

**T.C
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**MAKİNE ÖĞRENME YÖNTEMLERİNİ KULLANARAK SİRİAL TEST İLE
PARKİNSON HASTALIĞININ TANI TAHMİNİ**

Mustafa GERGER

ELEKTRİK – ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**ŞANLIURFA
2019**

Dr. Öğr. Üyesi Abdulkadir GÜMÜŞÇÜ danışmanlığında, Mustafa GERGER'in hazırladığı “**Makine Öğrenme Yöntemlerini Kullanarak Spiral Test ile Parkinson Hastalığının Tanı Tahmini**” konulu bu çalışma 18/06/2019 tarihinde aşağıdaki jüri tarafından oy birliği ile Harran Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

İmza

Danışman : Dr. Öğr. Üyesi Abdulkadir GÜMÜŞÇÜ

Üye : Doç. Dr. İ. Berkan AYDİLEK

Üye : Dr. Öğr. Üyesi Seydi KAÇMAZ

Bu Tezin Elektrik Elektronik Mühendisliği Anabilim Dalında Yapıldığını ve Enstitümüz Kurallarına Göre Düzenlendiğini Onaylarım.

Doç. Dr. İsmail HİLALİ
Enstitü Müdürü

Bu çalışma HÜBAK Tarafından Desteklenmiştir.
Proje No: 18097

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

İÇİNDEKİLER

	Sayfa No
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	iv
ÇİZELGELER DİZİNİ	v
SİMGELER DİZİNİ	vi
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	5
3. MATERYAL VE METOT	11
3.1. Veri Seti	11
3.2. Makine Öğrenme Algoritmaları	13
3.2.1. k - en yakın komşu (k-EK) algoritması	14
3.2.2. Karar ağacı sınıflandırma metodu	16
3.2.3. Genetik algoritma ile öz nitelik seçme	19
3.2.3.1. Genetik algoritmalar temel kavramlar	21
3.2.3.2. Genetik algoritmanın uygulanması	21
3.2.3.3. Seçim (Çoğalma)	24
3.2.3.4. Çaprazlama	24
3.2.3.5. Mutasyon	24
3.2.4. Birini dışarıda bırak çapraz doğrulama	25
4. ARAŞTIRMA BULGULARI ve TARTIŞMA	27
4.1. Deneysel Çalışmalar	27
4.2. Araştırma Bulguları	33
5. SONUÇLAR ve ÖNERİLER	38
KAYNAKLAR	41
ÖZGEÇMİŞ	44

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENME YÖNTEMLERİNİ KULLANARAK SİRİAL TEST İLE PARKİNSON HASTALIĞININ TANI TAHMİNİ

Mustafa GERGER

Harran Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik – Elektronik Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Abdulkadir GÜMÜŞÇÜ
Yıl: 2019, Sayfa: 44

Parkinson hastalığı son zamanlarda ciddi bir sağlık sorunu haline gelmiştir. Etkileri özellikle 65 yaş üstü hastalarda gittikçe artmaktadır. Hastalığın motor aktiviteler üzerindeki olumsuz etkileri hastaları sosyal ortamlardan uzaklaştırmaktadır. Tanısı oldukça zor olan Parkinson hastalığının teşhisinde bilgisayar destekli sistemlerin kullanımı artmaktadır. Parkinson spiral testi ile teşhis bu yöntemlerden biridir. Bu çalışmada, Parkinson hastalığının tanısında çizilen spirallerin örüntü tanıma yöntemleriyle analiz edilmesi önerilir. Hastalar tarafından çizilen spiraller karakter olarak kabul edilmiştir. Bu şekilde, veri setine dönüştürüldüğünde yüksek p düşük n problemi ortaya çıkmaktadır. Bu nedenle, öznelik seçiminin genetik algoritma ile etkisi araştırılmıştır. Sınıflandırma sürecinde k-EK (k - en yakın komşu) ve karar ağaçları yöntemleri kullanılmıştır. Sonuçlar, birini dışarda bırak çapraz doğrulama yöntemiyle doğrulanmıştır. Bu çalışmada, İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı tarafından hazırlanan veriler kullanılmıştır. Elde edilen deneysel sonuçlar önerilen yöntemin başarılı sonuçlar verdiğini göstermektedir. Spesifik olarak, doğruluk değeri, genetik algoritma ile öznelik seçme yöntemi karar ağacı sınıflandırmasında 1.00 olarak elde edilmiştir.

ANAHTAR KELİMELEER: parkinson tanısı, spiral testi, örüntü tanıma, sınıflandırma, genetik algoritma

ABSTRACT

MSc Thesis

DIAGNOSIS OF PARKINSON DISEASE WITH SPIRAL TEST USING MACHINE LEARNING METHODS

Mustafa GERGER

**Harran University
Graduate School of Natural and Applied Sciences
Department of Electrical and Electronics Engineering**

**Supervisor: Assist. Prof. Dr. Abdulkadir GÜMÜŞÇÜ
Year: 2019, Page: 44**

Recently, Parkinson's disease is a serious health problem. The incidence is increasing especially in people over 65 years of age. The adverse effects of the disease on motor activities distract patients from social environments. The diagnosis of Parkinson's disease is quite difficult. Therefore, computer-aided systems are used. Diagnosis by the Parkinson spiral test is one of these methods. In this study, it is recommended to analyze the spirals drawn in the diagnosis of Parkinson's disease by pattern recognition methods. The spirals drawn by the patients were assumed to be characters. In this way, a high p low n problem occurs when the data set is converted. Therefore, the effect of feature selection by genetic algorithm has been investigated. In the classification process, k-NN and decision trees were used. The results are validated by a leave one out cross validation method. In this study, the data set prepared by the Department of Neurology in Cerrahpasa Faculty of Medicine, Istanbul University was used. The experimental results obtained show that the proposed method gives successful results. Specifically, accuracy value obtained 1.00 in the decision tree classification by selecting the feature with the genetic algorithm.

