları, zor optimizasyon problemlerini çözmek için kademeli ve yavaş tavlama tekniğine dayalı bir yöntem önermişlerdir. Bu algoritmadaki kademeli ve yavaş tavlama, cevap alanında arama sırasında daha kötü cevapların seçilme olasılığının kademeli olarak azalması olarak görüleb (S. Mirjalili et al., 2014).

Meta-sezgisel algoritmanın üçüncü ana dalı kalabalık zeka yöntemleridir. Bu algoritmalar çoğunlukla doğadaki grupların, sürülerin veya canlıların sosyal davranışlarını taklit ederler. Bu mekanizma neredeyse fizik tabanlı algoritmaya benzer, ancak arama araçları yaratıkların simüle edilmiş kolektif ve sosyal zekasını kullanarak gezinir.

En popüler kolektif zeka tekniği parçacık sürüsü optimizasyonudur. Parçacık sürüsü optimizasyon algoritması Kenedy ve Eberhart tarafından önerildi ve kuşların sosyal davranışlarından derlendi. Parçacık sürüsü optimizasyon algoritması, en iyi parçacığın konumunu ve o ana kadar elde edilen konumları takip eden birkaç parçacık kullanır.

Başka bir deyişle, bir parçacık hem kendi en iyi çözümü hem de kümenin en iyi çözümünü dikkate alarak hareket ettirilir (Kennedy J & Eberhart R.1995).

Bir diğer popüler kolektif zeka algoritması ise 2006 yılında Drigo ve arkadaşları tarafından önerilen karınca kolonisi optimizasyonudur. Bu algoritma, karınca kolonisindeki karıncaların sosyal davranışlarından esinlenmiştir. Aslında karıncaların yuva ile besin kaynağı arasındaki en kısa yolu bulma konusundaki sosyal zekası, karınca kolonisinin optimizasyonunun ana ilham kaynağıdır.

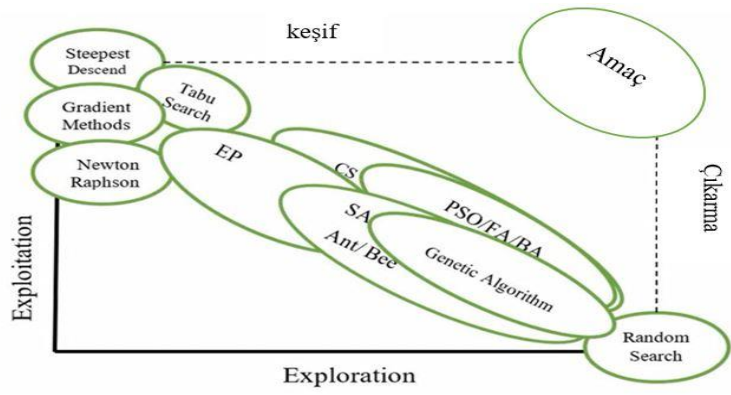
Bal arısı koloni algoritması (Artificial Bee Colony Algorithm (ABC)), arıların yiyecek kaynaklarını bulmadaki kolektif davranışlarını taklit eden bir diğer popüler algoritmadır. Arı kolonisi algoritmasında kaşif arılar, seyirci arılar ve işçi arılar olmak üzere üç tür arı vardır. Kaşif arılar, arama alanını keşfetmekle görevlendirilirken, gözücü ve işçi arılar, kâşif arıların bulduğu umut verici çözümleri kullanırlar.

Ayrıca son zamanlarda yarasaların konum davranışlarından esinlenerek yarasa davranışlarından ilham alan başka bir algoritma sunuldu. Doğada pek çok yarasa türü bulunmaktadır. Bunlar boyut ve ağırlık bakımından farklılık gösterirler, ancak yön bulma ve avlanma sırasında hepsi oldukça benzer davranırlar. Yarasalar bunun için doğal sonar kullanırlar. Yarasa algoritmasının (Bat Algorithm) tasarımında yarasaların av bulmadaki iki temel özelliği benimsenmiştir. Yarasalar avlarını kovalarken sesin şiddetini azaltma ve yayılan ultrasonik ses miktarını artırma eğilimindedirler. Bu davranış yarasa algoritması için matematiksel olarak modellenmiştir.

3.4. Meta-Sezgisel Algoritmelerde Keşif Ve Çıkarma Kavramı

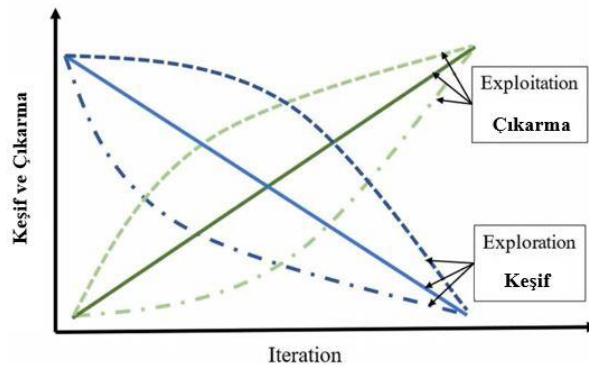
Keşif ve çıkarım, meta-sezgisel algoritmalarla optimizasyon yaklaşımında çok önemli iki kavramdır. Meta-sezgisel algoritmaların önemli özelliği ve temel rolü arama alanının ve arama yönteminin yönetimidir. Açgözlü algoritmalar olan ve başlangıç noktasına oldukça bağımlı olan gradyan azaltma algoritmaları (Steepest Descent) veya gradyan (Gradient Method) yöntemleri gibi yalnızca çıkarmanın olduğu algoritmalarda, uygun bir noktadan başlanırsa doğru ve uygun cevaba ulaşılırken, yanlış bir noktadan başlandığında da, yanlış cevaba ulaşılır. Yani, yerel minimumlara takılıp kalırlar. Meta-sezgisel algoritmanın sadece çıkarım kullanması halinde o algoritmanın açgözlü bir algoritmaya dönüşeceği söylenebilir. Çıkarma, bir başlangıç noktamızın olduğu ve yalnızca o noktanın yakınında arama yaptığımız anlamına gelir ve bu halde arama alanımız sınırlıdır.

Diğer taraftan keşif, yeni arayışlar ve buluşlarla ilgilidir. Bir algoritma yalnızca keşif kullanıyorsa, bu algoritma rastgele (Random Search) bir arama haline gelir. Bu arayışta hareket rastgele olduğu için doğru bir yol ya da hareket yoktur. Bu iki kavramı doğru bir şekilde bir arada kullanan ve her ikisinin uygun bir kombinasyonuna sahip olan bir algoritma doğru çalışır. Keşif ve madenciliğin tek bir merkezde nasıl bir araya getirildiği, farklı türdeki parçacık sürüsü optimizasyon algoritmaları, genetik algoritmalar, yarasalar vb. içeren algoritmaları birbirinden ayırır. Şekil 3.4'te meta-sezgisel algoritmalarındaki çıkarma ve keşif kavramları gösterilmektedir (S. Mirjalili et al., 2014).



Şekil 3.4 Meta-sezgisel algoritmalarındaki çıkarma ve keşfetme kavramı.

Şekil 3.5 ise, meta-sezgisel algoritmaların farklı yinelemelerinde keşif ve çıkarmanın rolünü göstermektedir. Probleme doğrusal olarak bakıldığında başlangıçta aşamasında problem çözmede keşfetmenin rolü daha fazla, çıkarmanın rolü ise daha azdır. Son yinelemelere yaklaşıldığında, bir yanıt son yanıt olarak bildirildiğinden çıkarmanın rolü de artar (S. Mirjalili et al., 2014).



Şekil 3.5 Meta-sezgisel algoritmalarındaki keşif ve çıkarma kavramı.

3.5. Gri Kurt Optimizasyon Algoritması

Gri kurt algoritması, gri kurtların davranışlarını, liderlik hiyerarşisini ve avlanma yöntemlerini taklit eden, doğadan ilham alan meta-sezgisel bir algoritmadır. Bu algoritma 2014 yılında Shahid Beheshti Üniversitesi'nden Mirjalili ve meslektaşları tarafından sunuldu. Aslında gri kurt optimizasyon algoritması, gri kurt avlama tarzından ilham alan meta-sezgisel bir optimizasyon yöntemidir. Bu algoritma, parçacık sürüsü optimizasyon algoritmaları, karınca kolonisi optimizasyonu gibi kolektif zeka algoritmalarının bir parçasıdır ve problem çözme alanındaki kurtların konumunu belirlemek için yalnızca bir (kombinasyon) operatör kullanır. Bu algoritma birçok öğrencinin ve araştırmacının dikkatini çeken güçlü bir optimizasyon aracıdır. Bu algoritmada, gri kurt sürüsünde liderlik hiyerarşisinin davranışını simüle etmek için alfa, beta, delta ve omega olmak üzere dört tür gri kurt kullanılmıştır. Bu algoritma popülasyon tabanlı, basit bir prosese sahip ve kolaylıkla büyük ölçekli problemlere genelleştirilebilir. Gri kurtlar besin zincirinin en üstünde yer alan yırtıcı hayvanlar (Apex) olarak kabul edilir. Gri kurtlar bir grup (sürü) halinde yaşamayı tercih ederler ve her grubun ortalama 5-12 üyesi vardır. Bu grubun tüm üyelerinin çok katı bir sosyal hakimiyet hiyerarşileri ve belirli görevleri vardır. Bu gruplar en tehlikeli avcılardan biridirler. Gri kurtlar avlanmak için önce avın etrafında daire çizerek etrafını sararlar ve kuşatma halkasını sıkılaştırarak avı yormaya başlarlar. Daha sonra, lider kurdun emriyle sırayla ava saldırırlar ve sonunda avı öldürürler.

3.6. Gri Kurtlarda Avlanma Yöntemi

Kurtların sosyal hiyerarşisine ek olarak toplu avlanma da gri kurtların bir başka ilginç sosyal davranışıdır. Gri kurt algoritmasının anlatımında ve eğitiminde bu algoritmanın sırasıyla 3 ana aşamadan oluştuğu söylenebilir:

- Avı gözlemlemek, takip etmek ve kovalamak (Şekil A)
- Avına yaklaşmak, etrafını sararmak (Çembere salmak) ve hareket etmeyi bırakıncaya kadar onu yoldan çıkarmak (Pursing and encircling) . Başka bir deyişle avı, hareket etmeyi bırakıncaya kadar kuşatmaya ve taciz etmeye çalışmak. (Şekil B-D)
- Ava saldırmak (Attacking) (Şekil E)

Gri Kurt algoritması da aynı süreçten esinlenmiştir. Şekil 3.6 gri kurtların bahsedilen üç aşamadaki davranışlarını göstermektedir (S. Mirjalili et al., 2014).



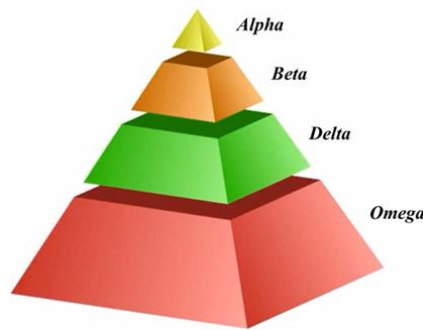
Şekil 3.6 Gri kurt avlanma şekli.

Gri kurtlar aşağıda belirtilen sosyal hiyerarşilerine göre yaşarlar. Şekil 3.7 aynı zamanda bu sosyal hiyerarşiyi de göstermektedir (S. Mirjalili et al., 2014).

1) Alfa 2) Beta 3) Delta 4) Omega

- İlk kategoride liderler, alfa olarak adlandırılan ve grubun lüzümen en güçlü üyeleri olmayan bir erkek ve bir kadından oluşur.
- Beta, emirlere itaat eder ve Alfa'nın emirlerini güçlendirir, grubun geri kalanına emirler verir ve danışman rolüne sahiptir.
- En düşük rütbe omegalara aittir, omegalara koruyucu rolüne sahiptir ve sürüde yemeği en son yiyenlerdir.

Bir kurt diğer üç kategoriye ait değilse delta kategorisine aittir. Yani; Yaşlılar, denetçiler, gözetmenler ve avcılar gibi.



Şekil 3.7 Gri kurtların sosyal hiyerarşisi.

3.7. Gri Kurt Algoritmasının Matematiksel Modellenmesi

GWO'yu sosyal hiyerarşi için tasarlarlarken:

- Maliyet fonksiyonu açısından değerlendirilen alfanın en iyi konumu (çözüm) dikkate alınır.

- İkinci en iyi çözüme beta denir.
- Üçüncü en iyi çözümün delta olduğunu varsayılır.
- Diğer çözümleri ise omega olarak algılanır.

Her yinelemede alfa, beta ve delta vardır. Her gri kurt bir arama faktörü olarak kabul edilir. Bu arama faktörleri maliyet fonksiyonuna göre konum açısından değerlendirilir. En iyi konum alfa'dır, daha sonra beta olur, sonra delta olur ve son olarak geriye kalan şey omega olur (S. Mirjalili et al., 2014).

3.8. Avın Kuşatılması İçin Matematiksel Denklemler

Av kuşatmasında kurtların konumunun güncellenmesi gerekir. Av sırasında gri kurtlar avın etrafını sarar. Kurtların kuşatma davranışını modellemek için aşağıdaki denklemler sunulmuştur.

Denklem (3.1), kurtların farklı yinelemelerdeki avı çevreleme konumuyla ilgilidir. Denklem (3.3) aynı zamanda A'nın hesaplamasını gösterir. A, hangi iterasyonda olduğuna göre değerlendirilir. İlişki (3.2) de önceki ilişkiden hesaplanır ve yem pozisyonuyla ilgilidir.

$$X(t + 1) = X_p(t) - A * D(t) \quad (3.1)$$

$$D(t) = |C * X_p(t) - X(t)| \quad (3.2)$$

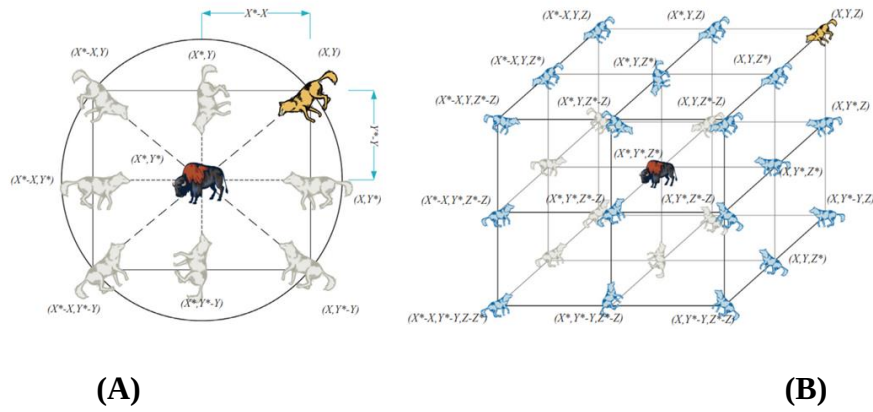
$$A = a(2r_1 - 1) \quad (3.3)$$

$$C = 2r_2, \quad r_1 \text{ \& } r_2 \text{ are random vectors } \in [-1,1] \quad (3.4)$$

burada t akış tekrarını temsil eder, A ve C katsayı vektörleridir, X_p av konum vektörüdür ve X gri kurt konum vektörünü temsil eder.

- $X(t)$: t'inci yinelemedeki gri kurdun konumu
- $X_p(t)$: t'inci yinelemedeki avın konumu
- a bileşeni doğrusal olarak 2'den sıfıra doğru azalır.

(X,Y) konumundaki gri kurt, avının (X^*,Y^*) konumuna göre konumunu güncelleyebilir. (X, Y) konumunda olan bir kurt, formüllerde rol oynayan A ve C rastgele parametrelerine göre bu durumlardan birinde yer alabilir. Şekil 3.8 iki boyutlu ve üç boyutlu konum vektörlerini ve kurtların olası konumlarını göstermektedir (S. Mirjalili et al., 2014).



Şekil 3.8 İki boyutlu (A) ve üç boyutlu (B) konum vektörleri ve olası konumlar.

3.9. Avlanma Denklemleri

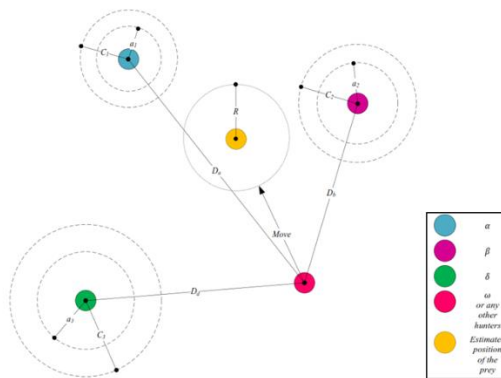
Son olarak gri kurdun konumunun nasıl güncelleneceği sorusu ortaya çıkıyor. Avlanma konumu veya $X_p(t)$ aslında sorunun çözümüdür. Gerçek durumda bu değer mevcut olmadığından varsayımsal durumlar dikkate alınmalıdır. Kurtların en iyi döngü durumu, avlanma için varsayımsal bir konum olarak kabul edilir. Çünkü gerçek durumda avlanma konumu veya $X_p(t)$ yoktur, bunun için bir alternatif olması gerekir. Bu nedenle avlanma pozisyonu α 'nın pozisyonu (elde edilen en iyi pozisyon) olarak kabul edilir (S. Mirjalili & S. M. Mirjalili, & A. Lewis, 2014).

$$X_1(t) = X_\alpha(t) - A_1 * D_\alpha(t) \quad , \quad D_\alpha(t) = |C_1 * X_\alpha(t) - X(t)| \quad (3.5)$$

$$X_2(t) = X_\beta(t) - A_2 * D_\beta(t) \quad , \quad D_\beta(t) = |C_2 * X_\beta(t) - X(t)| \quad (3.6)$$

$$X_3(t) = X_\delta(t) - A_3 * D_\delta(t) \quad , \quad D_\delta(t) = |C_3 * X_\delta(t) - X(t)| \quad (3.7)$$

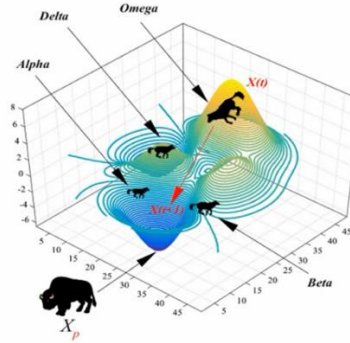
$$X(t+1) = \frac{(X_1(t) + X_2(t) + X_3(t))}{3} \quad (3.8)$$



Şekil 3.9 Gri kurt algoritmasında kurtların konumunun güncellenmesi.

Şekil 3.9 Omega kurdunun diğer kurtlara kıyasla etkinliğini göstermektedir (S. Mirjalili et al., 2014). Omega kurdu, alfa, beta ve delta kurtlarının konumuna ve sahip olduğu rastgele parametrelere göre hareket eder. Bu parametrelere göre Omega kurdunun konumu güncellenir. Aslında elde edilen yeni konum alfa, beta ve delta konumlarının bir sonucudur. Bu değer grafiksel olarak gösterilen avlanma denklemlerinin ifadesidir.

Şekil 3.10'da omega kurdunun konumu alfa, beta ve delta kurtlarının konumuna göre güncellenerek ava doğru hareket etmektedir. Aslında alfa, beta ve delta kurtları harekette iletken rolü oynar ve omega kurdunun daha iyi bir yere taşınmasına rehberlik ederler. Şekil 3.10 avlanma denklemlerini grafiksel olarak göstermektedir (S. Mirjalili et al., 2014).