KEYWORDS: parkinson diagnosis, spiral test, pattern recognition, classification, genetic algorithm

TEŞEKKÜR

Yüksek lisans öğrenimim boyunca bana her türlü desteği gösteren saygıdeğer danışman hocam Dr. Öğr. Üyesi Abdulkadir GÜMÜŞÇÜ'ye, çalışma hayatımın ilk dönemlerinde bana cesaret veren ve her türlü imkânı sağlayan ilk danışman hocam Dr. Öğr. Üyesi M. Akif NACAR'a ve hayatım boyunca bana maddi ve manevi desteklerini esirgemeyen babama, anneme ve eşime en içten duygularıyla teşekkür eder, sevgilerimi sunarım.

Bu çalışmamı, eşim Pınar GERGER'e ve kızım Bahar GERGER'e atfediyorum.



ŞEKİLLER DİZİNİ

Şekil 3.1. Veri Setinin Oluşturulması.....	12
Şekil 3.2. k-EK Algoritması Örnek Çizim	15
Şekil 3.3. Karar Ağacı Sınıflandırma Yöntemi	17
Şekil 3.4. Genetik Algoritma Genel Akış Diyagramı.....	23
Şekil 4.1. Parkinson Hastasının Statik Spiral Test Çizimi	27
Şekil 4.2. Parkinson Hastasının Dinamik Spiral Test Çizimi	28
Şekil 4.3. Sağlıklı İnsanın Statik Spiral Test Çizimi	28
Şekil 4.4. Sağlıklı İnsanın Dinamik Spiral Test Çizimi	29
Şekil 4.5. Çizim Verilerinden Elde Edilen Ham Veri Seti	30
Şekil 4.6. Gereksiz Verilerden Temizlenmiş Veri Seti	31
Şekil 4.7. Genetik Algoritma Tabanlı Öznitelik Seçme İşlemi Süreci	32
Şekil 4.8. GA Öznitelik Seçme ile k-EK Sınıflandırma Hata Oranı	36
Şekil 4.9. GA Öznitelik Seçme ile Karar Ağacı Sınıflandırma Hata Oranı.....	37



ÇİZELGELER DİZİNİ

Çizelge 3.1. Veri Setinin Tanımlanması	13
Çizelge 4.1. Karışıklık Matrisleri	34
Çizelge 4.2. Sınıflandırma Başarı Oranları	35
Çizelge 4.3. İşlem Süreleri	35
Çizelge 4.4. GA ile Seçilen Öznitelik Sayısı.....	36



SİMGELER DİZİNİ

DST	Dinamik Spiral Test
GA	Genetik Algoritma
KA	Karar Ağacı Algoritması
k-EK	k - En Yakın Komşu Algoritması
MKK	Matthews Korelasyon Katsayısı
PH	Parkinson Hastalığı
SST	Statik Spiral Test
STCP	Belirli Noktadaki Stabilitate Testi



1. GİRİŞ

Parkinson hastalığı (PH), sinir sistemini etkileyerek motor reflekslerinin, davranışların, konuşmaların ve diğer yaşamsal faaliyetlerin kısmen veya tamamen kaybolmasına neden olan nörolojik bir hastalıktır. 65 yaş ve üzerindeki her 100 kişiden birinde bu hastalık görülmektedir (Waters, 2000).

Yaşlanma ile birlikte ağırlaşan ve görülme oranının artış gösterdiği hastalığın en önemli belirtileri ses ve hareket fonksiyonlarındaki bozulmalardır. Okunaksız ve küçük yazmaya başlamak, konuşmalardaki bozukluklar, yavaş, ağır ve öne doğru eğik şekilde yürümek, depresyon ve yürürken kolların az sallanması gibi diğer belirtileri de vardır. Hastaların günlük işlerini yapmalarına engel olan bu belirtiler hastaların iş ve sosyal hayatından soyutlanmasına neden olurken hastalığın ileri dönemlerinde başkalarına muhtaç halde yaşamalarını gerektirmektedir. Bu durum hastaların psikolojilerini de olumsuz yönde etkilemektedir (Fahn ve Przedborski, 2000).

Parkinson hastalığı sinsi başlangıçlı olup, istirahat tremoru (istemli hareket), rijidite (pasif harekete karşı sürekli bir direnç bulunması), akinezi (insanda otomatik hareket yeteneğinin kaybolması), bradikinezi (istemli hareketlerin ileri derecede yavaşlaması) veya postural instabilite (postural refleks kaybı) gibi kardinal belirtileri ortaya çıkmadan önce nonspesifik semptomlardan oluşan prodromal bir dönem de söz konusu olabilir. Motor belirtilerden önce kişilik değişimleri ve çabuk yorulma da görülebilir. Hastalar kol ve omuzlarda ağrıdan ya da gerilmeden şikâyet edebilirler. Stres altında istirahat tremorun yakalanması, yürürken kolları az sallanması ve eş zamanlı tremorun meydana gelmesi gibi belirtiler muayene sırasında tanı için yardımcı bulgulardır.

Parkinson hastalığının başlangıcı genellikle çok tipiktir. Motor belirtiler çoğu zaman istirahat tremoru ile kendisini gösterir. Hastalık seyri boyunca değişik düzeylerde istemli hareketlerin ileri derecede yavaşlaması görülür. Hastalar dikkatle dinlenildiğinde günlük hayatlarında sıkıntılar yaşadıkları, alet kullanımında zorluk yaşadıklarını, yazı yazmada değişiklikler olduğu, vücutlarında genel olarak bir sertlik

ve yavaşlamanın var olduğu gözlemlenebilir. Bazı hastalarının konuşmalarında bozukluklar ve ses kayıpları yaşadıkları da belirlenebilir. Hastaların şikâyetleri arasında bradikinezi, rijidite, yürüyüş ve postür ile ilgili refleks mekanizmaların etkilenmesi ile ortaya çıkan oturduğu yerden zor kalkma, dengeyi korumada güçlük çekme, yüz ifadesinin değişmesi ve vücudunda kamburlaşma gibi yakınmalarda yer alır (Ertan, 2005).

İngiliz Doktor James Parkinson tarafından 1817 yılında ilk kez “shaking palsy” (titrek felç) şeklinde tanımlanan Parkinson hastalığı (PH), merkezi sinir sistemi bozukluğu olup çoğunlukla hastanın hareketlerinde, konuşmasında ve diğer vücut işlevlerinde zarara yol açar (Apaydın, 2013).

Hastalık 40-75 yaş civarında, genellikle de 60 yaşın üzerinde başlar ve görülme sıklığı da yaş arttıkça giderek artar. 65 yaş ve üzerindeki her 100 kişiden birinin PH olduğu kabul edilmektedir. Hastalık daha genç yaşlarda da görülebilir. Hastaların %5’inde semptomların 40 yaşından önce başlama durumu söz konusudur (Stephen ve ark., 2003).

Nörolojik bir sağlık sorunu olan Parkinson hastalığı, dünyada 4 milyondan fazla kişiyi etkilemektedir. Alzheimer’dan sonra belli bir yaş grubunda en sık görülen hastalık olan Parkinson hastalığının ülkemizde 130 binden fazla kişide var olduğu tahmin edilmektedir.

Parkinson, beyindeki “dopamin” denilen bir maddenin eksikliğiyle kendini gösterip sinsi seyreden ve sürekli ilerleyen bir hastalıktır. Öyle ki hastalar çoğu zaman hastalığın başlangıç tarihini kesin olarak söyleyemezler (Little ve ark., 2009). PH, ruhsal durum çöküntülerine ve bilişsel bozukluklara sebep olur. Kişinin motor kabiliyetlerine, konuşmasına ve diğer fonksiyonlarına zarar verir. Araştırmalar hastalığın en önemli semptomlarının ses kısıklığı ve bozukluğu ile yürüyüş değişikliği olduğunu göstermiştir. Özellikle artikülasyon ve solunum fonksiyonlarının Parkinson hastalığından etkilenmesi sonucu konuşma kalitesinde belirgin azalmalar görülmektedir.

Parkinson hastalığının, tremor (titreme), rijidite (kaslarda sertleşme) ve bradikinezi (hareketlerde yavaşlama) olmak üzere üç temel belirtisi bulunmaktadır. Bu temel belirtilerin yanında yürüme bozukluğu, yutma güçlüğü, otonom fonksiyon bozukluklarına ilişkin belirtiler (ağızda salya birikmesi, kabızlık, idrar yapma sorunları, cinsel fonksiyon bozuklukları gibi), gözlerde yanma, görme bozuklukları, ağrı ve duyuşal yakınmalar ve depresyon ile kişiyi etkilemektedir (Tugwell, 2008).

Hastalıkta görülen bu belirtiler bireylerin işlerini ve günlük faaliyetlerini yalnız başına sağlıklı bir şekilde yapmalarına engel oluşturmaktadır. Bu sebeple hastalığın ilk evrelerinde hastanın iş hayatından ve sosyal hayattan soyutlanması, ileri evrelerinde ise yaşamlarını başkalarına muhtaç olarak sürdürmek zorunda kalmaları hastaların psikolojilerinin bozulmasına sebep olmaktadır (Fahn ve Przedborski, 2000).

Belirtileri kişiden kişiye göre değişen ve belirtilerin ilerlemesi farklılık gösteren Parkinson hastalığında ortaya çıkan ilk belirtilerden biri el hareketlerinde yavaşlama ve yürürken kolların savrulmasındaki azalmadır. Bu belirtilere omuz ağrısı da eklenebilir. Çoğu kişi en çok dinlenme halinde ve başlangıçta hafif düzeyde olmak üzere titremeler yaşar. Genellikle elde oluşan bu titremeler kollar ve bacaklarda da görülebilir. Bununla birlikte, Parkinson hastalarının bir kısmı hastalıkları boyunca hiçbir zaman titreme yaşamayabilirler.

Belirtiler genellikle vücudun bir tarafında oluşur. Vücutta baskın taraf etkilendiğinde yazı yazma gibi alışkanlıklar gerçekleştirilirken semptomlar fark edilir. Erken teşhis ile tedaviden yararlanma oranı daha yüksektir. Parkinson hastalığının ilk aşamalarında hastalar denge problemi yaşayabilirler. Ayakta dururken, arkaya dönerken veya ani hareketler yaparken zorlanabilirler. Parkinson hastaları genellikle daha yavaş konuşurken mimiklerinde azalma meydana gelir. Depresyon ve uyku bozuklukları motor belirtilerden önce görülebilir. Parkinson hastalığının teşhisi için özel bir test yoktur. Diğer hastalık olasılıklarının elenmesi veya hastanın Parkinson ilaçlarına yanıt vermesi durumunda teşhis konulur.

Parkinson hastalığının teşhisi özellikle hastalığın ilk evrelerinde zordur. Bu sebeple son yıllarda tıbbi karar verebilmeyi desteklemek amaçlı bilgisayar tabanlı çözüm araştırmaları oldukça artmıştır. Bu çalışma ile bu alanda çalışan uzmanların Parkinson teşhisi sürecinde karar vermelerine yardımcı olmak ve konulan teşhislerde doğruluk oranını arttırmak hedeflenmektedir. Türkiye’de nöroloji ve bilgisayar bilimleri literatürüne katkı sağlamak amaçlanmaktadır.