Şekil 3.10 Kurtların konumu nasıl güncellenir ve ava doğru nasıl hareket edilir.

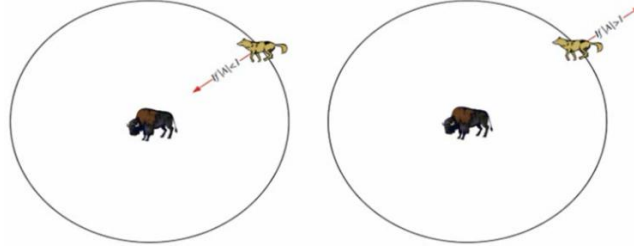
3.10. Av'a Erişim

Gri kurtların avına yönelik saldırısı, değeri $[-a, a]$ arasında olan rastgele bir vektör olan $\underline{a}$ vektörüne göre ifade edilir. $\underline{a}$ 'nın değeri 2'den 0'a yinelemeler sırasında doğrusal olarak azalır. Son olarak $\underline{a}$ vektörü, $\underline{a}$ parametresini güncelleyerek keşif ve kullanım arasındaki dengeyi kontrol eder. Her yinelemede bir parametre doğrusal olarak güncellenir (S. Mirjalili et al., 2014).

$$\vec{a} = 2 - t \cdot \frac{2}{\text{max_iter}} \quad (3.9)$$

- **max_iter**: Optimizasyon için izin verilen toplam yineleme sayısı
- **t**: yineleme sayısı

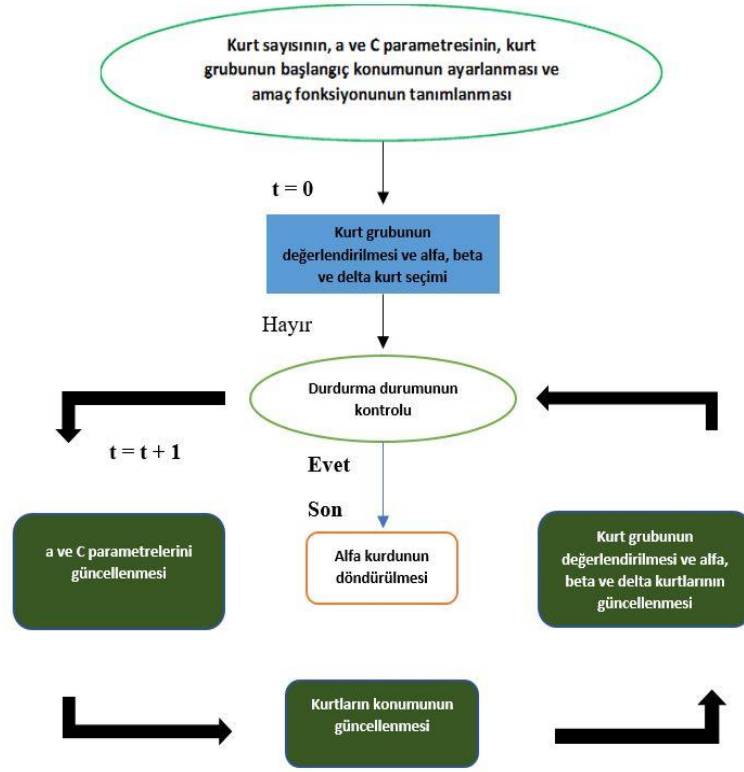
Şekil 3.12'de dairelerin yarıçapı da 1 olarak kabul edilir ve A 'nın mutlak değeri 1'den küçük olduğunda çıkarma rolü artar ve kurt ava doğru, hatta en iyi yönlendirme çözümüne doğru hareket eder. A 'nın mutlak değeri 1'den büyük olduğunda keşif modu oluşturulur ve kurt yeni keşiflere doğru hareket eder ve mutlaka en iyi çözüme doğru ilerlemez. Kurdun daire içine yerleştirilmesi tamamen rastgele olup A ve r parametrelerine dayanmaktadır (S. Mirjalili et al., 2014).



Şekil 3.12 Gri kurdun hareketinde A parametresinin rolü ve keşif ve çıkarma konuları.

3.12. Gri Kurt Algoritmasının Akış Şeması

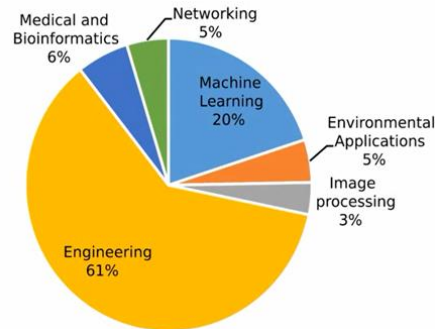
Şekil 3.13 gri kurt algoritmasının akış şemasını göstermektedir (S. Mirjalili et al., 2014). Bu şekle göre ilk önce başlatma değerlendirilmesi yapılır, daha sonra kurt sayısı ve a ile C rastgele parametreleri belirlenir. Kurt grubunun başlangıç konumu belirlenir. Daha sonra, değerleri rastgele atayan üniforma fonksiyonunu (Uniform) kullanılır. En önemli adım amaç fonksiyonunun (Objective function) veya maliyet fonksiyonunun (Cost function) belirlenmesidir. İkinci adımda amaç fonksiyonlarının hesaplanmasına göre her bir kurdun uyumu belirlenir. Ardından, alfa birinci en iyi, beta ikinci en iyi, delta üçüncü en iyi, geri kalanlar ise omega olarak kabul edilirler. Alfa, beta ve delta belirlendikten sonra algoritma ana döngüye döndürülür. Döngü için farklı durumlara sahip olabilen bir durma koşulu vardır. Örneğin en kolay ve en yaygın yollardan biri 100 iterasyona kadar çıkıp ardından durup en iyi sonucu döndürebilmesidir. Buradaki en iyi sonuç, finalde konumunu ve maliyetini gösteren alfa kurttur. İlk yinelemede a ve C parametreleri güncellenir. a parametresi 2'den 0'a doğrusal olarak hareket eder ve C , -2 ile 2 arasında rastgele bir parametredir. Daha sonra mevcut denklemlere göre kurtların konumları güncellenmektedir. Yapılan değerlendirmeden sonra alfa, beta ve delta konumları güncellenir. Son olarak durma koşulunu oluşturduktan sonra alfa kurdunun konumunu döndürür (S. Mirjalili et al., 2014).



Şekil 3.13 Gri Kurt Algoritmasının Akış Şeması.

3.13. Gri Kurt Algoritmasının Kullanımları

Gri kurt algoritmasının sunulduğu 2014 yılından bu yana bu algoritma çeşitli alanlarda kullanılmıştır. Aşağıdaki şekil farklı alanlardaki kullanım yüzdesini göstermektedir. Bu algoritmanın kullanıldığı alanlarda ilk sırayı yaklaşık %61'lik payla mühendislik alanı alırken, ikinci sırayı ise makine öğrenmesi konuları oluşturmaktadır. Gri kurt algoritması çok sayıda özelliğe sahip olan mühendislik işleri için oldukça pratik ve kullanışlıdır. Şekil 3.14'de aynı zamanda gri kurt algoritmasının farklı alanlardaki kullanımını göstermektedir (S. Mirjalili et al., 2014).



Şekil 3.14 Gri kurt algoritmasının farklı alanlardaki kullanımı.

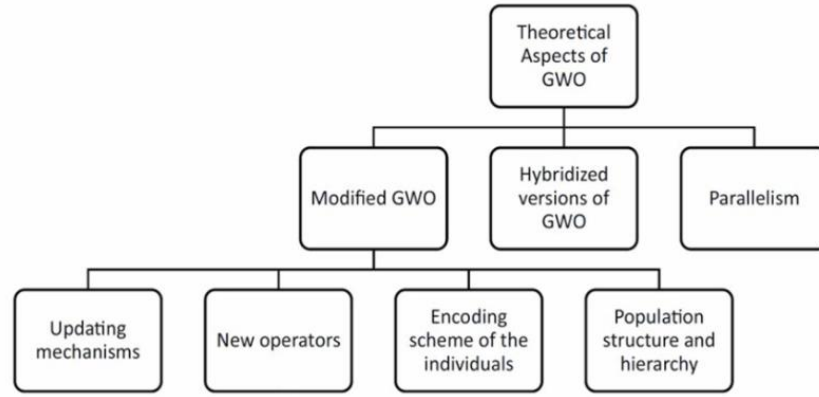
3.14. Gri Kurt Algoritmasının Teorik Yönleri

Bu bölümde algoritma üzerinde yapılan çalışmaların teorik olarak incelenmektedir. Bu çalışmalar aşağıdaki 3 gruba ayrılmıştır:

- 1) Algoritma paralelleştirme alanı
- 2) Gri kurt algoritmasının diğer meta-sezgisel algoritmalarla birleştirilmesi
- 3) Gri kurt algoritmasının kalbindeki küçük düzeltmeler

Paralel modda kurtları alfa, beta ve delta gruplarına sınıflandırırken grup sayısını artıracak ve o gruplarda da bu sınıflandırmalara sahip olacak şekilde algoritmayı hızlandırılmaya çalışılır.

Gri kurt algoritmasının diğer algoritmalarla birleştirilmesi kısmında gri kurt algoritmasının güçlü yönlerinden yararlanılmaktadır. Örneğin gri kurt algoritmasının genetik algoritma ile birleştirilmesiyle genetik algoritmanın verimliliği artırılmaktadır. Üçüncü kategoride gri kurt algoritmasının modifikasyonu ve değişimi ele alınmıştır. Bu reformlar arasında güncelleme mekanizmaları, yeni operatörlerin kullanılması, nüfus yapısı ve kalıtım vb. yer almaktadır. Şekil 3.15, gri kurt algoritmasının teorisinin bir görünümünü göstermektedir (Tu Q et al., 2019)



Şekil 3.15 Gri Kurt Algoritmasının Teorik Yönleri .

Karar ağaçları (Decision Tree) rastgele orman algoritmasının yapı taşları olduğundan, burada karar ağaçlarının kısa bir incelemesi yer almaktadır. Ayrıca birleşik sınıflandırıcı (Ensemble Classifier) boosting ve bagging konusu da tartışılmaktadır.

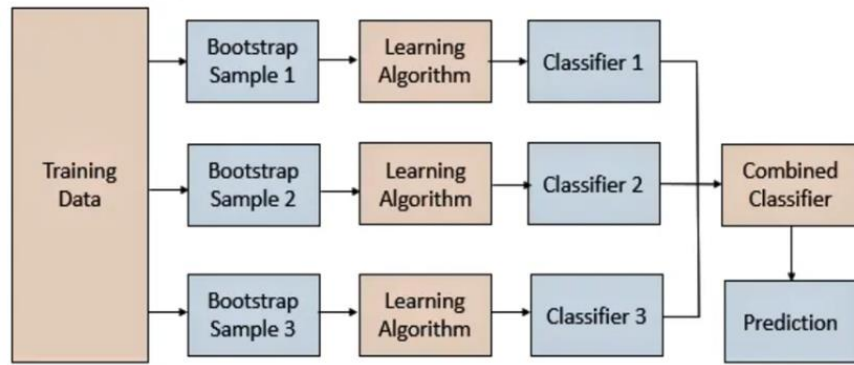
3.15. Birleşik Sınıflandırıcı Ve Boosting Ve Bagging Konuları

Hibrit sınıflandırıcılar çeşitli sınıflandırıcıların kombinasyonunu kullanırlar. Aslında bu sınıflandırıcıların her biri veriler üzerinde kendi modelini oluşturur ve bu modeli

saklar. Son olarak nihai sınıflandırma için bu sınıflandırıcılar arasında oylama yapılır ve en çok oyu alan sınıf, son sınıf olarak kabul edilir (Kaviani, 2020).

3.15.1. Bagging Yöntemi

Bagging yönteminde, sınıflandırıcıların her birine orijinal veri setinin bir alt kümesi verilir. Yani, her sınıflandırıcı veri setinin bir kısmını gözlemler ve kendisine sağlanan verinin aynı kısmına dayalı olarak modelini oluşturması beklenir. Şekil 3.16 sınıflandırma yönteminin bir görünümünü göstermektedir (Kaviani, 2020).



Şekil 3.16 Sınıflandırma yöntemi.

3.15.2. Boosting Yöntemi

Zayıf öğrenenlerin doğruluğunu artırmanın genel bir yöntemi yinelemeli bir süreçtir. Shapier 1990 yılında rastgele tahminlerden biraz daha iyi performans gösteren zayıf öğrenenlerin, birlikte güçlü ve doğru öğrenen bir sistem oluşturabileceklerini kanıtladı. Bagging yöntemi gibi boosting yöntemi de yeniden veri örnekleme ve çoğunluk oyu kullanarak öğrencilerin çıktılarını birleştirme avantajından yararlanır. Ancak Boosting yöntemi, yeniden örnekleme işlemini, öğrenmenin odağının daha zor örneklerle yönlendirilmesini sağlayacak şekilde gerçekleştirir.

3.16. Karar Ağacı Algoritması

Karar ağacı, veri madenciliği algoritmaları arasında en yaygın kullanılan algoritmalarından biridir. Bir karar ağacı, kökünden (yapraklarından) aşağıya doğru büyümesi dışında tıpkı bir ağaca benzer. Karar ağacı algoritmasında örnekler gruplandırılır ve aslında yaprak düğümlerin sonuna yerleştirilir. Karar ağacı, kategori veya sınıf adı şeklinde tek bir cevap verecek şekilde sunulabilen problemlerde kullanılır. ID3 algoritması karar ağaçları oluşturmaya yönelik temel algoritmalarından biridir. Karar ağaçlarını oluşturmaya yönelik ana algoritmaya Quinlan tarafından ID3

adı verilir. Bu algoritma, açgözlü yukarıdan aşağıya ve geri dönüşsüz olası dallar alanında açgözlü aramayı kullanır. ID3, bir karar ağacı oluşturmak için entropi (Entropy) ve bilgi kazancını (Information Gain) kullanır. Entropi, bir özellikten elde edilebilecek bilgi miktarıdır.

Yukarıdan aşağıya bir karar ağacı, bir kök düğümden oluşturulur ve verileri benzer (homojen) değerlere sahip örnekleri içeren alt kümelere bölmeyi içerir. ID3 algoritması bir numunenin homojenliğini hesaplamak için entropiyi kullanır. Örnek tamamen homojen ise entropi sıfırdır ve örnek eşit şekilde dağılmışsa entropi birdir. Karar ağaçlarının uyma olasılığı daha yüksektir. Veri kümesini bir özelliğe göre bölümlendirdikten sonra elde edilen kazanç (Gain), entropi azaltımına dayanır. Karar ağacı algoritmasında karar ağacı oluşturmak, bilgiden en yüksek faydayı sağlayan özelliğin (yani en homojen dalların) bulunmasıyla ilgilidir. ID3 algoritmasının görevi daha fazla bilgiye (daha fazla kâr) sahip özellikleri bulma ve bunları ağacın daha yüksek seviyelerine yerleştirmektir.

4.5C çıkarım algoritması Quinlan tarafından 1993 yılında hazırlanmıştır. Bu algoritma, hesaplamalarına sürekli ve ayrık değişkenleri dahil eder ve algoritmasında eksik değerleri dikkate alır. Bu algoritmanın mutlaka ikili olması gerekmez. Karar ağacının yolu boyunca en uygun ayırıcıyı seçmek için bilgi kazanımı veya entropi azaltma indeksini kullanır (Kaviani, 2020). 4.5C karar ağacı algoritmasının ID3 algoritmasına göre avantajları aşağıdaki şekilde değerlendirilebilir;

- 1) ID3 algoritması, karar ağacının oluşturulmasında yalnızca nominal özellikleri dikkate alır, ancak 4.5C algoritması, hem nominal hem de sayısal özellikleri dikkate alır.
- 2) ID3 algoritmasında öncelikle eksik verilerin giderilmesi gerekir. Ancak 4.5C, algoritmasındaki eksik verileri ile kendi karşılaşır.
- 3) 4.5C algoritması, ID3'ün aksine ağacı budar. Ağaç budama ağacın boyutunu ve karmaşıklığını azaltır. 4.5C karar ağacı algoritması, her nominal özellik değeri için varsayılan dalları ayırarak karar ağacının yoğunlaşmasına neden olur (Kaviani, 2020).

3.17. Rastgele Orman Algoritması

Rastgele orman, bir dizi ağaç sınıflandırıcının birleşimidir. Rastgele orman, sınıflandırma ve tahminde yaygın olarak kullanılmaktadır ve aynı zamanda

regresyonda da kullanılmaktadır. Rastgele ormanın uygulama alanı oldukça geniş ve büyük ölçülüdür.

1996 yılında Leo Breiman henüz başlangıç aşamasındaki algoritmalarından biri olan Bagging algoritmasını geliştirdi. Amit ve Goman çok sayıda geometrik özellik tanımlar ve bu durumlarda en iyi bölümü seçmek için rastgele bir seçim arar. Rastgele orman, hibrit bir makine öğrenme algoritmasıdır. Her ağaç bir dizi ağaç sınıflandırıcıyla birleştirilir. Her ağaç en popüler sınıfa bir oy verir ve ardından bu sonuçların birleştirilmesiyle nihai sonuç elde edilir. Rastgele orman algoritması yüksek sınıflandırma doğruluğuna sahiptir. Aykırı değerleri ve istenmeyen verileri (Noise) iyi tolere eder. Rastgele orman algoritması, biyolojik alanla ilgili veri madenciliği ve bilgi alanında en popüler araştırma yöntemlerinden biridir (Leo Breiman, 2001).

3.17.1. Rastgele Orman Algoritmasının Performansı

Rastgele orman, ağaç yapısına dayalı bir dizi sınıflandırıcıdan oluşan bir sınıflandırma algoritmasıdır. $\{h(x, \theta_k), K = 1, \dots\}, \{\theta_k\}$ bağımsızdır.

Rastgele vektörler aynı şekilde dağıtılır ve x girişindeki her ağaç, en popüler sınıfa bir oy verir. Bu tanım, rastgele orman algoritmasının ağaç yapısı formundaki birçok sınıflandırıcının birleşiminden oluştuğunu göstermektedir.

Breiman rastgele orman modelinde her ağaç, bir dizi eğitim örneğine ve bir rastgele değişkene dayalı olarak oluşturulur. K 'inci ağaca karşılık gelen rastgele değişken θ_k olarak gösterilir. K ardışık yürütme sonrasında $\{h_1(x), h_2(x), \dots, h_k(x)\}$ sınıflandırma dizisi elde edilir ve birden fazla sınıflandırma modeli sistemi oluşturmak için kullanılabilir. Bu sistemin nihai sonucu çoğunluk ile karşılaştırılır. Sistemin karar fonksiyonu aşağıdaki gibidir (Leo Breiman, 2001).

$$H(x) = \operatorname{argmax} \sum_{i=1}^k I(h_i(x) = Y) \quad (3.11)$$

$H(x)$ sınıflandırma modelinin bir birleşim, h_i bir karar ağacı modelidir, Y çıktı değişken ve $I(\cdot)$ indeks fonksiyonudur (Leo Breiman, 2001).