Yapılan çalışmalar incelendiğinde spiral çizme ile Parkinson teşhisinde genellikle elde edilen çizimler olması gereken çizim ile karşılaştırılıp farkların bulunması yönteminin kullanılmış olduğu görülmektedir. Bu çalışmada ise diğer çalışmalardan farklı olarak spiraller birer karakter olarak düşünülmüş ve teşhis edilme problemi karakter tanıma problemi olarak ele alınmıştır. Böylelikle makine öğrenme algoritmaları Parkinson hastalığı teşhisinde kullanılmış ve başarılı sonuçlar elde edilmiştir. Çalışmada k – En Yakın Komşu ve Karar Ağaçları metotları makine öğrenme yöntemleri olarak kullanılmış olup, Birini Dışarıda Bırak Çapraz Doğrulama yöntemi ile sonuçlar doğrulanmıştır.

2. ÖNCEKİ ÇALIŞMALAR

Parkinson hastalığının belirtileri hastaları fiziksel olarak etkilediği için teşhisin bilgisayar ile yapılabilirliği birçok çalışma için araştırma konusu olmuştur. Teknolojik gelişmelerin hızla artmasıyla günümüzde Parkinson hastalığını uzaktan teşhis ve takibi için bilimsel çalışmalar yapılmaktadır. Bu çalışmalar hastalığın etkilediği fonksiyonlara yönelik olarak özellikle çizim, el yazısı, yürüyüş ve konuşma üzerinde yoğunlaşmıştır.

Parkinson hastalığının belirtileri hastalarının yaşamlarını olumsuz yönde etkilemeye kadar başlangıçta anlaşılmayabilir. Bu belirtiler hastalar tarafından ya fark edilmez ya da göz ardı edilir (Waters, 2000). Belirsiz semptomlardan sonra istirahat tremoru, akinezi/bradikinezi, rijidite gibi semptomlar meydana gelmeye başlar (Ertan, 2005). Hastalarda çabuk yorulma ve kişilikteki değişimler gözlemlendikten sonra motor belirtiler ortaya çıkar. Hastalar çoğu zaman omuz ve kol ağrılarından şikâyet ederler. Stres altında olduklarında el parmaklarında ve ayaklarında istirahat tremorunda artmalar meydana gelir. Ayrıca kolları az sallayarak yürümek ve ellerde meydana gelen titreme semptomu olan tremorun oluşması hastalık teşhisi için dikkat edilmesi gereken saptamalardır (Delil, 2001).

Genellikle istirahat tremoru ile Parkinson hastalığının semptomları başlar. Tremordaki bu düzensizlik hastalık boyunca devam eder. Hastaların her birinde değişik oranlarda bradikinezi görülür. Hastalar dikkatle takip edildiğinde günlük işlerde zorluk yaşama, küçük ve yavaş yazma, yavaş hareket etme ve alet ve donanım kullanımında zorluklar yaşama gibi belirtileri uzun süredir yaşadığı gözlemlenebilir. Hastaların şikâyetlerinde refleks mekanizmalarının etkilenmesi sonucu oturduktan sonra zor kalkma, dengeyi sağlamada zorluk ve yatakta dönme gibi yakınmalar da yer alabilir. Yüz ifadelerindeki donukluk ve kamburlaşma da hastalığın semptomlarından biri olarak görülebilir. Ses değişimleri, ses şiddetindeki kayıplar ve sesin yapısının bozulması (disfoni) gibi belirtiler de mevcuttur (Little ve ark., 2009).

Little ve ark. (2009) Parkinson hastalığıyla meydana gelen ses yapısındaki bozulmaları ölçerek hastalığın durumunu derecelendirmeyi hedefledikleri çalışmada 23'ü Parkinson hastası olmak üzere 31 denek kullanmışlardır. Bu kişilere mikrofon aracılığıyla konuşma testleri uygulayıp, sesteki titreşim, ton ve ses düzeyi gibi kişiye özgü özellikleri çıkartmışlardır. Bunlardan hangilerinin Parkinson hastalığının durumunu ifade etmede kullanılabileceğini belirlemişlerdir. Belirlen özellikleri kullanarak kişinin hastalığının hangi aşamada olduğunu bulmayı hedeflemişler ve %94 gibi başarılı bir sonuca ulaşmışlardır.

Tsanas ve ark. (2009) Parkinson hastalığının durumundaki olumlu ya da olumsuz değişim oranını tahmin etmeyi hedefledikleri çalışmalarında, 42 Parkinson hastasından veri toplamışlardır. Mikrofon yardımıyla topladıkları yaklaşık 6000 ses kaydında konuşmadaki ses özelliklerini sinyal işleme algoritmaları ile karakterize etmişler ve kullanılabılır özellikleri belirlemişlerdir.

Revelt ve ark. (2009), Little ve ark.'nın (2009) çıkardıkları ses örneklerinden elde ettikleri çeşitli frekansları kullanarak kurallar çıkarmışlardır. Ortaya çıkardıkları bu kurallar, filtreleme algoritmalarıyla elenmiş ve seçilen kuralların başarı oranları sınıflandırma algoritması ile hesaplanmıştır ve %90 başarı oranını sağlamışlardır. Bu çalışma aslında yöntemi denemek için yapılmıştır. Elde edilen bu sonuçlar çalışmanın bir sonraki aşamasının temelini oluşturmuştur.

Tsanas ve ark. (2010), yaptıkları diğer çalışmada ses sinyallerini logaritmik olarak dönüştürmüşlerdir. Bütün testler hastanın evinde kendi yönetiminde gerçekleştirilmiş ve elde ettikleri sonuçları hastalığın teşhisinde doktorların sıklıkla kullandığı Birleşik Parkinson Hastalığı Değerlendirme Ölçeğiyle eşleştirmişlerdir. Ham ses sinyallerine kıyasla logaritmik olarak dönüştürülmüş ses sinyallerinin çok iyi klinik bilgiler sağladığı görülmüştür.