3.17.2. Rastgele Orman Algoritmasının Çalışma Yöntemi

Rastgele orman algoritması, basit ve zayıf algoritmalar için karar ağaçlarını kullanan hibrit bir algoritmadır. Rastgele ağaç algoritması, veri örnekleri üzerinde karar ağaçları

oluşturur ve bunların her birinden tahminler yapar ve sonunda oylama yoluyla en iyi çözümü seçer. Sonuçların ortalamasını alarak aşırı uyumu azalttığı için, tek bir karar ağacından daha iyi olan bir topluluk yöntemidir. Rastgele orman algoritmasında istenen adımlar şunlardır:

Birinci aşama – Bir veri kümesinden rastgele örnekler seçilir.

İkinci aşama - Her örnek için bir karar ağacı oluşturulur. Daha sonra her karar ağacından tahmin sonucu elde edilir.

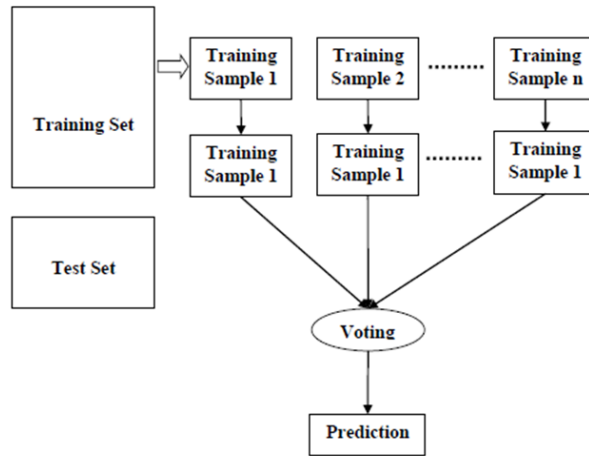
Üçüncü aşama - Bu adımda tahmin edilen her sonuç için oylama yapılır.

Dördüncü aşama, Nihai tahmin sonucu olarak en yüksek tahmin sonucu seçilir.

3.17.3. Rastgele Orman Algoritmasının zellikleri

- Güncel algoritmalar arasında doğruluk açısından benzersizdir.
- Çok büyük verilere uygulanabilir.
- Değişkenleri silmeden binlerce eğitim örneğini işleyebilir.

Şekil 3.17 rastgele orman algoritmasının performansı göstermektedir (Leo Breiman, 2001).



Şekil 3.17 Rastgele ormanın performansı.

Her birine bir veri alt kümesi enjekte edilmiş bir dizi karar ağacı vardır. Algoritmaların her biri öğrenme işlemlerini gerçekleştirir. Tahmin sırasında, yani tahmin için algoritmaya bir dizi yeni veri verildiğinde, öğrenilen bu algoritmaların her biri bir sonucu tahmin eder. Son olarak, rastgele orman algoritması en çok oyu alan sınıfı

oylama suretiyle seçebilir ve onu sınıflandırma için son sınıf olarak yerleştirebilir (Leo Breiman, 2001).

Algoritma 1- Gri Kurt Ve Rastgele Orman Algoritmasını Kullanan Sözde Kod Hibrit Yaklaşımı:

Initialize the population X_i ($i=1, 2, \dots, n$)

Initialize vectors a , A and C

Calculate the position of wolfs

X_α = the best search agent

X_β = the second best search agent

X_δ = the third best search agent

While ($t < L$) **do**

Update the position of wolfs

Update a , A and C

Calculate the positions (X_α , X_β , X_δ)

$t = t + 1$

end while

return X_α

Specify the total number of attributes N

Specify the length of the selected attributes L

Initialize α , β ($\alpha \in [0,1], \beta = 1 - \alpha$)

Calculate Fitness function

While ($t < N$)

$$Fitness = \alpha P + \beta \frac{L}{N}$$

Evaluate accuracy by using Random Forest algorithm (P)

Store the outcome of each decision tree

Compute the total votes for every tree

Declare majority class

$$(H(x) = \max \sum_{i=1}^k I(h_i(x) = Y))$$

end while

Önerilen sözde koda göre ilk önce başlatma işlemi yapılır, ardından (3.3) ve (3.4) numaralı ilişkilerde kurt sayısı ile $\underline{a}$ ve $\underline{C}$ rastgele parametreleri belirlenir. Kurt grubunun başlangıç konumu seçilir ve rastgele atanır. En önemli adım amaç fonksiyonunun (Objective function) veya maliyet fonksiyonunun (Cost function) belirlenmesidir. Bir sonraki adımda amaç fonksiyonlarının hesaplanmasına göre her bir kurdun uyumu belirlenir. Daha sonra alfa birinci en iyi olarak seçilir, ikinci en iyi beta ve üçüncü en iyi de delta olup geri kalanlar omega olarak kabul edilirler. Alfa, beta ve delta belirlendikten sonra ana döngüye döndürülür. Önerilen algoritma için dikkate alınan tekrar sayısı 50 tekrara eşittir. Buradaki en iyi sonuç, sonuçta konumunu ve maliyetini gösteren alfa kurdudur. İlk yinelemede $\underline{a}$ ve $\underline{C}$ parametreleri güncellenir. $\underline{a}$ parametresi 2'den 0'a doğrusal olarak hareket eder ve $\underline{C}$, -2 ile 2 arasında rastgele bir parametredir. Bundan sonra mevcut denklemlere göre kurtların konumları güncellenir. Daha sonra bunlar değerlendirilir ve alfa, beta ve deltanın konumu güncellenir. Son olarak tekrarlar bittikten sonra alfa kurdunun konumu geri döndürülür. Döngü durdurma koşulu, seçilen özelliklerin uygunluk değerinin optimal olduğu veya başka bir deyişle seçilen özelliklerin optimal özelliklerin bir alt kümesi olduğu durumun ortaya çıkması ile mümkündür. En iyi özellikler seçildikten sonra rastgele orman algoritması uygulanır. Bunun yanı sıra değerlendirme kriterleri de incelenerek kalp hastalıklarının tahmini ve tanısı yapılır.

3.18. Sonuç

Bu bölümde meta-sezgisel algoritmaların tanıtımı ve model türlerinin yanı sıra keşif ve çıkarma kavramı ile bunların meta-sezgisel algoritmalarda nasıl birleştirileceği anlatılmıştır. Daha sonra ise gri-kurt algoritması ve matematiksel modeli gösterilmiştir. Bir sonraki bölümde önerilen tasarım ve gri kurt algoritması kullanılarak öznelik seçim yöntemi incelenmiştir. Önerilen model aslında gri kurt algoritmaları ile rastgele orman algoritmasının birleşimidir.

Bu bölümde sunulan içeriklere göre meta-sezgisel algoritmaların yapısı ve matematiksel modellemesi ile keşif ve çıkarma kavramının bu algoritmadaki uygulamaları hakkında bilgi sahibi olmak mümkündür. Ayrıca platformların farklı yetenekleri ve giyilebilir sensörlerden nasıl bilgi alıp web sunucularına gönderilip analiz edileceğini anlamak da mümkündür.

4. ÖNERİLEN ALGORİTMA VE UYGULANMASI İLE ÇIKTI SONUÇLARININ GÖRÜNTÜLENMESİ

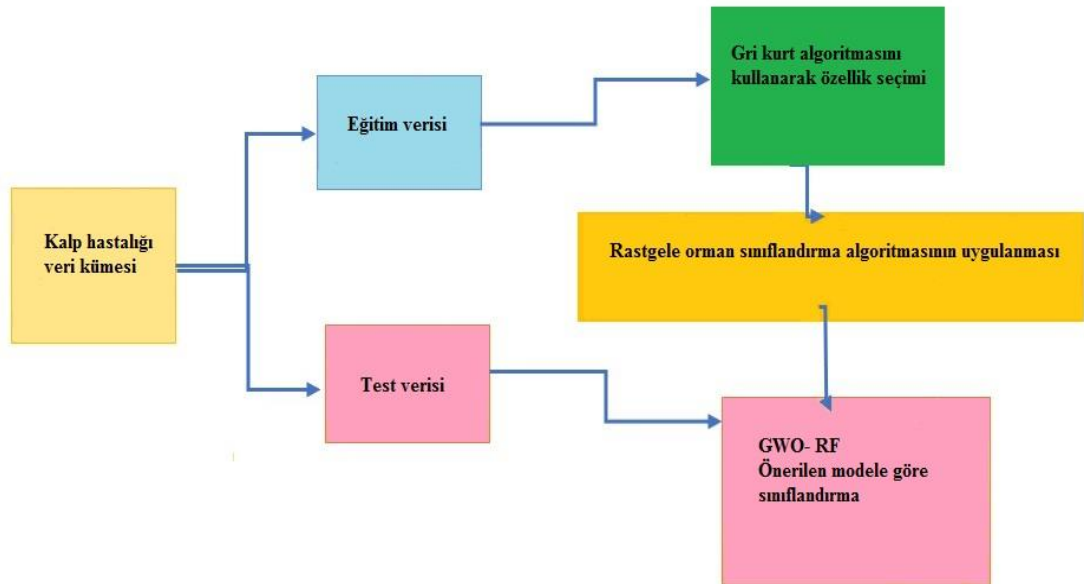
4.1. Giriş

Bu bölümde, öncelikle önerilen algoritma tanıtılmakta ve algoritmanın işleyişine genel bir bakış gösterilmektedir. Ayrıca önerilen algoritmanın performans artışı, Cleveland verileri üzerinde koroner arter hastalığı tanısı alanında yapılan çalışmalarla karşılaştırılarak bir tabloda tartışılmaktadır. Ardından iki gri kurt ve rastgele orman algoritmasının birleşiminden oluşan önerilen modelin uygulaması tartışılacaktır. Önerilen model şu şekildedir: öncelikle gri kurt algoritması kullanılarak özellik seçimi yapılır, daha sonra kalp hastalığı tanısı için seçilen özelliklere rastgele orman sınıflandırma algoritması uygulanır. Ayrıca önerilen modelin doğruluğu, duyarlılık ve f₁-score gibi diğer değerlendirme kriterleri, veriler üzerinde yalnızca rastgele ağaç algoritmasının kullanıldığı durumla karşılaştırılmıştır. Değerlendirme kriterleri üzerinde yapılan karşılaştırmada, önerdiğimiz model üzerinde kriterlerin doğruluğunun arttığını ve iyileştiğini göreceğiz. Tüm kriterler çıktılarıyla birlikte bu bölümde verilmektedir. Önerilen algoritmayı uygulamak için MATLAB yazılım sürümü R2020b kullanıldı. Bu tezde , biri 15 özellikli UCI deposundan alınan Cleveland verileri, diğeri ise verilerini internete gömülü çeşitli sensörlerden elde eden 14 özellikli M Cerbram platformuna ilişkin veriler olmak üzere iki farklı veri setinden yararlanılmıştır. Bu platformdaki bazı özellikler Framingham platformuna benzer. Bu bölümde öncelikle veriler ve özellikleri incelenmiş, daha sonra gri kurt ve rastgele orman hibrit algoritması için kullanılan fonksiyonların gerçekleşmesinin tanımı tartışılmıştır. Bu uygulamada kullanılan doğrulama tekniği çapraz doğrulamadır. Bu teknik, makine öğrenimi modellerinin verimliliğini değerlendirmek için kullanılır. Son olarak hibrit algoritmanın çıktıları rastgele orman algoritmasıyla karşılaştırılmıştır. Ayrıca performans karakteristik grafikleri (Receiver Operating Characteristic (ROC)) ve bunun altındaki seviye de görüntülenir.

4.2. Önerilen Algoritma

Bu çalışma, koroner arter hastalığını teşhis etmek için gri kurt algoritması ve rastgele orman algoritmalarının bir kombinasyonunu kullanarak özellik seçimi için etkili bir yöntem önermiştir. Bu ölümcül hastalığın nedenlerinin araştırılmasında büyük

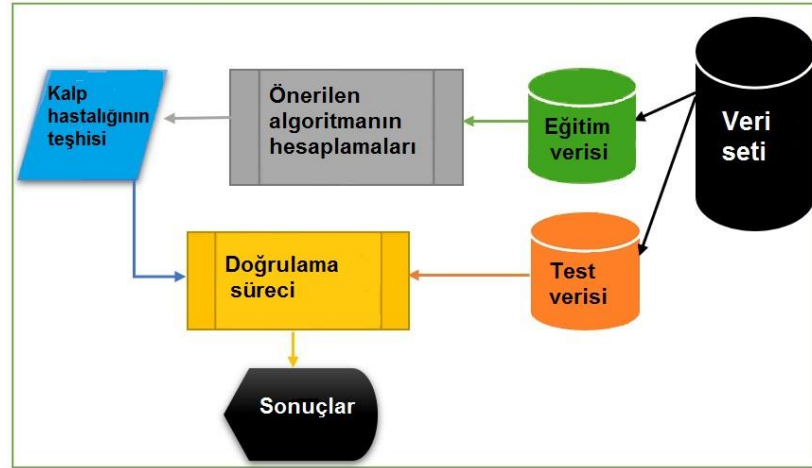
miktarda veri setinin analizi oldukça etkili olmuştur. Bu veri kümeleri, sınıflandırma doğruluğunu ve işlem hızını etkileyen gereksiz ve ilgisiz özellikler içerebilir. Bu nedenle söz konusu ilgisiz özelliklerin fazlalığının ortadan kaldırılması için özellik seçimi yönteminin kullanılması gerekmektedir. Hastalık tanısı için en uygun özellik alt kümesini belirlemek amacıyla bir özellik seçme yöntemi önerilmiştir. Önerilen bu yöntem, özellik seçimi ve sınıflandırmanın iki ana aşamasını içerir. İlk aşamada hastalıktaki en iyi özellikleri bulmak için gri kurt optimizasyonu kullanılır, ikinci aşamada kalp hastalığını sınıflandırmak ve tanımlamak için rastgele orman sınıflandırma algoritması uygulanır ve ardından önerilen algoritmanın performansı aşağıdaki yöntemler kullanılarak kontrol edilir. İlk adımda, gri kurt algoritması popülasyonun başlangıç konumlarını oluşturur ve daha sonra ayırık arama uzayındaki popülasyonun mevcut konumlarını günceller. İkinci aşamada, birinci aşamada elde edilen optimal özellik alt kümesine dayalı olarak oldukça etkili rastgele orman sınıflandırması gerçekleştirilir. Şekil 4.1 verilerin eğitim verilerine ve test verilerine nasıl bölüneceğini gösterir, ardından gri kurt algoritması kullanılarak özellik seçimi yapılır ve verilere rastgele orman sınıflandırma algoritması uygulanır (S. Mirjalili et al., 2014).



Şekil 4.1 Önerilen gri kurt ve rastgele orman modeline dayalı sınıflandırma sistemi.

Şekil 4.2 aynı zamanda önerilen algoritmanın uygulanması için istenen veri üzerindeki verilerin genel yapısını da göstermektedir(S. Mirjalili et al., 2014).

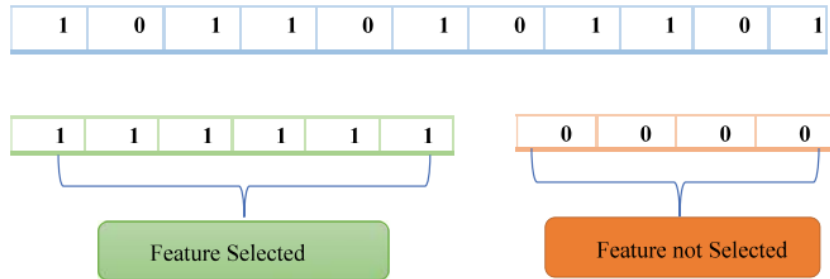
Bu şekilde göre veri seti iki kategoriye ayrılabilir: eğitim verileri ve test verileri. Eğitim veri seti üzerinde önerilen gri kurt ve rastgele orman algoritması uygulanmıştır. Önerilen bu algoritmayı kullanarak kalp hastalığı örüntüsü tanımlanır. Daha sonra bu konuda ayrıntılı olarak tartışılacak olan önerilen algoritmanın performansını kontrol etmek için değerlendirme kriterleri kullanılır.



Şekil 4.2 Önerilen gri kurt ve rastgele orman algoritması için veri madenciliği yapısı.

4.3. Gri Kurt Algoritmasının Özellik Seçme Mekanizması

Gri kurt özellik seçimi problemini çözmenin ana adımı, çözüm sunumunda özelliklerin bir alt kümesini görüntülemektir. Şekil 4.3 çözümün sunumunu göstermektedir. Çözüm uzunluğu d ile gösteriliyor. Burada d , özellikleri gösteren tamsayı değeridir. Çözüm konumu 1 veya 0 olabilir. Bit değeri 0 ise özellik seçilmez. Bit değeri 1'e eşitse özellik seçilir. 1 değeri alt küme özelliğinin boyutunu belirtir. Şekil 4.3 özellik seçimi çözümünün temsili göstermektedir (S. Mirjalili et al., 2014).



Şekil 4.3 Çözümün özellik seçiminde gösterimi.

4.4. Gri Kurt Algoritmasında Özellik Seçiminin Değerlendirilmesi

Bu adımda, gri kurt algoritması kullanılarak belirli bir sonucun önemli özelliklerini tanımlamak için özellik seçimi yapıldı. Daha sonra kriterlerin sınıflandırılması ve değerlendirilmesi için rastgele orman algoritması kullanılmıştır. Sınıflandırma performansının iyileştirilmesindeki temel amaç, gri kurt algoritmasını kullanarak özneliliklerin boyutlarını azaltarak tanıma doğruluğunu arttırmaktır.

Gri kurt algoritması, en iyi özelliği seçmek amacıyla özellik uzayını verimli bir şekilde aramak için kullanılır. En iyi özellik, sınıflandırma doğruluğu yüksek ve seçilen özellik sayısı daha az olan özelliktir. (4.1) Denklemin değerlendirme fonksiyonu, seçilen özellikleri değerlendirmek için gri kurt algoritmasında kullanılan sınıflandırma doğruluğunu en üst düzeye çıkarmak için kullanılır.

$$Fitness = \alpha P + \beta \frac{L}{N} \quad (4.1)$$

Burada P sınıflandırma doğruluğu, L seçilen özellik uzunluğu, N genel veri kümesindeki toplam özellik sayısıdır; burada alfa ve beta, sınıflandırma doğruluğu ağırlığı ve özellik seçim kalitesiyle ilgili parametrelerdir.

$\alpha \in [0,1]$, $\beta = 1 - \alpha$ rastgele orman sınıflandırması, denklem (4.1)'teki (P) değerini hesaplamak için kullanılır.

Gri kurt algoritmasını kullanarak bir özellik seçme problemini çözmenin ilk ve en önemli adımı, çözüm gösteriminde bir özellik alt kümesini temsil etmektir. Şekil 4.4 çözümün görüntüsünü göstermektedir. Önerilen özellik seçme yöntemi için, bir özellik alt kümesini temsil etmek için ikili bir kromozom kullanılır. Kromozom uzunluğu d olarak gösterilir; burada d, özelliklerin toplam sayısını temsil eder. Kromozom konumu 1 veya 0 değerini alabilir. Eğer i bitinin değeri bire eşitse özellik seçilir. Aksi takdirde bu özellik seçilmez ($i = 1,2,\dots$). Dolayısıyla değerleri aynı olan bit sayısı, bir özellik alt kümesinin boyutunu gösterir (S. Mirjalili et al., 2014).