Tablet bilgisayar yardımıyla hastalar üzerinden alınan dijital yazım örnekleri son zamanlardaki çalışmaların temelini oluşturmaktadır. Gemmert ve ark. (2003) 54 ile 82 yaş arasında olan 13 adet sağlıklı insanı ve Parkinson hastasını kullanarak yaptıkları

çalışmada, her hastadan, 1,0, 1,5, 2,0, 3,5 cm uzunluklarında 'lllllll' şeklinde 12 tane şablon çizmelerini istemişlerdir. Yaptıkları ölçümler sonucunda Parkinson hastalarının 1,5 cm üzerindeki şekilleri çizmede sağlıklı insanlara göre zorlandıkları ve başarısız olduklarını gözlemlemişlerdir. Parkinson hastalarının yazma hızlarının sağlıklı insanlara göre daha düşük olduğunu da tespit etmişlerdir. Sonuç olarak, Parkinson hastalığının ayırt edilmesinde çizilen karakter boyutlarının önemli olduğunu ve bu farkın hasta ve sağlıklı insanların ayırt edilmesinde kullanılabileceğini belirtmişlerdir.

Ünlü ve ark. (2006) yazım aşamasında oluşan basıncı ve el hareketlerini bir elektronik kalem kullanarak kaydettikleri çalışmalarında 28 Parkinson hastası ve 28 sağlıklı birey kullanmışlardır. Yazım sırasında üç koordinat ekseninde meydana gelen basıncı ve hareketi kaydederek farklı çizimlerde görülen basınç farkına dayanan bir özellik çıkarmışlardır. Bu özellik yardımıyla Parkinson hastalarında görülen titremeler ile normal titreme belirtileri arasındaki farkı ortaya çıkarmaya çalışmışlardır.

Aly ve ark. (2007) geleneksel yöntemle yapılan çizim ortamını koruyarak tablet bilgisayar yardımıyla bir çalışma yapmışlar. Bu çalışmada elektronik kalem yardımıyla çizimlerinin tamamlanma süresini, koordinatlarını, dijital ekran üzerindeki basıncı ve kayma miktarını ölçmüşlerdir. Elde ettikleri verileri Fourier dönüşümü ile inceleyip Birleşik Parkinson Hastalığı Değerlendirme Ölçeği'ne göre sonuçları değerlendirmişlerdir. Parkinson hastası ve aynı yaştaki sağlıklı insanları kıyasladıkları değerlendirmede kullanılan sistemin geleneksel muayenenin yerini alamayacağını ama tanı aşamasında yardımcı olabileceğini belirtmişlerdir.

Wang ve ark. (2008) yaptıkları çalışma ile hastaların hareket reflekslerinin uyumunu ve performansını spiral analiz yöntemi ile tablet bilgisayar yardımıyla ölçmüşlerdir. Bu reflekslerin fonksiyonunu yitirip yitirmediğini, ne kadar kayıp yaşadığını ya da ne kadar çalıştığını, hastalığın teşhisinde kullanılıp kullanılamayacağını, hastalık hakkında bilgi elde etme olanağını öğrenmeyi hedeflemişlerdir. Tablet bilgisayar üzerine bir kâğıt yardımıyla Arşimet spiralleri çizdirerek elde ettikleri bilgiler doğrultusunda Parkinson hastaları ile sağlıklı bireyleri birbirinden ayırarak sınıflandırmayı başarmışlardır.

Sakar ve ark. (2013) yaptıkları çalışmada Parkinson hastalarından bir dizi ses örnekleri toplamışlardır. Her konu için çoklu konuşma kayıtları alarak sesli harfleri seslendirmenin Parkinson hastalığı teşhisinde ne kadar belirleyici olduğunu ve merkezi eğilim ve dağılım ölçütlerinin tüm örnek kayıtlarının temsilcisi olarak ne derece işe yaradığını araştırmışlardır. Makine öğrenme algoritmaları kullanarak yaptıkları araştırmada uzun süreli ünlülerin Parkinson hastalığı için ayırt edici bilgi taşıdığını tespit etmişlerdir. Ayrıca merkezi eğilim ve dağılım ölçümleri olan bir konunun örneklerini temsil etmenin öngörücü modelin genelleştirilmesini iyileştirdiğini bulmuşlardır.

Luciano ve ark. (2016) 138 Parkinson hastası ve 150 kontrol kişilerine tablet üzerine spiraller çizdirerek x, y, z koordinat verileri, basınç ve zaman bilgilerini toplamıştır. İkinci dereceden düzgünlük, birinci dereceden sıfır geçiş, şekil ve kinematik düzensizlikler, sıklık, ortalama hız ve spiral genişlik gibi endeksleri araştırmıştır. Sıfır geçişi dışında tüm endekslerin Parkinson hastaları ve kontroller arasında anlamlı olarak farklı olduğunu ve tüm endeksleri kullanan bir modelin, yüksek düzeyde ayırt edici geçerliliğe sahip olduğunu tespit etmişlerdir.

Pereira ve ark., (2018) yaptıkları çalışmada, el yazısı dinamikleri tarafından üretilen görüntülerin özelliklerini öğrenmek için evrimsel sinir ağlarını tanıtmışlar. Ek olarak, bilgisayar destekli Parkinson Hastalığının teşhisi ile ilgili araştırmaları teşvik etmek için görüntülerden ve sinyal tabanlı verilerden oluşan bir veri setini kullanıma sunmuşlardır. Erken evre tespiti bağlamında başarı oranı %95 olmuştur. El yazısı dinamiklerini derin öğrenme teknikleri kullanarak analiz etmek, otomatik Parkinson hastalığını tanımlamak ve klasik teşhis yöntemlerine göre daha iyi performans elde etmesi bakımından faydalı olduğunu göstermişlerdir.