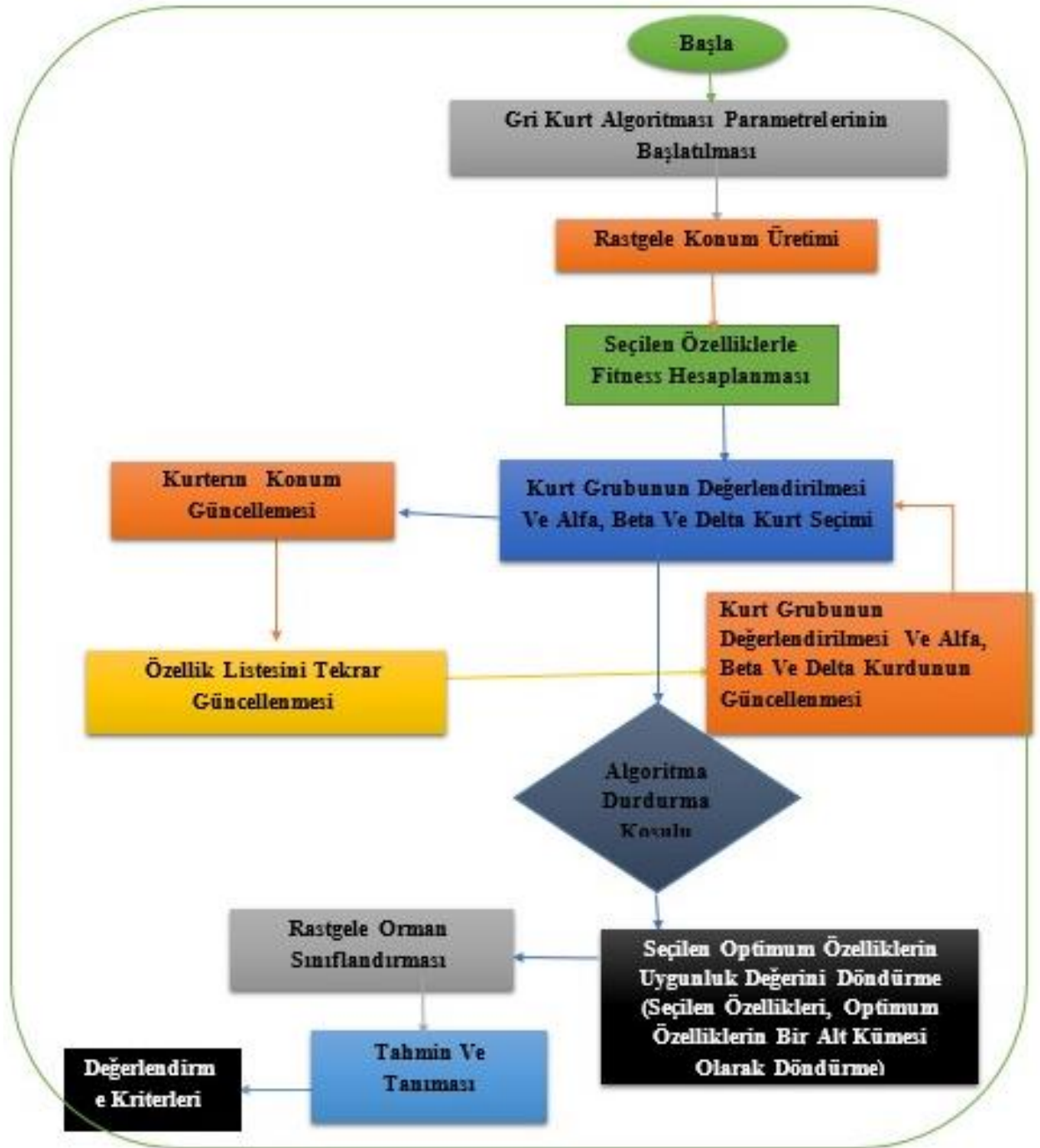
Şekil 4.4 Özellik seçim ekranı.

1 sayısı bir özelliğin seçildiğini, 0 sayısı ise özelliğin seçilmediğini gösterir. Bu konu Şekil 4.4'te gösterilmektedir. Ayrıca önerilen algoritmanın akış şeması Şekil 4.5'te gösterilmektedir.

Aşağıda önerilen algoritmanın uygulanma yöntemi tartışılmış ve algoritmanın başarımının nasıl artırılacağı gösterilmiştir.

Verilen akış şemasında öncelikle kurtlara rastgele bir başlangıç değeri verilir. Daha sonra bu konumlar kullanılarak seçilen özellikler elde edilir ve bu özellikler değerlendirme fonksiyonu (4.1) ile değerlendirilir. Aslında bu fonksiyon en iyi özellikleri seçmek için kullanılır. Denklem (4.1)'ün değerlendirme fonksiyonu, seçilen özellikleri değerlendirmek için kullanılan gri kurt algoritmasında sınıflandırma doğruluğunu en üst düzeye çıkarmak için kullanılır. Sonraki iterasyonlarda bu konumlar tekrar güncellenerek alfa, beta ve delta kurtları belirlenir. Son olarak değerlendirme fonksiyonu kullanılarak en iyi özellikler belirlenir. Daha sonra özelliklere rastgele orman sınıflandırması uygulanarak koroner arter (kalp) hastalığı tespit edilir ve değerlendirme kriterleri uygulanır.

Tablo 4.1'de ayrıca son yıllarda Cleveland verileri üzerinde koroner kalp hastalığını öngörmek amacıyla yapılan çalışmalar da gözden geçirilmekte olup bunlardan bir kısmına bu bölümde yer verilmektedir. Ayrıca değerlendirme kriterlerinde önerilen algoritmanın etkinliğinin artırılması da bu tabloda gösterilmektedir.



Şekil 4.5 Önerilen algoritmanın akış şeması

Tablo 4.1 Önerilen algoritmanın verimliliğinin önceki çalışmalarla karşılaştırılması

Makalenin yazarı	Makalenin yayınlanma yılı	Makalenin konusu	Accuracy Kesinlik	Sensitivity hassasiyet	Specificity duyarlık	İncelenmekte olan veri seti	Makalede kullanılan yöntem	Değerlendirme kriterleri 10-fold
aravind Akella ve arkadaşları (A. Akella & V.Kaushik, 2020)	2020	Machine Learning Algorithms for Predicting Coronary Artery Disease: Efforts Toward an Open-Source Solution	%87.64 %86.52 %79.78 %84.10 %93	%82.61 %79.59 %74.47 %78.72 %%93.1	----- ----- ----- ----- -----	Cleveland UCI	Random forest Support-Vector Machine Decision Tree k-Nearest Neighbors Neural Network	10-fold
Sima dehnavi , majid emamipour , amin golabpour (S. Dehnavi & M. Emamipour, 2020).	2020	Combining Random Forest and Neural Networks Algorithms to Diagnose Heart Disease	%87.65	%87.5	%86	Cleveland UCI	Random Forest +Neural Network	10-fold

Munir Mokhtari ve arkadaşları (F. Abdeldjouad & M. Brahami, 2020).	2020	A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques	%79.13 %78.77	%79.96 %78.22	%75.1 %71.2	Cleveland UCI	Multi-Objective Evolutionary Fuzzy Classifier (MOEFC), Logistic Regression (LR)	10-fold
Parvin SGAME, Vinoud Vaze, Emmanuel M (P. S. Game & V. Vaze & M. Emmanuel, 2019).	2019	Induction of Decision Trees based on Gray Wolf Optimizer for heart disease classification	%90	%89	%88.9	Cleveland UCI	GWO + ID3	10-fold
Qasem Al-Tashi ve arkadaşları (Q. Al-Tashi & H. Rais & S. Jadid, 2019).	2019	Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification	%89.83	%93	%91	Cleveland UCI	GWO+SVM	10-fold
Prf. Paria R. Patil	2017	Automated Diagnosis of	%83.6	%82.55	%81.7	Cleveland UCI	Weighted Random	10-fold

(P. P. R. Patil & P. S. A. Kinariwala , 2017).		Heart Disease using Random Forest Algorithm					Forest	
S. Kumar ve G. Sahoo (S. Kumar & G. Sahoo, 2017).	2017	A Random Forest Classifier based on Genetic Algorithm for Cardiovascular Diseases Diagnosis	%93	%90	%91	Cleveland UCI	Random forest + GA	10-fold
Ahmad Musa ve arkadaşları	2016	A Comparative Study for Predicting Heart Diseases Using Data Mining Classification Methods	%93	%90.8	%91.78	Cleveland UCI	Decision tree + Random forest	10-fold
Challet KS, Sugumaran V (S. Vachiravel , 2015).	2015	Diagnosis of Heart Disease Using Decision Tree and SVM Classifier	%84.3	%81	%81.5	Cleveland UCI	Decision tree + SVM Classifier	10-fold
Araştırmada	2023	Önerilen algoritma	%97,93	%98,21	%97,2	Cleveland UCI	GWO+RF	10-fold

4.5. Önerilen Algoritmayı Uygulamak İçin Kullanılan Veriler

Önerilen algoritmayı uygulamak için iki veri modeli kullanılmıştır; verilerden biri UCI koleksiyonu için Cleveland verileridir. UCI Makine Öğrenimi Koleksiyonu, makine öğrenimi topluluğu tarafından makine öğrenimi algoritmalarının ampirik analizi için kullanılan veritabanları, etki alanı teorileri ve veri oluşturuculardan oluşan bir koleksiyondur. Bu bilgi arşivinin önemli noktaları arasında 1000'den fazla atıf yapılmış olması da sayılabilir. Web sitesinin mevcut versiyonu 2007 yılında Arthur Asuncion ve David Newman tarafından tasarlanmıştır. Bu verilere 303 kişi katılıyor ve bunların 15'i aşağıda detaylı olarak anlatılacak olan kalp hastalıkları ile ilgilidir. Diğer veriler M-Serbram platformuyla ilgilidir. Bir takım özellikleri bahsedilen cihazlardan elde edilen verilerle ilgilidir. Kalp hastalıklarıyla ilgili bilgilerin alınmasında kullanılan Bluetooth tansiyon aleti, Zephyr BT gibi cihazlardır. Düşük enerjili bluetooth teknolojisini kullanan söz konusu cihazlar, cep telefonlarına bağlanma özelliğine sahip olup, bu bilgiler internet üzerinden sunuculara aktararak doktorun hastayla web arayüz portalı üzerinden iletişim kurması sağlanmaktadır. Bu veritabanı 14 özellik içermektedir ve bu veritabanına katılan kişi sayısı 254 kişidir. Aşağıda bu özellikleri tam olarak inceleyeceğiz.

4.5.1. Cleveland Verilerinin Özellikleri

Şekil 4.6, bir dizi özelliğin gösterildiği Cleveland verilerinin bir bölümünü göstermektedir. Ardında her bir özellik ve önemi incelenecektir. Tablo 4.2'te Cleveland verilerinin özellikleri, sayısal (Numerical), nominal ve ayrıca bunların değeri ve ortalamasını içeren özellik türleri ile birlikte gösterilmektedir.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
sex	age	cp	trestbps	thal	fbs	restecg	thalach	exang	oldpeak	slope	ca	chol	target	currentSmcigs	PerDay
1	63	3	145	1	1	0	150	0	2.3	0	0	233	1	0	0
1	37	2	130	2	0	1	187	0	3.5	0	0	250	1	0	0
0	41	1	130	2	0	0	172	0	1.4	2	0	204	1	1	20
1	56	1	120	2	0	1	178	0	0.8	2	0	236	1	1	30
0	57	0	120	2	0	1	163	1	0.6	2	0	354	1	1	23

Şekil 4.6 Cleveland verileri.

Tablo 4.2 Cleveland verilerinin özelliklerinin açıklaması

Özelliklerin Miktarı	Özellik Türü	Öznitellikler
28 - 77 Ortalama: 51.9	Sayısal	Yaş
Kadın: 0 Erkek: 1	Nominal	Cinsiyet
1 ve 2 ve 3 ve 4	Nominal	Göğüs ağrısı
80 - 200 Ortalama: 61.8	Sayısal	Tansiyon
100 - 603 Ortalama: 204	Sayısal	Kolesterol
0 ve 1	Nominal	hızlı kan şekeri
0 ve 1 ve 2	Nominal	kalp atışı
60 - 602 Ortalama: 140.6	Sayısal	Maksimum kalp atış hızı
0 ve 1	Nominal	Egzersiz neden olduğu anjina
-2,6 ve 6.2	Sayısal	Dinlenme sırasında ST derinliğinde azalması
1 ve 2 ve 3	Nominal	Egzersiz sırasında ST segment zirvesi
1 ve 2 ve 3	Nominal	Talasemi
0-3	Sayısal	Floroskopi ile boyanan büyük damarların sayısı
0 ve 1	Nominal	sigara içen biri olmak
0-30	Sayısal	Günde tüketilen sigara sayısı

- **Cinsiyet**

(Erkek:1 ve Kadın:0) Erkeklerin kalp hastalıklarına yakalanma riski kadınlara göre daha fazladır.

- **Yaş**

Yaş, kardiyovasküler veya kalp hastalıklarının gelişmesinde en önemli risk faktörüdür ve bu oran her on yılda neredeyse üç katına çıkar. Ergenlik

döneminde damarlarda yağ birikintileri oluşabilmektedir. Koroner kalp hastalığından ölenlerin yüzde 82'sinin 65 yaş ve üzerinde olduğu tahmin ediliyor. Aynı zamanda felç riski 55 yaşından sonra her on yılda bir iki katına çıkıyor. Kullanıcı mevcut uygulamalara yaşını girebilmektedir.

- **Göğüs Ağrısının Türü**

Angina aslında kalp kasınızın oksijenli kan almaması durumunda ortaya çıkan göğüs ağrısıdır. Göğsünüzde baskı hissedebilirsiniz. Bu rahatsızlık aynı zamanda omuzlarınızda, kollarınızda, boynunuzda, çenenizde veya sırtınızda da ortaya çıkabilir. Aşağıdaki formatı kullanan kişinin gösterdiği göğüs ağrısının türü aşağıdaki gibidir.:

- 1) Stabil anjina
- 2) Stabil olmayan anjina
- 3) Varyant anjina

- **Tansiyon**

Zamanla yüksek tansiyon kalp damarlarına zarar verebilir. Obezite, yüksek kolesterol veya diyabet gibi diğer durumların neden olduğu yüksek tansiyon riski artırır. Cleveland'ın verilerinde bahsedilen kan basıncı, yüksek tansiyon (sistolik) ile ilgilidir.

- **Akdeniz Anemisi**

Akdeniz anemisi (Thalassemia), genetik bir hastalık ve bu hastalıktan muzdarip bir kişinin vücudundaki hemoglobin ve kırmızı kan hücrelerinin sayısının normalden düşük olduğu kalıtsal bir kan bozukluğudur.

- 1 ise normal moddur.
- 2 ise hafif (Minor) tir.
- 3 ise şiddetli (Major) dir.

- **Hızlı Kan Şekeri**

Hızlı kan şekeri (Fast Blood Sugar) 0 ve 1 olarak gösteriliyor.

1 ise izin verilen sınırın üzerinde ve kan şekeri var demektir.

0 ise sınırın altındadır ve kan şekeri yoktur.

- **Kalp Atış Hızı Kaydı (Elektrokardiyografi)**

Kalp atış hızı kaydı 3 kategoriye ayrılır:

- 1) 0 ise normal demektir.

2) 1 ise ST-T dalgası anormalliği var demektir.

3) 2 ise sol ventrikül hipertrofisidir.

Sol ventrikül hipertrofisi, kan basıncının artması veya aort kapak darlığının bir sonucu olarak kalbin sol ventrikül kaslarının genişlemesidir. Sol ventriküler hipertrofi nispeten yaygın bir genetik hastalıktır (nüfusun %1'i).

- **Ulaşılan Maksimum Kalp Atış Hızı**

Elde edilen maksimum kalp atış hızı (Thalach) bir tam sayıdır.

- **Egzersiz Neden Olduğu Anjina**

Egzersiz testinde egzersizin neden olduğu anjinayı (Exang) hesaplamak için göğüs ağrısı olursa sayı 1, aksi takdirde sayı 0'dır.

- **Dinlenme Sırasında ST Derinliğinde Azalma**

dinlemede ST segment derinliğindeki azalma (Old peak) tam sayı veya ondalık sayı olarak ifade edilir.

- **Egzersiz Sırasında ST Segment Zirvesi**

Egzersiz sırasında ST segmentinin zirvesi (Slope) aşağıdaki 3 çeşite sahiptir:

1) Yükselen durumda ise 1'e eşittir.

2) Düz durumda 2'ye eşittir.

3) Depresif durumda ise 3'e eşittir.

- **Floroskopi İle Boyanan Büyük Damarların Sayısı**

Floroskopi ile boyanan ana damarların sayısı (Ca) bir tamsayı olarak görüntülenir. Bu özellik sayısal bir özelliktir.

- **Kolesterol**

Düşük yoğunluklu lipoprotein kolesterolün (LDL - kötü kolesterol) yüksek seviyelerinin arterlerin daralmasına neden olma olasılığı daha yüksektir. Kandaki bir tür diyet yağı olan trigliseritlerin yüksek seviyeleri kalp krizi riskini artırır. Ancak yüksek yoğunluklu lipoprotein kolesterol (HDL- iyi kolesterol) seviyesinin yüksek olması kalp krizi riskini azaltır.

- **Sigara İçen Biri Olmak**

Değer 1'e eşitse sigara içen tespit edilir.

Değer 0'a eşitse sigara içen tespit edilmez.

- **Günde Tüketilen Sigara Sayısı**

Bu bir tam sayıdır.

- **Kalp Hastalığının Teşhisi**

Bu özellik hedef özellik olarak sunulmaktadır. Kişinin kalp hastası olup olmadığını gösterir. İki değer 0 ve 1'dir.

Değer 1'e eşitse kalp hastası tespit edilir.

Değer 0'a eşitse kalp hastası tespit edilmiyor.

4.5.2. Farmingham Verilerinin Özellikleri

Framingham Kalp Çalışması, Ulusal Sağlık Enstitüleri ve Boston Üniversitesi Tıp Fakültesi'nin bir parçası olan Ulusal Kalp, Akciğer ve Kan Enstitüsü tarafından desteklenmektedir. Bu

veriler bu merkezden elde edilmiştir. Bu veriler içerisinde 2124 kişi incelenmiş olup, verileri incelenmektedir. Bu veriler, 14'ü bağımsız değişken (risk faktörü) ve bir diğeri bağımlı değişken (hedef) olmak üzere 15 değişken içermektedir. Şekil 4.7 Farmingham'dan alınan verilerin bir kısmını bir takım özellikler ile birlikte gösterilmektedir.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
age	male	currentSmoker	See a cardiologist	Occurrence of Stroke	prevalentHyp	diabetes	totChol	BMI	glucose	sysBP	diaBP	BPMeds	heartRate	TenYearCHD
39	1	0	0	0	0	0	195	26.97	77	106	70	0	80	0
46	0	0	0	0	0	0	250	28.73	76	121	81	0	95	0
48	1	1	1	0	0	0	245	25.34	70	127.5	80	0	75	0
61	0	1	1	0	1	0	225	28.58	103	150	95	0	65	1

Şekil 4.7 Farmingham verileri.

Bu özellikler birkaç farklı kategoriye ayrılmıştır. İlk kategori, insanların cinsiyetini ve yaşını içeren demografik istatistiklerle ilgilidir.

- **Yaş**

Verilerde yer alan kişilerin yaşları tamsayı şeklindedir.

- **Cinsiyet**

Erkek olan: 1

Kadın olan: 0

İkinci kategori davranışsal özelliklerle ilgilidir.