Braga ve ark. (2019) yaptıkları çalışmada kontrolsüz arka plan koşullarında serbest konuşma yoluyla Parkinson hastalığının erken belirtilerini tespit etmek için bir yöntem önermişlerdir. Makine öğrenme algoritmalarıyla entegre sinyal ve konuşma işleme tekniklerini kullanarak erken algılama mekanizması oluşturmuşlardır. Eğitim ve değerlendirme aşamalarında parametrelerin tahmininde, Parkinson hastalarının

farklı aşamalarındaki kayıtlarını içeren üç ayrı konuşma veri tabanı kullanmışlardır. Sonuçlar, Random Forest (RF) veya Support Vektör Machine (SVM) tekniklerinin kullanma potansiyelini ortaya koymuştur. Önerilen algoritma ile Parkinson Hastalığının varlığını çok yüksek bir doğrulukla ve güvenilir bir hesaplama yöntemi ile tahmin etmişlerdir.

Agarwal ve ark. (2016) konuşma bozuklukları analizinin Parkinson hastalığının erken teşhisinde önemli bir araç olduğunu düşünerek bir çalışma yapmışlardır. Doğru konuşma örneklerini kullanarak Parkinson hastalığını tahmin etmek için Extreme Learning Machine kullanarak etkili bir yaklaşım önermişlerdir. Yöntemin performansı, UCI deposundan alınan güvenilir bir veri kümesi ile değerlendirilmiştir. Önerilen yöntem, Parkinson hastalığı olan hastaları ve sağlıklı bireyleri, eğitim veri seti için % 90.76 ve 0.81 MKK doğrulukta ayırt etmiştir. Parkinson hastalığı olan hastalardan oluşan bağımsız bir veri seti ile test edildiğinde ise önerilen yöntem % 81,55 doğruluk oranı vermiştir. Metodun performansı Yapay Sinir Ağı ve Destek Vektör Makinesi gibi mevcut tekniklerle karşılaştırılmış ve elde edilen sonuçlar, önerilen yöntemin Parkinson hastalığını tanımlamak için güvenilir olduğunu göstermiştir.

Danna ve ark. (2018) yaptıkları çalışmada, Parkinson hastası bireylerde ve medikal tedavisi olmayan sağlıklı bireylerde, kalem hareketini ve dijitalleştirilmiş spiral çizim verilerini değerlendirmeyi amaçlamıştır. Veri odaklı üç aşamalı yaptıkları analizlerde ilk olarak, mekânsal ve zamansal kısıtlamaların çeşitli değişkenler üzerindeki etkilerini belirlemişlerdir. İkinci olarak, Parkinson hastalığı semptomlarının baskınlığını ve baskınlığı arasındaki ilişkiyi en uygun değişkenler için incelemişlerdir. Son olarak, Parkinson hastalığı ile ilişkili değişikliklerin oluşumunu değerlendirmek için üçüncü bir analiz yapmışlardır. İlk analiz, hız tepe noktalarının ve kalem irtifa değişikliklerinin, hastalık ve ilacın etkisini değerlendirmek için spiral çizimde en alakalı değişkenler olduğunu ortaya koymuştur. İkinci analiz, ilacın etkisinin sadece hareket akıcılığı için mevcut olduğunu ortaya koymuştur. Son olarak, üçüncü analiz, ilacın etkisinin, çizimin başlangıcında olduğundan daha fazla olduğunu göstermiştir. Dijitalleştirilmiş spiral çizim, hastalık sırasında kinematik değişikliklerin

görev sırasında gerçekleştiği zaman kesin olarak gözlemlenmesini sağlamıştır. Bu çalışmanın Parkinson hastalığının tanısına ve izlenmesine katkıda bulunmak için büyük önem taşıyacağı sonucuna ulaşmışlardır.

Sharma ve ark. (2019) yaptıkları çalışmada, özellik seçimi için bir arama stratejisi görevi gören geleneksel Gri Kurt Optimizasyonuna (Grey Wolf Optimize - GWO) dayanan Modifiyeli Gri Kurt Optimizasyonu (Modified Grey Wolf Optimization - MGWO) adlı yeni bir model önermişlerdir. GWO, kurtların avlanma davranışlarıyla büyülenen bir meta-sezgisel algoritmadır. Rastgele orman, k en yakın komşu sınıflandırıcı ve karar ağacı sınıflandırma yöntemlerini kullanmışlardır. Önerdikleri model çeşitli ses veri setleri, el yazısı ve konuşma kullanılarak değerlendirilmiştir. Öne sürülen algoritma, Parkinson hastalığının tahmininde % 94,83'lük bir tahmin doğruluğu, % 98,28'lik tespit oranı, % 16,03'lük sahte alarm oranı ve daha erken bir aşamada fonksiyonel bir tedavi görmeleri için bireylere yardımcı olmaktadır. Veri setlerinde önerilen algoritmanın değerlendirilmesinden elde edilen sonuçlar, Optimize Mürekeplebalığı Algoritmasının (Optimized Cuttlefish Algorithm - OCFA) sonuçlarıyla karşılaştırılmıştır. Deneysel sonuçlar önerilen algoritmanın keskinliği en üst düzeye çıkarmaya ve seçilen özellik sayısını en aza indirmeye yardımcı olduğunu göstermiştir.

3. MATERYAL VE METOT

3.1. Veri Seti

Veri seti, California Irvine Üniversitesi Makine Öğrenimi Deposundan (University Of California Irvine Machine Learning Repository) indirilmiş olup, İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı'na başvuran 62 Parkinson hastası ile 15 sağlıklı bireyin çizim verilerinden oluşmaktadır. Tüm deneklerden üç tip el yazısı kaydı (Statik Spiral Test (SST), Dinamik Spiral Test (DST) ve Belirli Noktadaki Stabilitate Testi) (STCP) alınmıştır. (Isenkul ve ark., 2014)

El yazısı veri seti Wacom Cintiq 12WX grafik (Hahne ve ark., 2009) tablosu kullanılarak oluşturulmuştur. Temelde bir grafik tablet ve içinde bulunan bir LCD monitörden yararlanılmıştır. El yazısı çizimlerini kaydetmek ve kayıtları kullanarak Parkinson hastalarının koordinasyonunu test etmek için özel bir yazılım tasarlanmıştır. Yazılım, cihazın API işlevlerini kullanır ve Windows sistemlerinde çalışabilen C# platformunda geliştirilmiştir.

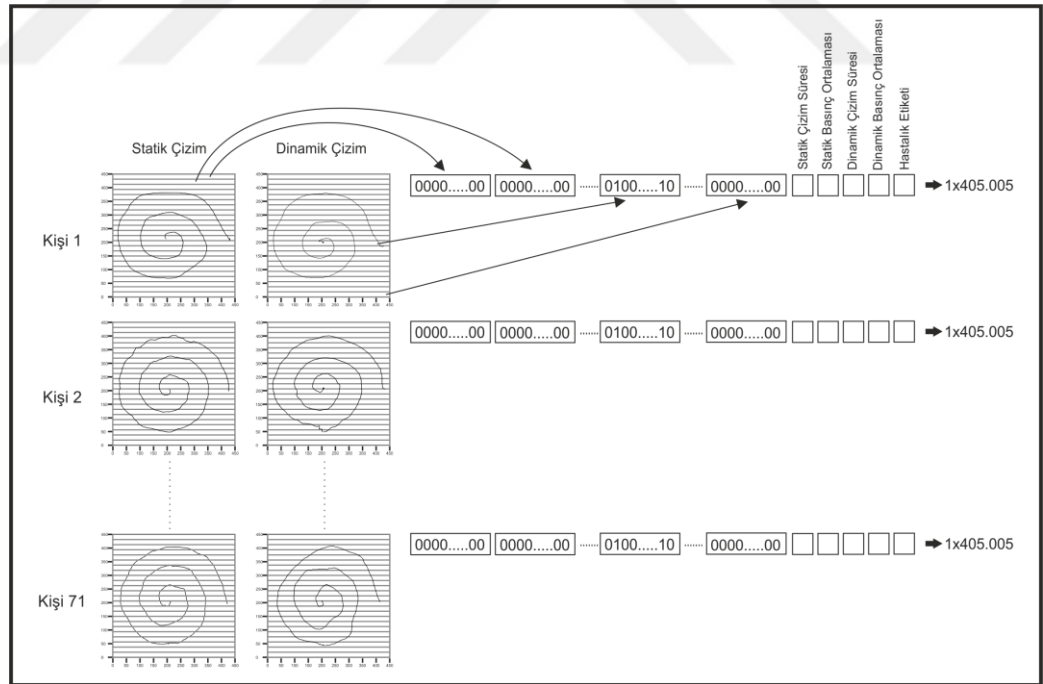
Isenkul ve ark. (2014) yaptıkları çalışmada, grafik tablet aracılığıyla veri toplanması için geliştirilen üç farklı test türü vardır. Bunlardan ilki, literatürde motor performansın belirlenmesi (Wang ve ark., 2008), titreme ölçümü (Pullman, 1998) ve Parkinson hastalığı teşhisi (Saunders ve ark., 2008) gibi farklı amaçlar için klinik araştırmalarda sıklıkla kullanılan Statik Spiral Test (SST) 'tir. Bu testte, yazılımı kullanarak grafik tablet üzerinde Arşimet spiralleri belirir ve hastalardan, dijital kalemi kullanabildikleri kadar aynı spirali yeniden çizmeleri istenir. Test sırasında yukarıda belirtilen özellikler ve hastayı belirtmek için diğer veriler veri tabanına kaydedilir.

İkinci test Dinamik Spiral Test (DST) 'tir. SST'nin aksine, Arşimet spirali belli zaman aralıklarında belirir ve kaybolur, bir başka deyişle Arşimet spirali yanıp söner. Bu, hastayı kalıbı aklında tutmaya ve çizmeye devam etmeye zorlar. Bu testin amacı, hastanın çizim performansındaki değişimin belirlenmesi ve duraklama süreleridir,

çünkü bu durumda Arşimet spiralini çizmek daha zordur. Bu test sonucunda hastaların çoğunun çizmeye devam ettiği, ancak neredeyse hepsinin deseni kaybettiği görülmüştür.

Üçüncü test, Belirli Noktadaki Stabilitate Testidir (STCP). Bu testte, ekranın ortasında belirli bir kırmızı nokta var ve deneklerden ekrana belli bir süre dokunmadan dijital kalemle noktadan tutmaları isteniyor. Bu testin amacı hastanın el stabilitesini veya el titremesini belirlemektir.

Bu tez çalışmasında ise, Isenkul ve ark. (2014) tarafından yukarıda elde edilme yöntemleri anlatılan 62 Parkinson hastasından ve 15 kontrol grubu sağlıklı insanlardan alınan spiral çizim ve dinamik spiral çizim verileri kullanılarak yeni bir veri seti oluşturulmuştur. Çizim verilerinde 6 Parkinson hastasının statik ya da dinamik çizim verilerinin olmaması nedeniyle bu veriler veri setinden çıkarılmıştır. 71 kişiye ait verilerle çalışma yapılmıştır.