- **Katılımcının Sigara İçmesi**

Katılımcının sigara içip içmediği (Current smoker) iki değer ile gösterilmektedir:

Sigara içiyor olmak: 1

Sigara içmiyor olmak: 0

Üçüncü kategori hastalığın geçmişi ile ilgilidir.

- **Kişinin Bir Kardiyoloğa Gidip Gitmemesi**

kişinin kardiyoloğa gidip gitmediği (See a cardiologist) iki değer olarak gösteriliyor.

Evet, ziyaret ettim: 1

Hayır, ziyaret etmedim:0

- **Kalp Krizi Geçmişi**

kişinin kalp krizi geçirip geçirmediği (Occurence of stroke) iki değer şeklindedir.

Evet: 1

Hayır: 0

- **Yaygın Yüksek Tansiyon**

Kişinin devamlı yüksek tansiyon (Prevalent Hyp) geçmişi var mı, yok mu?

Evet: 1

Hayır: 0

- **Diyabet**

Kişi şeker hastası (Diabetes) mı değil mi?

Evet: 1

Hayır: 0

Dördüncü kategori laboratuvar bilgileriyle ilgilidir.

- **Kan Kolesterol Düzeyi**

Kan kolesterol düzeyi bir tam sayıdır.

- **Vücut Kitle ndeksi**

BMI vücut kitle indeksi anlamına gelir. Bu indeks aslında fazla kilonun miktarını veya kaybedilen kiloyu veya ideal kilo aralığını belirlemek için kullanılabilecek bir tarama kriteridir. Dünya Sağlık Örgütü'ne göre yetişkinler için ideal vücut kitle indeksi 18,5 ile 24,5 arasındadır.

Ondalık sayı şeklindedir.

BMI hesaplama formülü aşağıdaki şekildedir:

$$BMI = \frac{m}{h^2} \quad (4.2)$$

- **Kan Şekeri Seviyesi**

Kan şekeri seviyesi (Glucose) bir tam sayıdır.

Beşinci kategori giyilebilir cihazlardan elde edilen bilgilerdir.

- **Sistolik Ve Diyastolik Kan Basıncı**

Atardamarlarınızdan akan kanın neden olduğu basınç sabit değildir; dinamikdir ve kalbin herhangi bir anda ne yaptığını temsil eder. Kalp aktif olarak attığında (sistol adı verilen bir olay), kanı damarlara pompalar. Kanın atardamarlara bu dinamik akışı, atardamarların içindeki basıncı artırır. Aktif kalp kasılması sırasındaki yüksek tansiyona sistolik (Systolic) kan basıncı denir. Bir kişi sessizce ve sakin oturduğunda normal sistolik kan basıncı 120 mmHg veya daha azdır. Duygusal stres dönemlerinde veya kalbin dinlenme halinden daha fazla attığı herhangi bir zamanda, bir kişi egzersiz yaparsa; Kalbin kasılma kuvveti ve sistolik basınç artar. Kardiyak stres koşullarında sistolik kan basıncındaki artış tamamen normaldir.

Diyastolik (Diastolic) kan basıncı, her kalp atışı arasında kanın arterlerden boşaltıldığı basınçtır. Yani kalp aktif olarak atardamarlara kan pompalamadığındaki tansiyondur. Kalbin kasılması bittikten sonra kalbin karıncıkları şaşırtıcı bir şekilde gevşeyerek bir sonraki kasılmaya hazırlanır. Bu ventriküler gevşeme periyoduna diyastol adı verilir ve diyastol sırasındaki kan basıncına diyastolik kan basıncı denir. Dinlenme sırasında normal diyastolik kan basıncı 80 mmHg veya daha azdır. Bir kişinin yüksek tansiyonu varsa, diyastolik kan basıncı sıklıkla dinlenme halinde artar. Sistolik ve diyastolik kan basıncını ölçmek için Omron kablosuz brakiyal kan basıncı adı verilen ve aynı zamanda 200 kayıt saklama özelliğine sahip bir cihaz kullanılıyor.

- **Tansiyon İlacı Kullanımı**

Evet: 1

Hayır: 0

- **Kalp Atışı**

BT Zephyr cihazından kalp atış hızını ölçmek için kullanılır. Bu cihaz, sürekli izlemeye ihtiyaç duyan hastaların kalp atış hızı bilgilerini çıkarmak için kullanılan giyilebilir bir sensördür. Zephyr BT sensörleri doğruluk, dikkat, güvenilirlik, maliyet ve kullanışlılığa göre seçilir. Bluetooth düşük enerji teknolojisi, bu sensörün kullanımının daha uzun süre çalışmasını sağlar.

4.5.3. Framingham Verilerinde Koroner Kalp Hastalığını Tanımaya Yönelik Özelliklerin Seçilmesindeki Öncelik

Tahran Kalp Merkezi uzmanlarından Sayın Dr. Kyomarth Majdi'nin görüşlerine göre söz konusu vakalarda bahsedilen özellikler önem sırasına göre şöyledir:

Yaş, sistolik ve diyastolik kan basıncı, kalp atış hızı, kandaki kolesterol düzeyi, kalp krizi geçmişi, yüksek tansiyon geçmişi, sigara kullanımı, vücut kitle indeksi, şeker hastalığı, Cinsiyet, Glikoz, Tansiyon ilacı kullanmak, Kalp doktoruna gitmek

4.5.4. Cleveland Verilerinde Koroner Kalp Hastalığını Tanımaya Yönelik Özelliklerin Seçilmesindeki Öncelik

Tahran Kalp Merkezi uzmanlarından Sayın Dr. Kyomarth Majdi'nin görüşlerine göre söz konusu vakalarda bahsedilen özellikler önem sırasına göre şöyledir:

Yaş, kan basıncı, üç kategoriye ayrılmış kalp hızı (0 normal durum, 1 ST-T dalgası anormalliği, 2 sol ventriküler hipertrofi), ulaşılan maksimum kalp hızı, kolesterol, 4 kategoriye ayrılan göğüs ağrısının türü, floroskopi ile boyanan ana damar sayısı, egzersizin neden olduğu anjina, sigara kullanımı ve günlük tüketilen sigara sayısı, ST segmentinin zirvesi egzersiz sırasında (3 mod dikkate alınır. a- Yükselen durumda, b- düz durumda, c- depresif durumda), dinlenme sırasında egzersizden kaynaklanan ST derinliğinde azalma.

4.6. Önerilen Gri Kurt Ve Rastgele Orman Algoritmasını Uygulamak İçin Kullanılan Fonksiyonlar

Bu algoritmanın uygulanmasında MATLAB versiyon R2020b yazılımı kullanılmıştır. Önerilen algoritmanın kodunu çalıştırmak için kodun ana fonksiyonu (Main code) kullanılır. Ana fonksiyonda diğer fonksiyonlar da çağrılır ve her biri ayrı ayrı açıklanacak.

- **MATLAB Yazılımında Veri Okuma**

Öncelikle csv uzantılı bir dosya olan veriler MATLAB'da csv okuma dosyası (Read CSV) olarak saklanır.

- **Veri Yükleme Fonksiyonu (Load Data)**

Bu fonksiyonda veriler programa girilir. Heart.mat adı verilen uzantı dosyası ile matris olarak kaydeder. Özelliklerle birlikte tüm bilgiler bu veritabanında mevcuttur.

- **Veri Yönetim Fonksiyonu (Manage Data)**

Bu fonksiyonda test ve eğitim verileri belirtilir ve eğitim verileri üzerinde çapraz doğrulama gerçekleştirilir. Verilerimiz kalp bilgisine ilişkin 15 özellik (sütun) ve hedef olarak 1 özellik (sütun) içermektedir. Verilerin %80'i eğitim için kullanılır. Rastgele fonksiyon (Randperm) kullanılarak endeksler rastgele hareket ettirilmiştir. Doğrulama için eğitim verilerine uygulanan çapraz yöntem kullanılır. Bu tezde k sayısı 10'a eşittir. Eğitim verileri 10 kategoriye bölünmüş ve her seferinde biri geçerli veri (Valid data) olarak, diğer 9 veri ise eğitim için kullanılmıştır.

- **Nitelik Değerlendirme Fonksiyonu (Evaluate Features (Cost Function))**

Bu fonksiyon, gri kurt algoritmasında özellik seçimi için bir maliyet fonksiyonu olarak kullanılır. s, 0 ve 1 olarak görüntülenen özellikleri içerir ve ns sa'nın özelliklerinin sayısıdır. Ayrıca seçilen özellik sayısının yüzdesi de belirtilir. Rastgele orman sınıflandırma algoritmasını test ve eğitim verilerine uygulamak için ağaç toplayıcı fonksiyonu (Treebagger) kullanılmıştır. Uydurma fonksiyonunu (Fitness) yazmak için gri kurt ve destek vektör algoritmasının birleşimi ile ilgili makalede verilen formül kullanılmıştır. İki bölümden oluşan ağırlıklı bir fonksiyon kullanılmış olup, ilk bölüm her bir bölümde sınıflandırma doğruluğunun hesaplanması, ikinci bölüm ise özellik seçimi ile ilgilidir.

$$Z = (1 - w_1 * mean(accuracy)) + (w_2 * prediction) \quad (4.3)$$

- **Gri Kurt Algoritmasını Kullanan Özellik Seçim Fonksiyonu (Feature Selection Using GWO)**

Özellik seçimi görevini gerçekleştiren ana fonksiyon bu bölümde tanımlanmıştır. Özellik değerlendirme fonksiyonu burada aynı zamanda bir maliyet fonksiyonu olarak da değerlendirilmektedir. Bu fonksiyonda gri kurt denklemleri kullanılarak alfa, beta ve delta kurtlar için farklı konumlar belirlenmiş ve son olarak farklı konumların ortalaması son konum olarak sunulmuştur. Landa katsayısını kullanarak kurtların konumu için elde edilen sürekli değerleri 0 ve 1 değerlerine dönüştürür. Son olarak maliyet fonksiyonunun değerleri belirlenir ve tekrar sayısına göre maliyet fonksiyonuna erişim sayısını gösteren değişken değerleri güncellenir.

- **Basit Rastgele Orman Fonksiyonu (Simple Random Forest)**

Bu bölümde rastgele orman sınıflandırmasını uygulamak için bir ağaç toplayıcı fonksiyonu kullanılmıştır. Bu fonksiyonda istenilen ağaç sayısı belirlenir ve ayrıca bu fonksiyonun son kısmında rastgele orman algoritmasının doğruluğu hesaplanır.

- **Değerlendirme Grafikleri Fonksiyonu (evaluateplot)**

Bu fonksiyonda dikkat, doğruluk, okunabilirlik ve F_1 -score hesaplanmasını içeren değerlendirme kriterleri ile performans karakteristik diyagramı ve altındaki seviye de hesaplanır. Bu fonksiyon, gri kurt algoritmasına göre değerlendirme için kullanılır ve bu algoritma tarafından seçilen özellikleri kullanır.

- **Değerlendirme Grafikleri Fonksiyonu1 (evaluateplot1)**

Bu fonksiyonda dikkat, doğruluk, okunabilirlik ve F_1 -score hesaplanmasını içeren değerlendirme kriterleri ile performans karakteristik diyagramı ve altındaki seviye de hesaplanır. Bu fonksiyon rastgele orman sınıflandırma algoritmasını değerlendirmek için kullanılır.

- **Ana Fonksiyon (Main Code)**

Program kodunun ana yürütülmesi bu fonksiyonda yapılır ve adı geçen tüm fonksiyonlar, parametreleriyle birlikte bu ana fonksiyonda çağrılır. Proje koduna ilgili görüntüleme aşamasında tüm değerlendirme kriterleri incelenir ve son olarak seçilen özellikler görüntülenir.

- **Ağaç Toplayıcının Fonksiyonunun İncelenmesi (Tree Bagger)**

Bireysel karar ağaçları aşırı uyum sağlama (Overfitting) eğilimindedir. Toplu karar ağaçları, ağaç prototipleri (kendi kendine önyükleme (Bootstrap)), birçok karar ağacının sonuçlarını birleştirir, bu da aşırı uyumun etkilerini azaltır ve genellemesini artırır. Ağaç birleştirici fonksiyonu, ağaç prototiplerini kullanarak verileri gruplayarak karar ağaçlarını büyütür. Ayrıca, ağaç toplayıcı fonksiyonu, rastgele orman algoritmasına benzer şekilde, karar ağacının her bölümünde kullanılacak tahmin edicilerin rastgele bir alt kümesini seçer.

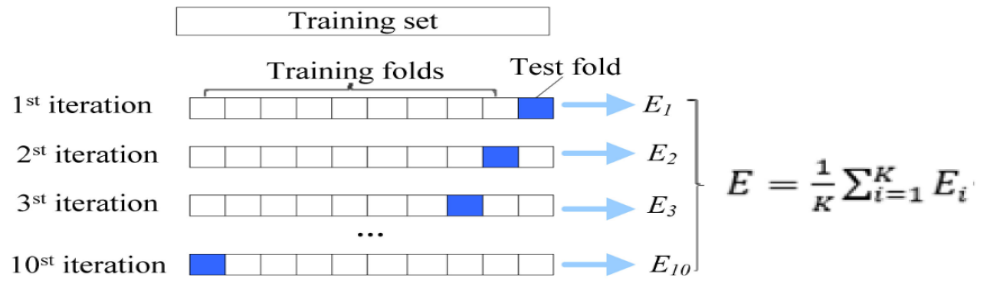
4.6.1. Gri Kurt Algoritmasının Uygulanması

Bu bölümde ilk gri kurt algoritmasının bir kısmı uygulanmış olup alfa, beta ve delta kurtların nasıl tanımlanacağı tartışılmıştır. Problem tanımlama aşamasında öncelikle arama uzayı belirlenir ve bu uzayda en uygun çözüm de belirlenir. Daha sonra parametreler ayarlanarak başlatma aşamasına (Initialization) geçilir. Bu aşamada gri kurtların kopya sayısı ve popülasyonu belirlenir. Ayrıca kurtların konumu ve maliyet fonksiyonu (Cost function) da bu aşamada belirlenir. Sonra alfa, beta ve delta kurtları belirlenerek değerlendirilir. Daha sonra ise kurt sayısına eşit olan bir halkada kurtların konumları rastgele bir fonksiyon kullanılarak belirlenir. Döngü durdurma koşulu, alfa kurt olan en iyi sonucun konumu ve maliyetinin gösterildiği zamandır. En iyi cevaba ulaşılamazsa kurtların konumu tekrar güncellenmelidir.

4.7. Çapraz Doğrulama Kriteri (K-fold Cross Validation)

Çapraz doğrulama, veri bilimcileri tarafından yaygın olarak kullanılan en önemli tekniklerden biridir. Makine öğrenimi modelleriyle çalışmanın sorunu,

model bağımsız bir veri kümesinde test edilene kadar modelin performansını elde edememektir. Çapraz doğrulama, makine öğrenimi modellerinin performansını değerlendirmek için çok yararlı bir tekniktir. Bu yöntem, makine öğrenimi modelinin nasıl oluşturulduğunu ve bağımsız bir veri kümesine nasıl genişletildiğini anlamaya yardımcı olur. Bu teknik, modelin gerçek dünyadaki doğruluğunu ve verimliliğini tahmin etmek için kullanılır. Bir makine öğrenmesi problemi verildiğinde genellikle bilinen veriler (Training data set - eğitim verileri) ve bilinmeyen veriler (Test data - test verileri) olmak üzere iki veri kümesi görülür. Çapraz doğrulama, makine öğreniminde yaygın olarak kullanılan en yaygın çapraz doğrulama türlerinden biridir. Bu doğrulama türünde veriler K alt kümeye bölünür. Bu K alt kümelerden her seferinde biri doğrulama için kullanılırken, diğer K-1 adet küme eğitim için kullanılır. Bu prosedür K kez tekrarlanır ve tüm veriler tam olarak bir kez eğitim ve bir kez doğrulama için kullanılır. Son olarak, bu K kez doğrulamanın ortalama sonucu, nihai bir tahmin olarak seçilir. Elbette sonuçları birleştirmek için başka yöntemler de kullanılabilir. Normalde K-fold yöntemi kullanılır. Tabakalı K-Fold yönteminde her sınıfa ait verilerin oranı, her alt kümede ve ana kümede aynı tutulmaya çalışılır. Şekil 4.8 aslında bu doğrulama yöntemini göstermektedir.



Şekil 4.8 Doğrulama yöntemi.

Bu çalışmada eğitim verilerinin doğrulanması için K-fold yöntemi kullanılmış olup, istenilen K sayısı 10'a eşittir. Bu yöntemi MATLAB'da uygulamak için eğitim verilerinde crossvalind fonksiyonu kullanılmıştır.

4.8. Değerlendirme kriterleri

Bu bölümde önerilen modele ilişkin değerlendirme kriterleri incelenmiştir. Ayrıca önerilen model için bu kriterlere göre rastgele orman sınıflandırma algoritmasının tek başına kullanıldığı durumla karşılaştırma yapılmıştır. Önerilen algoritmanın gelişimini karşılaştırmak için değerlendirme kriterleri kullanılmıştır. Bu amaçla eğitim verileri, test ve toplam veriler üzerinde doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Sensitivity) veya okunabilirlik kriterleri ile F_1 -score gelişimi gösterilmekte, ayrıca performansın karakteristik grafikleri ve grafiklerin düzeyi karşılaştırılmaktadır. Bu nedenle değerlendirme kriterleri incelenerek her birinin önerilen model üzerindeki etkisi incelenmiştir. Daha sonra açıklanacak olan karışıklık matrisine göre aşağıdaki kriterler tanımlanabilir.

- **Doğruluk**

Genel olarak doğruluk, modelin çıktısı ne kadar iyi tahmin ettiğini ifade eder. Yakından bakarak modelin uygun şekilde eğitilip eğitilmediğini ve genel olarak nasıl performans gösterdiği hemen görülebilir.

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.4)$$

- **Kesinlik**

Yanlış pozitif değeri yüksek olduğunda kesinlik ölçüsü uygun bir ölçü olacaktır. Model sonucu pozitif olarak tahmin ettiğinde, pozitif olarak tahmin edilen tüm durumlara karşı doğru pozitif tahminlerin sayısını gösterir.

$$PPV = \frac{TP}{TP+FP} \quad (4.5)$$

- **Duyarlılık**

Tüm pozitif çıktılara karşı doğru tahmin edilen değerleri temsil eder. Diğer adı ise yeniden okumaktır. Yanlış negatif değer (False negative) yüksek olduğunda duyarlılık kriteri uygun bir kriter olacaktır.

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR \quad (4.6)$$

- **F1 Score**

F1 kriteri bir testin doğruluğunu değerlendirmek için uygun bir kriterdir. Bu kriter doğruluk ve duyarlılığı bir arada dikkate alır. F1 kriteri en iyi durumda bir, en kötü durumda ise sıfırdır.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.7)$$

- **Ayrıcalıklı Olmak (Specificity, Selectivity Or True Negative Rate (TNR))**

Tüm negatif çıktılara karşı tahmin edilen tüm negatif değerleri temsil eder.

$$\text{TNR} = \frac{\text{TN}}{\text{N}} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR} \quad (4.8)$$

- **Negatif Tahmin Değeri (Negative Predictive Value (NPV))**

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (4.9)$$

- **Kayıp Oranı (Miss Rate Or False Negative Rate (FNR))**

$$\text{FNR} = \frac{\text{FN}}{\text{P}} = \frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{TPR} \quad (4.10)$$

4.8.1. Performans Karakteristik Eğrisi

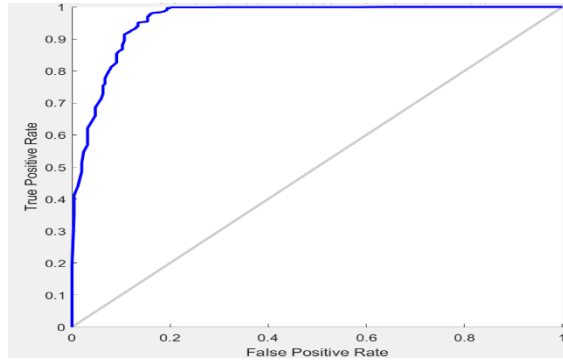
Makine öğreniminde performansı ölçmek ve hesaplamak önemli ve temel bir görevdir. Bu nedenle sınıflandırılan veya kategorize edilen problemler için performans karakteristik eğrisi kullanılır. Performans karakteristik eğrisi, algılama eşiği de değişken olan ikili sınıflandırma sisteminin değerlendirme yeteneğini temsil eden bir grafikdir. Performans karakteristik eğrisi en ünlü model değerlendirme yöntemlerinden biridir. Bu grafik sınıflandırma problemlerindeki verimliliğin bir ölçüsüdür. Gerçek pozitif oranın (True Positive Rate (TPR)) yanlış pozitif oranına (False Positive Rate (FPR)) oranının grafiği çizilerek bir performans karakteristik eğrisi oluşturulur. Ancak bu değerlere ilişkin eşiğin değişken olduğunu unutmayın. Bu nedenle sürekli bir grafik oluşturulacaktır. Gerçek pozitif oran terimi duyarlılıkla, yanlış pozitif oran terimi ise özgünlükle ilgilidir (Specificity).

Performans karakteristik eğrisi, ayırım eşiğinin değişken olduğu ikili bir sınıflandırma sisteminde duyarlılığın veya doğru tahmine karşı yanlış tahminin grafiksel bir

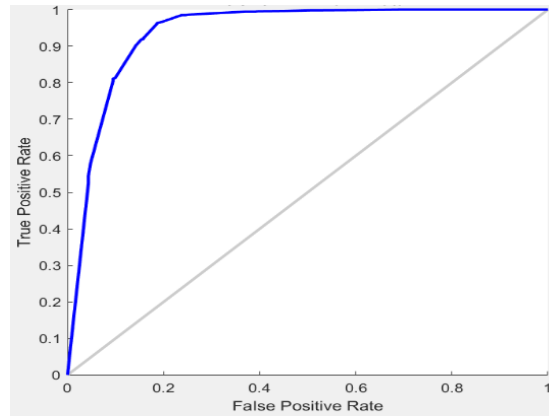
temsilidir. Eğrinin performansını açıklamak için en iyi sınıflandırma modelinin eğrinin bire en çok yaklaştığı model olduğu söylenebilir. Bu, bire ne kadar yakınsa ölçümün o kadar doğru ve uygun olduğu anlamına gelir. Ayrıca en kötü model eğrinin sıfıra yakın olduğu modeldir. Sıfıra ne kadar yakınsa ölçüm doğruluğunda bir zayıflık olduğunu gösterir.

Bu tezde performans karakteristik eğrisi kullanılmış olup, performans karakteristik eğrisi bir kişinin kalp hastalığı olup olmadığını belirlemede diğer ölçümlere göre daha doğrudur.

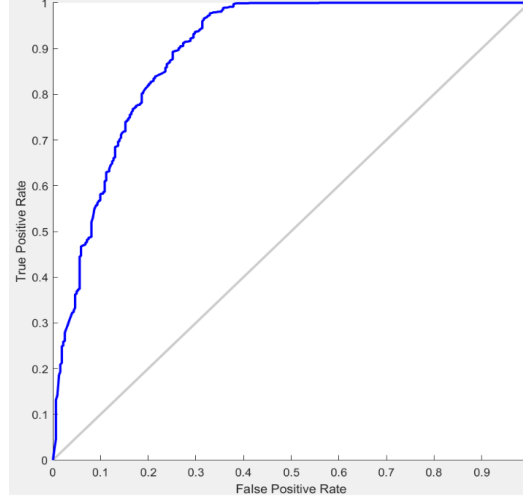
Rastgele orman algoritması ve gri kurt algoritmasının birlikte kullanıldığı duruma ilişkin performans karakteristik eğrisi Şekil 4.9'de gösterilmektedir. Ayrıca Şekil 4.10 rastgele orman algoritmasının tek başına kullanıldığı durumu göstermektedir. Bu iki performans karakteristik eğriler Cleveland verilerin içindir . Sonraki iki şekil, yani Şekil 4.11 ve Şekil 4.12, Framingham verilerine ilişkin performans karakteristik eğrileridir.



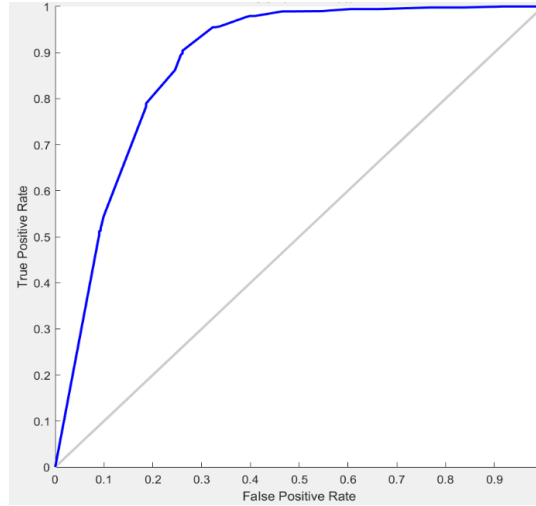
Şekil 4.9 Önerilen gri kurt ve rastgele orman algoritmasının performans karakteristik diyagramı



Şekil 4.10 Rastgele orman algoritması için performans karakteristik diyagramı



Şekil 4.11 Önerilen gri kurt ve rastgele orman algoritmasının performans karakteristik diyagramı



Şekil 4.12 Rastgele orman algoritması için performans karakteristik diyagramı

4.8.2. Performans Karakteristik Diyagramının Altındaki Seviyenin Kriteri (Area Under The Cover (AUC))

Bir sınıflandırıcı etkinliğini belirlemek için kullanılan bir diğer önemli kriter, performans özelliğinin alt diyagramının düzeyi kriteridir. Performans özelliği alt diyagramının seviye kriteri, performans özelliği diyagramı altındaki seviyeyi gösterir. Bu sayının değeri bir sınıflandırıcı için ne kadar yüksekse, sınıflandırıcının nihai verimliliği de o kadar olumlu olur. Performans karakteristik diyagramı, sınıflandırıcıların etkinliğini kontrol etmeye yönelik bir yöntemdir. Performans karakteristik eğrisinin altındaki alan, performansın bir yönünü ölçen bir sayıdır.

Performans karakteristik diyagramının altındaki seviye sıfır ile bir arasında bir değerdir. 0,5 değeri rastgele bir tahmindir ve 1 değeri mükemmel bir tahmindir. İncelenen örneğin kategorisini rastgele belirleyen bir sınıflandırıcı için performans özelliği alt diyagramının düzey değeri 0,5'e eşittir. Ayrıca bu kriterin maksimum değeri bire eşittir ve sınıflandırıcının ideal olduğu ve herhangi bir yanlış alarm olmadan tüm pozitif örnekleri tanıyabildiği bir durum için ortaya çıkar. Performans özelliği alt diyagramının seviye kriteri, sınıflandırıcıların etkinliğini belirleyen diğer kriterlerden farklı olarak sınıflandırıcının karar eşiğinden bağımsızdır. Dolayısıyla bu kriter, bir sınıflandırıcının çıktısının farklı veri setleri için güvenilirliğini gösterir; ve bu, sınıflandırıcının etkinliğini değerlendirmeye yönelik diğer kriterlerle hesaplanamayacağı anlamına gelir. Performans karakteristik diyagramının altındaki seviyeyi gösteren performans karakteristik alt diyagramının seviye kriteri, üst sınıflandırıcının tanıtılmasında belirleyici rol oynayabilir.

4.8.3. Karışıklık Matrisi (Confusion Matrix)

Sınıflandırma konularında bir modelin etkinliğinin analiz edilmesi ve değerlendirilmesine temel olan bu önemli tabloya karışıklık matrisi adı verilmektedir. Bu matrisin satır ve sütun sayısı kategori sayısına bağlıdır. Yukarıdaki tablo, kalp hastalığı olmak veya olmamak, şartlı olmak veya olmamak, işten ayrılmak veya ayrılmamak gibi ikili bir sınıflandırma problemi için 2x2'lik bir matrisi göstermektedir. Genel olarak, n-modlu bir sınıflandırma problemi için $n \times n$ işlem matrisinin oluşturulması gerekir. Şekil 4.13 karışıklık matrisini ve doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif değerlerin yerleşimini göstermektedir. Kalp hastalığını tahmin etme hedefinin bir hasta olduğunu varsayarsak, yani tahmin pozitifse hastanın kalp hastalığı olduğu, tahmin negatifse hastanın kalp hastalığı olmadığı anlamına geldiği durumda bu matrisin hücreleri analiz edilir.

Gerçek pozitif (True positive (TP)): Hastanın gerçekten kalp hastalığı varsa ve tahmin edilen değer aynı zamanda kalp hastalığını da gösteriyorsa.

Yanlış pozitif (False positive (FP)): Kalp hastalığı yoksa ancak tahminimizin sonucu kalp hastalığına işaret ediyorsa.

Yanlış negatif (False negative (FN)): Kalp hastalığı varsa ancak tahminimiz kalp hastalığının olumsuz olduğunu gösteriyorsa.

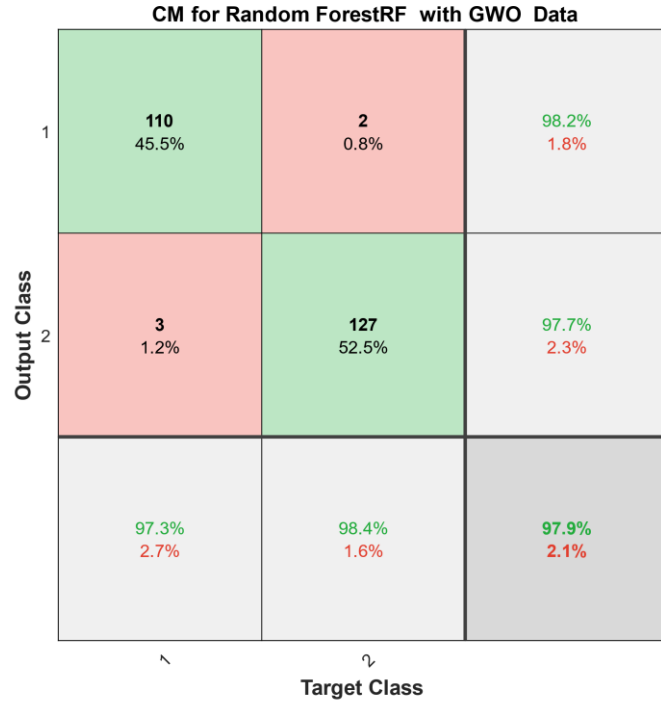
Gerçek negatif (True negative (TN)): Hastanın kalp hastalığı yoksa ve bizim tahminimiz de aynıysa gösteriyorsa.

Bilindiği gibi ideal olan, yanlış durumların (yanlış pozitifler ve yanlış negatifler) sıfır olmasıdır ancak pratikte bu gerçekleşmez ve verilerden oluşturulan modelin doğruluğunu ve verimliliğini kontrol edecek mekanizmalar ve kriterler gerektirir .

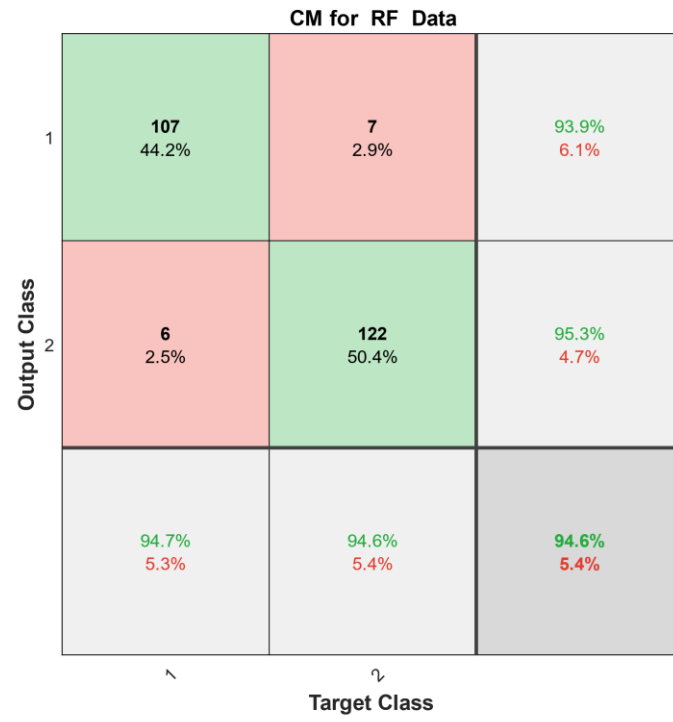
		Gerçek Değerler	
		Pozitive	Negative
Tahmin Edilen Değerler	Pozitive	Gerçek pozitif TP	Yanlış pozitif FP Tip I hatası
	Negatif	yanlış negatif FN Tip II hatası	gerçek negatif TN

Şekil 4.13 Karışıklık matrisi

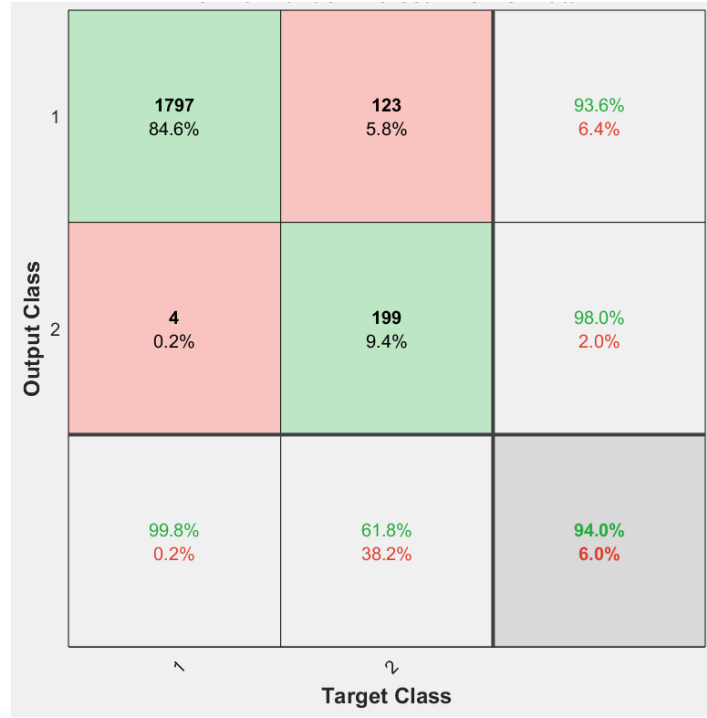
Bu tezdeki karışıklık matrisleri iki sınıflı veriler içindir. Bunlardan biri aslında gerçek sınıflarımız olan hedef sınıf (Target class), diğeri ise sınıflandırmamızın çıktısına karşılık gelen çıktı sınıfı (Output class)dır. Matristeki 4 sayıya dayanarak: doğru pozitif, doğru negatif, yanlış pozitif, yanlış negatif, istenen tüm kriterler elde edilebilir. Doğruluk, sınıflandırmanın kaç tanesini doğru bir şekilde tanıdığı ve tüm veri üzerinde tanımlandığı anlamına gelir. Bir bakıma doğruluk, kesinlikle aynı şeydir ancak özel bir sınıftadır. Şekil 4.14 ve Şekil 4.15, Cleveland verileri için karışıklık matrisleridir. Ayrıca, Şekil 4.16 ve Şekil 4.17, Framingham verileri için karışıklık matrisleridir.



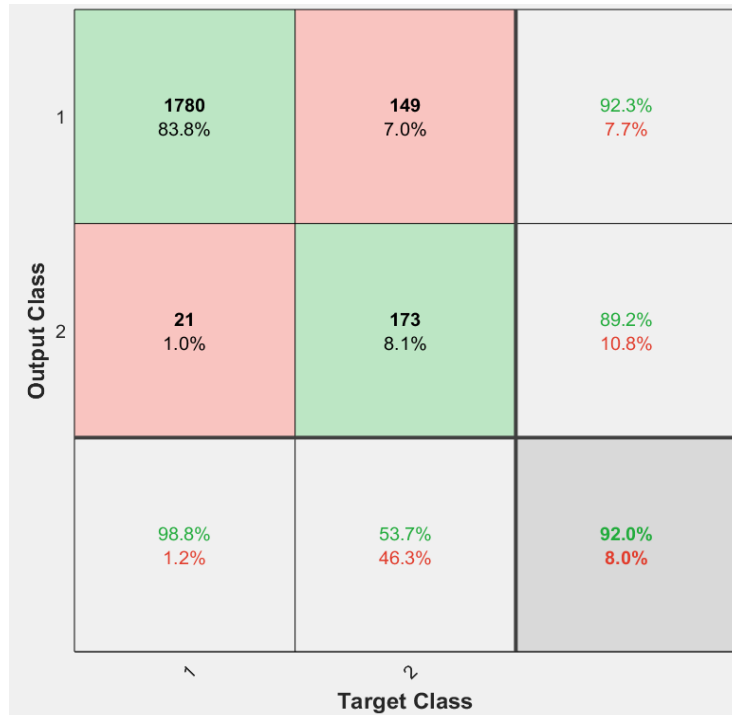
Şekil 4.14 Önerilen gri kurt ve rastgele orman algoritması için karışıklık matrisi



Şekil 4.15 Rastgele orman algoritması için karışıklık matrisi



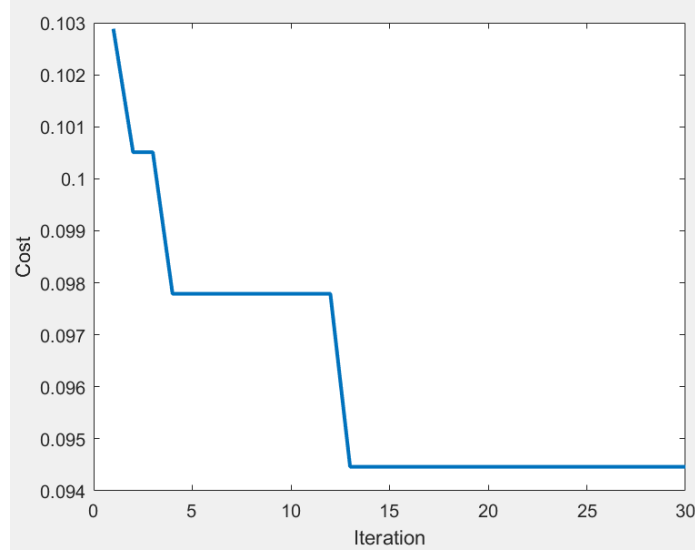
Şekil 4.16 Önerilen gri kurt ve rastgele orman algoritması için karışıklık matrisi



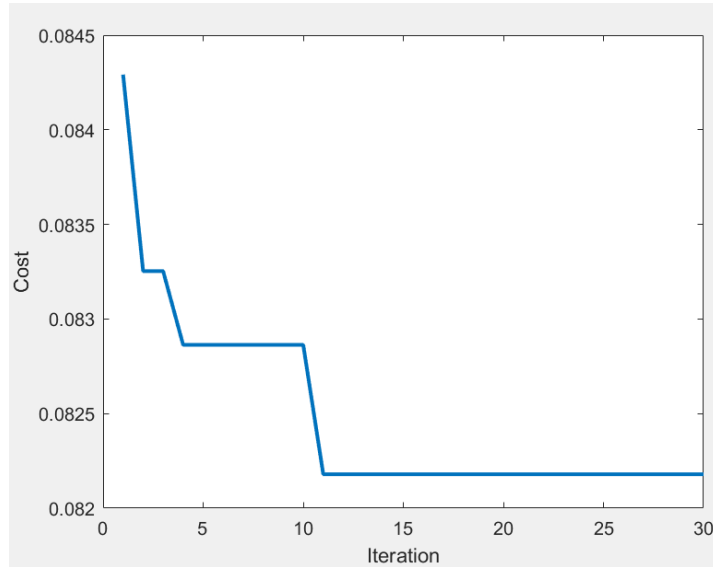
Şekil 4.17 Rastgele orman algoritması için karışıklık matrisi

4.8.4. Önerilen Gri Kurt Ve Rastgele Orman Algoritmasının Tekrarına Göre Maliyet Fonksiyonunun Diyagramı

Grafik 4.1 ve Grafik 4.2'de tekrara dayalı maliyet fonksiyonu gösterilmekte olup, Grafik 4.1 Cleveland verilerinin tekrarına göre maliyet fonksiyonu grafiği, Grafik 4.2 ise Farmingham verilerinin tekrarına göre maliyet fonksiyonu grafiğidir.



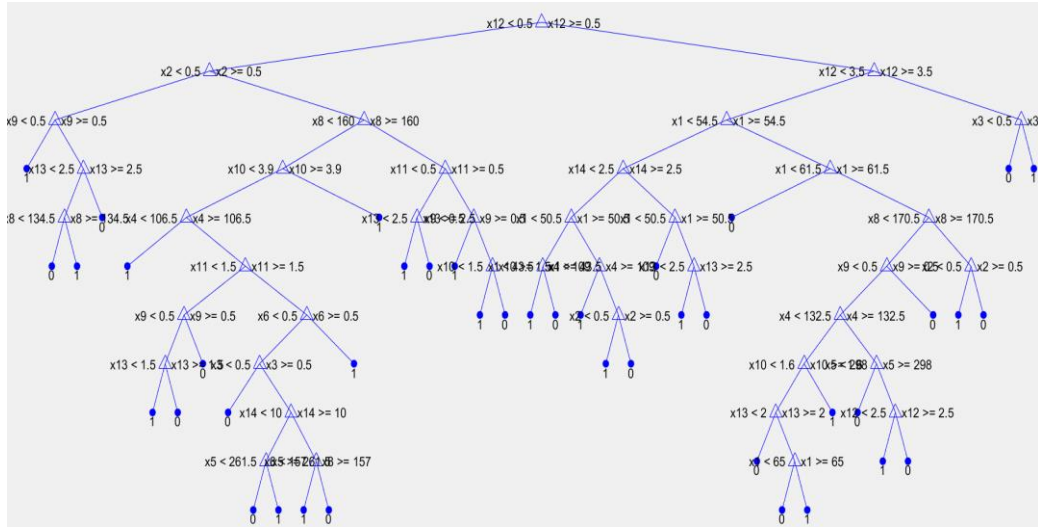
Grafik 4.1 Cleveland verilerinin tekrarına göre maliyet fonksiyonu grafiği



Grafik 4.2 Farmingham verilerinin tekrarına göre maliyet fonksiyonu grafiği

4.8.5. Önerilen Algoritmada Rastgele Orman Algoritmasını Oluşturan Karar Ağaçları

Üçüncü bölümde bahsedildiği gibi rastgele orman algoritması çok sayıda karar ağacından oluşur. Rastgele orman algoritması zayıf öğrenme şeklinde çalışmakta ve son olarak karar ağaçları arasında oylama yapılmaktadır. Bu amaçla rastgele orman algoritmasında kullanılan karar ağaçları, ağaç görüntüleme komutu kullanılarak kodun çıktısında görülebilmektedir. Rastgele orman algoritmasında kullanılan ağaç sayısı 10'dur. Şekil 4.18 'de bu rastgele ormanın 1 numaralı ağacı gösterilmektedir.



Şekil 4.18 Rastgele orman algoritmasında karar ağacı

4.9. Önerilen Gri Kurt Ve Rastgele o-Orman Algoritmasının Farklı Uygulamalardaki sonuçları

Önerilen algoritmanın performansını farklı uygulamalarda başlangıç durumuna kıyasla karşılaştırıyoruz. Her uygulamada kaldırılan özellikler farklıdır. Tablo 4.3, Cleveland verileri için değerlendirme kriterlerini kullanarak önerilen algoritmanın performansını rastgele orman algoritmasıyla karşılaştırmaktadır. Rastgele orman algoritmasına sahip gri kurt algoritmasının kesinliği, doğruluğu, duyarlılığı, F1_score'u ve karakteristik performans grafiğinin altındaki alanı, rastgele orman algoritmasının kendisiyle karşılaştırıldığında Tablo 4.3'te gösterilmektedir.

Tablo 4.3 Önerilen algoritmanın Cleveland verilerinin rastgele orman algoritması ile performans karşılaştırması

Method	Accuracy	Precision	Recall	F1_Score	AUC
GWO-RF	0,9472	96,927	91,83	94,0149	0,986
RF	0,9241	91,9708	91,3043	91,644s	0,9674

Gri kurt ve rastgele orman algoritmasının Cleveland verileri üzerinde çalıştırılmasının kaldırılan özellikleri Tablo 4.4'te gösterilmektedir.

Tablo 4.4. Cleveland verilerinin önerilen algoritmasında kaldırılan özellikler

Kod yürütme	Kaldırılan özellikler
Önerilen algoritma	Cinsiyet, talasemi

Tablo 4.5 ayrıca önerilen algoritmanın performansını Framingham verileri için değerlendirme kriterlerini kullanan rastgele orman algoritmasıyla karşılaştırmaktadır.

Tablo 4.5 Önerilen algoritmanın Farmingham verilerinin rastgele orman algoritması ile performans karşılaştırması

Method	Accuracy	Precision	Recall	F1_Score	AUC
GWO-RF	0,9402	93,5911	99,7779	96,5869	0,899
RF	0,9199	92,2758	98,820	95,54	0,8765

Framingham verileri üzerinde gri kurt ve rastgele orman algoritmasında kaldırılan özellikler Tablo 4.6'da gösterilmektedir.

Tablo 4.6 Farmingham verilerinin önerilen algoritmasında kaldırılan özellikler

Kod yürütme	Kaldırılan özellikler
Önerilen algoritma	Cinsiyet, doktora gitmek, glikoz

4.10. Sonuç

Bu bölümde, gri kurt optimizasyon algoritması ile özellik seçimi kullanılarak ve bunu rastgele orman algoritması ile birleştirerek, rastgele orman algoritmasının kalp hastalıklarını tahmin etmedeki performansı artırılmıştır. Doğruluğun ve diğer değerlendirme kriterlerinin iyileştirilmesi de gri kurt algoritması kullanılarak özellik seçiminin avantajlarından biridir. Bu araştırmanın sonuçları, önerilen yöntemin geçmişte yapılan diğer çalışmalara göre doğruluk, duyarlılık ve verimlilik açısından oldukça iyi çalıştığını ve rastgele orman sınıflandırma yönteminden daha iyi performans gösterdiğini sergilemiştir. Ayrıca önerilen algoritmanın değerlendirmelerine ilişkin tüm grafikler Cleveland ve Farmingham olmak üzere iki farklı veri seti üzerinde çizilmiştir. Öte yandan sonuçları daha da güçlendirmek için gri kurt algoritması ile farklı sınıflandırıcılar kullanılabilir. Şu ana kadar gri kurt algoritmasıyla birlikte çok sayıda sınıflandırma algoritması kullanıldı. Bu algoritmaların birbirleriyle birleştirilmesi başarılı olmuş, doğruluk ve diğer değerlendirme kriterlerinin iyileştirilmesine yol açmıştır. Ayrıca, önerilen yöntemin sağlamlığını daha fazla araştırmak için gelecekte başka bazı veri setleri de kullanılabilir.

5. SONUÇ VE ÖNERİLER

5.1. Giriş

Bu bölümde gri kurt algoritması ile rastgele orman algoritmasının birleşiminden elde edilen sonuçlar analiz edilecek ve hedef veriler üzerinde tek başına uygulanan rastgele orman algoritması ile karşılaştırılacaktır. Dördüncü bölümde verilen değerlendirme kriterleri ve çıktı grafikleri bu bölümde incelenmektedir. Kombine algoritmanın, rastgele orman sınıflandırma algoritmasının tek başına kullanıldığı duruma göre performans artışı grafik ve tablolarda incelenmiştir. Devamında gri kurt algoritmasının diğer sınıflandırma algoritmalarıyla kombinasyonuna ilişkin önerilerde bulunulmuştur. Ayrıca M-Serbram platformu ve onun diğer araçlarla bağlantılı olarak geliştirilmesi hakkında bazı noktalara değinilmiştir.

5.2. Performans Karakteristik Diyagramının Analizi

Performans karakteristik eğrileri ile ilgili olan dördüncü bölümde yer alan Şekil 4.9 ve Şekil 4.10'e göre, önerilen gri kurt algoritması ve rastgele orman algoritmasına ilişkin performans karakteristik eğrisinin, rastgele orman algoritmasının performans karakteristik eğrisi bire yakındır. Eğrinin performansını açıklamak için en iyi sınıflandırma modelinin eğrinin bire yakın olduğu model olduğu söylenebilir. Bu, bire ne kadar yakınsa ölçümün o kadar doğru ve uygun olduğu anlamına gelir. Dolayısıyla önerilen gri kurt algoritması ile rastgele orman algoritmasının performans karakteristik eğrisinin bire yakın olması, önerilen bu algoritmanın daha doğru bir ölçüm yaptığını ve daha iyi tahmin ettiğini göstermektedir. Hem Cleveland hem de Farmingham verileri üzerindeki performans karakteristik eğrisinin incelenmesi, önerilen algoritmanın performansının rastgele orman algoritmasına kıyasla arttığını doğrulamaktadır.

5.3. Performans karakteristik diyagramının altındaki seviyenin kriter analizi

Önerilen gri kurt algoritması ile rastgele orman algoritması için Cleveland verilerinin performans karakteristik diyagramları altındaki seviyenin karşılaştırılması 0,98'e, ayrıca rastgele orman algoritmasına ilişkin performans karakteristik diyagramı altındaki seviyenin 0,96'ya eşit olduğu görülmektedir. Bu da algoritmanın ne kadar verimli göstermektedir. Alt seviyenin daha fazla grafiğe sahip olması nedeniyle önerilen algoritma rastgele orman algoritmasından daha iyidir. Ayrıca Framingham

verileri için önerilen gri kurt algoritması ve rastgele orman algoritmasının performans karakteristik grafikleri altındaki düzey 0,9'a eşittir. Rastgele orman algoritmasına ilişkin performans karakteristik diyagramının altındaki düzey de 0,89'a eşittir. Bu karşılaştırma, önerilen algoritmanın verimliliğinin, grafiğin altındaki yüzeyin daha fazla olması nedeniyle rastgele orman algoritmasından daha iyi olduğunu göstermektedir. bu sayının değeri ne kadar yüksekse nihai performans o kadar olumlu değerlendirilir. Performans özelliği alt diyagramının seviye kriteri, performans özelliği diyagramı altındaki seviyeyi gösterir. Performans karakteristik diyagramı, sınıflandırıcıların etkinliğini kontrol etmeye yönelik bir yöntemdir. Performans karakteristik eğrisinin altındaki alan, performansın bir yönünü ölçen bir sayıdır. Hem Cleveland hem de Framingham verileri üzerinde performans karakteristik diyagramının altındaki seviyenin incelenmesi, önerilen algoritmanın performansının rastgele orman algoritmasına kıyasla arttığını doğrulamaktadır.

5.4. Karışıklık Matrisi Analizi Ve Değerlendirme Kriterleri

Dördüncü bölümde yer alan karışıklık matrisleri ve bu matrislerde ifade edilen değerlendirme kriterlerine göre, gri kurt ve rastgele ormanı birleştiren önerilen algoritmanın performansı, rastgele orman algoritması ile karşılaştırmalı olarak incelenmiştir. Bilindiği gibi önerilen algoritmanın kesinlik, doğruluk, duyarlılık, hata oranı ve F₁-score kriterleri her iki veri setinde de iyileştirilmiştir. Bahsedilen tüm kriterler, karışıklık matrisinin satır veya sütunlarının toplamından elde edilir. Bir hastalığı veya olayı tahmin etme şeklindeki makine öğrenmesi problemlerinde karışıklık matrisinin temsili oldukça önemlidir. Gri kurt algoritması, gereksiz özellikleri ortadan kaldırarak, algoritma her çalıştırıldığında verilere uygulanan sınıflandırma algoritmasının verimliliğini artırabilir. Bu tezdeki karışıklık matrisleri iki sınıflı veriler içindir.

Bunlardan biri aslında gerçek sınıflarımız olan hedef sınıf, diğeri ise problem sınıflandırmasının çıktısına karşılık gelen çıktı sınıfıdır. Önerilen gri kurt ve rastgele orman algoritması için karışıklık matrisi olan Şekil 4.14'e göre, sınıf 1'in kalp hastası olmakla, sınıf 2'nin ise kalp hastası olmamakla ilgili olduğu anlaşılmaktadır. Bu matriste birinci ve ikinci satır ile birinci ve ikinci sütunda yer alan dört sayının toplamı, toplam kişi sayısına eşittir. İlk sütunda yer alan doğru pozitif ve yanlış negatif sayıların

toplamı, toplam kalp hastası sayısına eşit olup, ana sayılarımız olan hedef eksenine yerleştirilmiştir. Ayrıca ikinci sütunda yer alan doğru negatif ve yanlış pozitif sayıların toplamı, sağlıklı olan ve kalp hastalığı olmayan kişilerin toplam sayısına eşittir. Cleveland verilerine ilişkin karışıklık matrisi olan Şekil 4.14'te matristeki sayıların toplamı veride bulunan tüm kişilere eşittir, yani 303 kişiye eşittir. Bu kişilerden 138'i kalp hastası, 165'i ise sağlıklıdır. İlk satırda yer alan sayıların toplamı, aslında önerilen gri kurt ve rastgele orman algoritması ile elde edilen kalp hastalarının tahminiyle ilgili olan doğru pozitif ve yanlış pozitif sayıların toplamına eşittir. Önerilen algoritmada yanlış pozitif kişi sayısı 4 kişiye yani %1,3'e, yanlış negatif kişi sayısı ise 12 kişiye yani %4'e eşittir. Şekil 4.15 rastgele orman algoritmasının Cleveland verilerine uygulanmasına ilişkin karışıklık matrisiyle ilgilidir. Bu matriste yanlış pozitif kişi sayısı 11 kişi yani %3,6, yanlış negatif kişi sayısı ise 12 kişi yani %4'tür. Bilindiği üzere gri kurt algoritmasını rastgele orman algoritması ile birleştiren önerilen algoritmanın hata yüzdesi azalmıştır. Şekil 4.16 ve Şekil 4.17, Framingham verilerine ilişkin karışıklık matrisleridir. Framingham verilerine 2123 kişi katıldı. Bunlardan 1801 kişi sağlıklı, 323 kişi ise kalp hastasıdır. Şekil 4.16, gri kurt ve rastgele ormanı birleştiren önerilen algoritmanın uygulanmasına ilişkin karışıklık matrisiyle ilgilidir. Bu matriste yanlış pozitif kişi sayısı 123 kişi yani %5,8, yanlış negatif kişi sayısı ise 4 kişi yani %0,2'dir. Şekil 4.17 aynı zamanda rastgele orman algoritmasının karışıklık matrisini göstermektedir. Bu matriste yanlış pozitif kişi sayısı 149 kişiye yani yüzde 7'ye, yanlış negatif kişi sayısı ise 21 kişiye yani yüzde 1'e eşit oluyor. Bu nedenle bilindiği üzere gri kurt algoritmasını rastgele orman algoritmasıyla birleşiminden oluşan önerilen algoritmanın hata yüzdesi azalmıştır.

5.5 Tekrara Göre Maliyet Fonksiyon Grafiği Analizi

Grafik 4.1, Cleveland verileri için gri kurt ile rastgele orman algoritmasını birleştiren önerilen algoritmanın tekrarlama açısından maliyet fonksiyonuyla ilgilidir. Bu algoritmadaki maliyet fonksiyonu seçilen özelliklerin değişimine göre farklılık göstermektedir. Önerilen algoritmadaki yineleme sayısı 30 yinelemeye eşittir. Grafik 4.1'te görüldüğü gibi başlangıçta maliyet fonksiyonunun değeri 0,103'e eşitken, ikinci yinelemede maliyet fonksiyonunun değeri 0,1'e düşürülmektedir. Daha sonra beşinci yinelemede maliyet fonksiyonunun değeri 0,98'e düştü. On üçüncü yinelemede maliyet fonksiyonu tekrar değişir ve değeri 0,94 olur. Bundan sonra maliyet

fonksiyonu deęiřmeyi bırakır ve maliyet fonksiyonu yakınsar ve 0,94 sabit deęerinde kalır. Meta-sezgisel algoritmalarda önemli noktalardan biri bu algoritmaların yakınsama eğilimidir. Maliyet fonksiyonunun yazılmasındaki amaç, algoritmanın tahmindeki doğruluęunu arttırmak ve hata oranını azaltmaktır. Bu nedenle maliyet fonksiyonu yazılırken biri önerilen algoritmanın doğruluęu, dięeri ise seçilen özellikler olmak üzere iki ana faktör ele alınmıştır. Bu iki faktörün seçimini dengelemek için ağırlıklı katsayılar kullanılmıştır. Bu faktörün artırılıp azaltılmasıyla bu iki faktörün etkisi artırılabilir veya azaltılabilir. Bu ilişki bir önceki bölümde maliyet fonksiyonunun hesaplanmasında kullanılan denklem (4.1) başlığı altında belirtilmiştir.

Maliyet fonksiyonunun hesaplanmasında doğruluęun önemini arttırmak amacıyla ortalamayla çarpılan w_1 katsayısı daha büyük kabul edilir, seçilen özelliklerle çarpılan w_2 katsayısı ise daha küçük alınır. Bu şekilde w_1 katsayısı 0,995'e, w_2 katsayısı ise 0,005'e eşit kabul edilir.

Grafik 4.2, Framingham verileri için gri kurt ile rastgele orman algoritmasını birleřtiren önerilen algoritmanın yineleme açısından maliyet fonksiyonuyla ilgilidir. Grafik 4.2 de görüldüęü gibi başlangıçta maliyet fonksiyonunun deęeri 0,84'e eşit olup, ikinci yinelemede seçilen özellikler deęiřtirilerek maliyet fonksiyonunun deęeri 0,83'e düşürölür. Üçüncü yinelemede maliyet fonksiyonunun deęeri 0,825'e eşit olur. Son olarak on birinci yinelemede maliyet fonksiyonunun deęeri 0,82 olur ve bu yinelemede seçilen özellikler sabit kalır ve deęiřmez. Meta-sezgisel algoritmalar yakınsama eğilimindedir. Dolayısıyla önerilen algoritmanın maliyet fonksiyonu sabit kalmakta ve on birinci yinelemede deęiřmemektedir. Maliyet fonksiyonunun deęerleri de farklı yinelemelerde denklem (4.2)'ten elde edilir.

5.6. Karar Ağaçlarının Analizi

Rastgele orman algoritması çok sayıda karar ağacından oluşur. Rastgele orman algoritması zayıf öğrenme şeklinde çalışmakta ve son olarak karar ağaçları arasında oylama yapılmaktadır. MATLAB yazılımındaki view komutu kullanılarak karar ağaçları göröllebilmektedir. Şekil 4.18'te karar ağaçlarından biri çizilmiştir. Ağaç toplayıcı fonksiyonu kullanılarak karar ağaçları arasında oylama yapılır. Tahmin sırasında, yani tahmin için algoritmaya bir dizi yeni veri verildiğinde, öğrenilen bu algoritmaların her biri bir sonucu tahmin eder. Son olarak, rastgele orman algoritması

en çok oyu alan sınıfı oylayarak seçebilir ve onu sınıflandırma işlemi için son sınıf olarak yerleştirebilir. Önerilen gri kurt algoritması ile rastgele ormandan elde edilen ağaçların yanı sıra rastgele orman algoritmasının ağaçları da karşılaştırıldığında, önerilen algoritmadaki özellik sayısının azaldığı sezgisel olarak görülebilmektedir.

5.7. Önerilen Gri Kurt Ve Rastgele Orman Algoritması Tarafından Kaldırılan Özelliklerin Analizi

Cleveland'ın verilerinden çıkarılan özellikleri gösteren önceki bölümdeki Tablo 4.4'e göre, Cleveland'ın verilerinden çıkarılan özellikler cinsiyet ve Akdeniz anemisiye eşittir. Kardiyologlardan Sayın Kiyomarth Majed'in görüşüne göre cinsiyet, kalp hastalığının teşhisinde etkili ve doğru bir özellik değildir, çünkü bu özellik farklı millet ve kültürlerde farklı olabilmekte ve erkeklerde kalp hastası sayısının her zaman kadınlara göre daha fazla olduğu kesin olarak söylenemez. Kaldırılan ikinci özellik Akdeniz anemisi özelliği ile ilgilidir. Bu özellik, kalp fonksiyonlarında bozulmalara neden olarak kalp atışlarında ve kalbe kan akışında düzensizliğe neden olur. Çoğunlukla kalp damarlarının yağ biriktirmesi nedeniyle ortaya çıkan tıkanmayla ilişkili olan koroner arter hastalığının tanısında hayati bir faktör olarak değerlendirilemez. Framingham verilerinden kaldırılan özellikleri gösteren önceki bölümdeki Tabloya 4.6' dan, kaldırılan özelliklerden bazılarını görebiliriz. Framingham deneyinde kaldırılan özellikler cinsiyet, doktora görünme ve glikozdu. Cleveland verilerinde cinsiyet faktörünün çıkarılmasıyla ilgili olarak, cinsiyetin kalp hastalıkları tanısında etkili ve doğru bir özellik olmadığı, bu özelliğin farklı millet ve kültürlerde farklı olabileceği ve erkeklerde kalp hastası sayısının her zaman kadınlara göre daha fazla olduğu kesin olarak söylenemez.

Bir diğer husus da kardiyoloğa gitme meselesidir. Bu özellik kalp hastalıklarının etkenleri ve risk faktörlerinden biri olarak değerlendirilemez, Çünkü her insan yaşamı boyunca kalp doktoruna gidebilir ve bu kişinin kesinlikle kalp hastası olduğunun kanıtı değildir. Önerilen algoritmanın ilk uygulamasında verilerden çıkarılan bir diğer özellik ise glikoz özelliğidir. Bu özellik, Cleveland verilerindeki kan şekerinin hedef özelliği gibi, kişinin kalp hastalığına yakalanmasında tek başına bir etken olarak değerlendirilemez. Çünkü kan şekeri düzeyi tek başına kişinin diyabet hastası olup olmadığını belirleyemez. Bazen şeker oranı yüksek gıdaların tüketilmesi nedeniyle

kan testinde glikoz seviyesinin geçici olarak yüksek olduğu gösterilir fakat kişinin şeker veya kan şekeri problemi olmayabilir.

5.8. Öneriler Ve Gelecek Çalışmalar

Gri kurt adı verilen meta-sezgisel algoritmalarından birisinin rastgele orman adı verilen makine öğrenme algoritmalarından biri ile birleştirilmesi olan bu araştırmada gerçekleştirilen yeniliklere göre, meta-sezgisel algoritmaların makine öğrenmesi algoritmalarıyla birleştirilmesi alanında yeni fikirler ve planlar sunulabilir. Gri kurt algoritmasının 2014 yılında tanıtılan nispeten yeni bir algoritma olduğu göz önüne alındığında, üzerinde birçok tasarım ve fikir hayata geçirilebilir. Örneğin, bu algoritmanın diğer makine öğrenimi algoritmalarıyla birleştirilmesi, sınıflandırma algoritmalarının doğruluğunu ve verimliliğini artırabilir. Gri kurt algoritması verilerden uygun özelliği seçerek sınıflandırma algoritmasının etkinliğini artırabilir. Şimdiye kadar bu algoritmanın bir dizi sınıflandırma algoritmasıyla kombinasyonu kullanıldı. Örneğin, gri kurt algoritmasının destek vektör makinesi algoritmasıyla kombinasyonu meme kanserini teşhis etmek için kullanılmıştır. Ayrıca gri kurt algoritmasının ID3 karar ağacı algoritmasıyla birleştirilmesi sınıflandırma algoritmasının verimliliğini artırmaktadır.

Bu alanda önerilebilecek yeni fikirler gri kurt algoritmasının diğer sınıflandırma algoritmalarıyla birleşimi olabilir. Örneğin regresyon yöntemleri, doğrusal ve lojistik regresyon veya Bayes algoritması (Naïve Bayes) gibi denetimli öğrenme algoritmalarını gri kurt algoritmasıyla birleştirmek için kullanılabilir.

- 1) Gri kurt algoritmasının regresyon sınıflandırma algoritmasıyla birleştirilmesi
- 2) Gri kurt algoritmasının Bayes algoritması ile kombinasyonu
- 3) Gri kurt algoritmasının k-en yakın komşu algoritması ile kombinasyonu

Bu algoritma kombinasyonu şu şekildedir; öncelikle özellikleri seçmek için gri kurt algoritması kullanılır. Ardından makine öğrenmesi algoritmaları uygulanır. Gri kurt algoritması tarafından yapılan bu özellik seçimi ve ardından öğrenme algoritmalarının uygulanması, öğrenme algoritmalarının yalnızca verilere uygulandığı durumla karşılaştırıldığında algoritmanın doğruluğunu ve verimliliğini artırır.

Gelecekte üzerinde çalışılabilecek ve yenilik olarak sunulabilecek diğer plan ve fikirler arasında, çeşitli giyilebilir sensörlerin açık kaynak platformlara bağlantısının genişletilmesi de yer alıyor. Örneğin bu araştırmada açık kaynak kodlu bir platform

olan ve farklı giyilebilir sensörleri platforma bağlayabilme özelliğine sahip M Cerbaram platformu önerildi. Bu konuda Xiaomi bilekliğinin piyasadaki ucuz ve uygun maliyetli sensörlerden biri olan M Cerbram platformuna bağlanması konusunda çalışmalar yapılması mümkündür.

4) Xiaomi bileklik gibi giyilebilir sensörlerin açık kaynaklı platform M Cerbram'a bağlanması

M-Serbram platformuna ilişkin tüm kodlar ve uygulamalar bu platforma ilişkin sitede ve ayrıca GitHub sitesinde mevcuttur. Şu anda Microsoft bilekliği M Cerbram platformuna bağlanabilme özelliğine sahiptir. Ancak üretimlerinin tamamlanmış olması ve stoklarının çok az olması nedeniyle sensörlerle ilgili bilgilerin kaydedilmesi için Xiaomi bilekliklerinin kullanılması daha doğru olacaktır.

KAYNAKLAR

- Dehghan, M., Roostaei, M.(2013). Investigating wireless sensors for accurate and real-time remote health monitoring. The first general medicine conference, *Amirkabir University of Technology*. January 2013, Tehran.
- Vahdat, D. and Kaysari, M.(2017). Basics of Internet of Things. *Tehran, Ati Nagar Publications*, first edition.
- Gharib, H. and Sabzian, H.(2017). Fighting in the health system with the help of Internet of Things. *Donyaye Eghtesad newspaper*, June 2. news code: 3391694.
- Dr. Najafi, M.H.(2015). Heart angina or chest pain, , June 17, 2015. <https://drhsnajafi.com/2168>.
- Kaviani, M. (2020). Ensemble Classifier and the subject of Bagging and Boosting, <https://chistio.ir>
- Kaviani, M. (2020). ID3 decision tree algorithm and Entropy and Gain structure, 4.5C decision tree classification algorithm. <https://chistio.ir>.
- Google Fit activity and health tracking. <https://play.google.com/store/apps/>
- Examining the assessment criteria of the category. <https://tanoco.ir/datamining>
- Mann DL, Zipes DP, Libby P, Bonow RO. (2014). Braunwald's heart disease e-book: a textbook of cardiovascular medicine. *New York: Elsevier Health Sciences*; P. 2-10, 18-20.
- V. Chaurasia and S. Pal. Data Mining Approach to Detect Heart Diseases, *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*, Vol. 2, No. 4, Month Year, Page: 56-66, ISSN: 2296-1739.
- Q. Al-Tashi, S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian.(2019). Binary Optimization Using Hybrid Grey Wolf Optimization for Feature Selection, *IEEE Access*, vol. 7, pp. 39496–39508, doi: 10.1109/ACCESS.2019.2906757.
- P. S. Game, V. Vaze, and M. Emmanuel.(2019). Induction of decision trees based on gray wolf optimizer for heart disease classification, *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 4S2, pp. 168–173.

- Q. Al-Tashi, H. Rais, and S. Jadid.(2019). Feature selection method based on grey wolf optimization for coronary artery disease classification, *Adv. Intell. Syst. Comput.*, vol. 843, pp. 257–266. doi: 10.1007/978-3-319-99007-1_25.
- S. R. Kamel, R. Yaghoubzadeh, and M. Kheirabadi. (2019). Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer, *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0247-7.
- B. Sankara Babu, A. Suneetha, G. Charles Babu, Y. Jeevan Nagendra Kumar, and G. Karuna.(2018). Medical disease prediction using grey wolf optimization and auto encoder based recurrent neural network, *Period. Eng. Nat. Sci.*, vol. 6, no. 1, pp. 229–240. doi: 10.21533/pen. v6i1.286.
- S. Kumar and A. Ann Romalt. (2020). An Analysis On Feature Selection Methods, Clustering and Classification Used in Heart Disease Prediction- A Machine Learning Approach, *Journal of Critical Reviews*, Vol 7, Issue 6, 2020, doi: <http://dx.doi.org/10.31838/jcr.07.06.27>.
- P. Sharma, S. Sundaram, M. Sharma, A. Sharma, and D. Gupta.(2018). Diagnosis of Parkinson's disease using modified grey wolf optimization. *Cogn. Syst. Res.*, vol. 54, pp. 100–115, 2019, doi: 10.1016/j.cogsys.2018.12.002.
- S. Chatterjee, K. Hovsepian, et al.(2016). MCrave: continuous Estimation of craving During Smoking Cessation, *ACM Ubicomp*.
- A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar.(2012). MPuff: Automated detection of cigarette smoking puffs from respiration measurements. *IPSN'12 - Proc. 11th Int. Conf. Inf. Process. Sens. Networks*, pp. 269–280. doi: 10.1145/2185677.2185741.
- N. Saleheen et al.(2016). MSieve: Differential behavioral privacy in time series of mobile sensor data. *UbiComp 2016 - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 706–717. doi: 10.1145/2971648.2971753.
- Barnali Sahu, Satchidananda Dehuri, Alok Jagadev.(2018). A Study on the Relevance of Feature Selection Methods in Microarray Data. *The Open Bioinformatics Journal*, Volume 11, Issue 1, Pages: 117-139.
- Beni G, Wang J.(1993). Swarm intelligence in cellular robotic systems. In: *Robots and biological systems: towards a new bionics*. ed. Springer. p. 703–12.

- Holland JH.(1992). Genetic algorithms. *Sci Am*; 267:66-72.
- Kennedy J, Eberhart R.(1995). Particle swarm optimization, in *Neural Networks*, In: *Proceedings. IEEE international conference on*; p. 1942–1948.
- Abd AL, El Hafeez T, Zaki AM.(2018). Face recognition based on GreyWolf Optimization for feature selection. *In international conference on advanced intelligeny systems and informatics*, pp 273-283.
- Tu Q, Chen X, Liu X.(2019). Multi-strategy ensemble Grey Wolf Optimizer and its application to feature selection. *App Soft Comput*. 76:16-30.
- <https://corporatefinanceinstitute.com/resources/knowledge/other/bagging-bootstrap-aggregation/>, What is Bagging (Bootstrap Aggregation)?
- Leo Breiman.(2001). Random forest, *statistic department, university of Califrnia, Berkeley, CA 94720*, January 2001.
- <https://www.pcmag.com/reviews/google-fit-for-android>, Google Fit (for Android) Review.
- S. M. Hossain et al.(2017). MCerebrum: A mobile sensing software platform for development and validation of digital biomarkers and interventions. *SenSys 2017 - Proc. 15th ACM Conf. Embed. Networked Sens. Syst.*, vol. 2017-January. doi: 10.1145/3131672.3131694.
- S. Kumar and G. Sahoo.(2017). A random forest classifier based on genetic algorithm for cardiovascular diseases diagnosis. *Int. J. Eng. Trans. B Appl.*, vol. 30, no. 11, pp. 1723–1729. doi: 10.5829/ije.2017.30.11b.13.
- P. P. R. Patil and P. S. A. Kinariwala.(2017). Automated Diagnosis of Heart Disease using Data Mining Techniques. *Int. J. Adv. Res. Ideas Innov. echnology*, vol. 3, no. 2, pp. 560–567.
- S. Vachiravel.(2015). Diagnosis of Heart Disease Using Decision Tree and SVM Classifier. *Int. J. Appl. Eng. Res.*, vol. 10, no. March, pp. 598–602.
- F. Abdeldjouad and M. Brahami. (2020). A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques. *Lecture Notes in Computer Science book series, LNCS*, vol. 12157.
- S. Dehnavi and M. Emamipour. (2020). Combining Random Forest and Neural Networks Algorithms to Diagnose Heart Disease. *HamaraJournals Login*

Register Frontiers in Health Informatics, vol. 30. doi: 10.30699/fhi.v9i1.214.

A. Akella and V.Kaushik. (2020). Machine Learning Algorithms for predicting Coronary Artery Disease: Efforts Toward an Open Source Solution. *bioRxiv*, doi:10.1101/2020.02.13.948414.

<https://framinghamheartstudy.org/fhs-for-researchers/policies-and-procedures/data-sharing-principles>, Data Sharing Principles.

[https://www.diabetes.ca/managing-my-diabetes/tools---resources/body-mass-index-\(bmi\)](https://www.diabetes.ca/managing-my-diabetes/tools---resources/body-mass-index-(bmi)), Body Mass Index (BMI).

<https://www.mathworks.com/help/stats/treebagger.html>, TreeBagger.

<https://towardsdatascience.com/why-and-how-to-cross-validate-a-model>, Why and how to Cross Validate a Model?.

M. Niu and Y. Li, RFamyloid. (2018). A Web Server for Predicting Amyloid Proteins”, *Int. J. Mol. Sci.* 2018, 19, 2071; doi:10.3390/ijms19072071.

<https://medium.com/analytics-vidhya/complete-guide-to-machine-learning-evaluation-metrics>, Complete Guide to Machine Learning Evaluation Metrics.

<https://towardsdatascience.com/understanding-auc-roc-curve>, Understanding AUC - ROC Curve.

S. Mirjalili, S. M. Mirjalili, and A. Lewis. (2014). Grey Wolf Optimizer. *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014, 2013, doi: 10.1016/j.advengsoft.2012.007.

TÜRKÇE - İNGİLİZCE SÖZLÜK

Makine Öğrenimi Algoritmaları	Machine Learning Algorithms
Yapay Zeka	Artificial Intelligence
Hibrit Algoritmalar	Ensemble Algorithms
Meta-Sezgsel Algoritmalar	Metaheuristic Algorithms
Değerlendirme Kriterleri	Validation Metrics
Gri Kurt Optimizasyon Algoritması	Grey Wolf Optimization Algorithm
Veri Seti	Data Set
Kablosuz Vücut Sensörü Ağları	Wireless Body Area Network (WBAN)
Bulut Bilişim	Cloud Computing
İvmeölçer	Accelerometer
Bulaşıcı Olmayan Hastalık	Non-Communicable Diseases
Göğüs Ağrısı	Angina
Veri Analizi	Data Mining
Karar Ağacı	Decision Tree
Nöral Ağlar	Neural Network
Vektör Makineleri Desteklemek	Support Vector Machine
Doğrusal Regresyon	Linear Regression
Lojistik Regresyon	Logistic Regression
Rastgele Orman Algoritması	Random Forest Algorithm
K En Yakın Komşu Algoritması	K Nearest Neighbour
Genetik Algoritma	Genetic Algorithm
Kombine Sınıflandırma	Ensemble Classifier
Sınıflandırmak, Torbalama	Bagging

Güçlendirme, Artırma	Boosting
Kümeleme	Clustering
Kesinlik	Accuracy
Duyarlılık	Sensitivity
Karınca Kolonisi Optimizasyonu	Ant Colony Optimization
Parçacık Sürü Optimizasyonu	Particle Swarm Optimization
Sürü Zekası	Swarm Intelligence
Evrimsel Algoritmalar	Evolutionary Algorithms
Arı Kolonisi Optimizasyonu	Bee Colony Algorithm
Yarasa Algoritması	Bat Algorithm
Keşif	Exploration
Sömürü	Exploitation
Gradyan Azaltması	Steepest Descent
Rastgele Arama	Random Search
Maliyet Fonksiyonu	Cost Function
K-katlı Çapraz Doğrulama	K-fold Cross Validation
Performans Karakteristik Eğrisi	Receiver Operating Characteristic Curve
Performans Karakteristik Diyagramının Altındaki Seviye	Area Under Cover
Karışıklık Matrisi	Confusion Matrix
Düşük yoğunluklu Lipoprotein	Low Density Lipoprotein
Yüksek Yoğunluklu Lipoprotein	High Density Lipoprotein
Vücut Kitle İndeksi	Body Mass Index
Ortalama Parçacık Hacmi	Mean Corpuscular Volume

Benzetimli Tavlama	Simulated Annealing
Negatif Tahmin Deęeri	Negative Predictive Value
Pozitif Tahmin Deęeri	Positive Predictive Value
Gerçek Negatif Oranı	True Negative Rate
Gerçek Pozitif Oranı	True Positive Rate
Yanlış Negatif Oranı	Miss rate or False Negative Rate
Yanlış Pozitif Oranı	False Positive Rate
Bal Arısı Koloni Algoritması	Artificial Bee Colony Algorithm
Genetik Programlama	Genetic Programming
Tekrarlayan Sinir Ağları	Recurrent Neural Network