Şekil 3.1. Veri Setinin Oluşturulması

Şekil 3.1.'de görüldüğü gibi ham veriler, her bir kişi için spiral çizim ve dinamik spiral çizim verileri, çizimlerin tamamlanma süreleri, çizimlerin basınç ortalamaları ve

hastalık etiketinden oluşacak şekilde yeni bir veri setine dönüştürülmüştür. Bu da her bir kişi için yaklaşık 1x405.005’lik matrise, toplamda 71x405.005’lik matrise tekabül etmektedir.

Çizelge 3.1. Veri Setinin Tanımlanması

Öznitelik Türü	Öznitelik Sayısı
Statik Çizim Verisi	202.500
Dinamik Çizim Verisi	202.500
Statik Çizim Süresi	1
Statik Çizim Basınç Ortalaması	1
Dinamik Çizim Süresi	1
Dinamik Çizim Basınç Ortalaması	1
Hastalık Etiketi	1
Toplam	405.005

Isenkul ve ark. (2014) tarafından hazırlanan veri setinde çizimin yapıldığı noktaların X ve Y koordinatları, zaman verisi ve basınç değerleri bulunmaktadır. Bu değerler yardımıyla 450 x 450’lik bir matris oluşturularak kalemin geçtiği koordinatlara “1”, diğer noktalara “0” verisi eklenerek ikilik bir görüntü dizi elde edilmiştir. Aynı prosedür dinamik çizim verisi için de gerçekleştirilmiştir. Bu dizinin sonuna çizimlerin tamamlanma süresi ve çizim boyunca yapılan basınç ortalaması verileri eklenmiştir. Veri setinin son sütununa da hastalık etiketi eklenmiştir. Parkinson hastası olanlara “1”, sağlıklı insanlara “0” etiketi eklenmiştir. Böylece bir kişi için 1x405.005’lik bir matris, toplamda 71x405.005’lik bir matris verisi elde edilmiştir.

3.2. Makine Öğrenme Algoritmaları

Makine öğrenmesi mevcut veriler yardımıyla elde edilen tecrübelerin bilgisayar yazılımları ile gelecekteki olayları tahmin etmede ve modelle yapılmasına olanak sağlayan yapay zekâ alanıdır. Temel dayanağı giriş bilgilerini alan algoritmalar yardımıyla çıktılar oluşturmaktır ve bu çıktıları yeni veriler ile güncelleyip yeni bir

çıktıyı tahmin etmek için istatistiksel analizler kullanılmaktadır. Makine öğrenmesinin gelişimi hızla artarken bu yönde çok sayıda algoritmalar geliştirilmektedir.

Denetimli ve denetimsiz algoritmalar olarak makine öğrenme algoritmaları kategorize edilir. Algoritmalar, eğitim sırasında yapılan tahminlerin doğruluğunu teyit etmenin yanında, hem girdileri hem de çıktıları sağlamak için bir veri bilimcisi veya veri analisti gerektirir. Modelde hangi değişkenlerin kullanılacağını, modelin hangi özellikleri analiz edeceğini veri bilimci belirler. Eğitim sonunda algoritmanın öğrendikleri yeni verilere uygulanır.

Denetimsiz algoritmalarda sonuç verilerinin eğitim işlemine gerek yoktur. Denetleme işlemi yerine verileri gözden geçirmek ve sonuç elde etmek için yinelemeli bir yaklaşım olan derin öğrenmeyi kullanırlar. Sinir ağları olarak da adlandırılan Denetimsiz öğrenme algoritmaları daha karmaşık işlemler için kullanırlar. Milyonlarca eğitim verisini tarayan bu sinir ağları değişkenler arasındaki ilişkileri otomatik olarak tespit ederek çalışırlar. Eğitim sonucunda, yeni verileri yorumlamak için algoritma veri bankasını kullanır. Büyük miktarlarda eğitim verisi gerektiren işlemlerde kullanılan bu algoritmalar, büyük veri çağı ile birlikte uygulanabilir olmuştur.

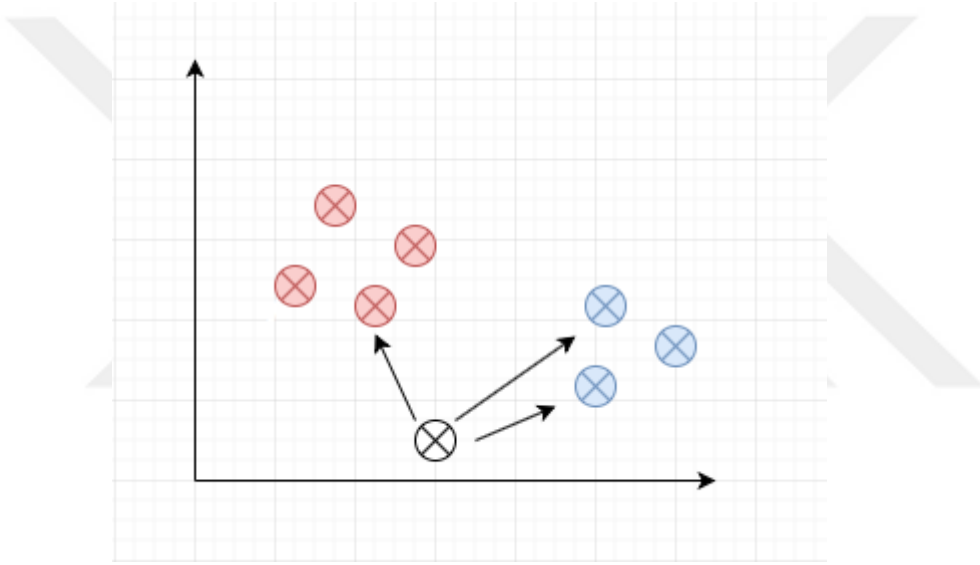
3.2.1. k - en yakın komşu (k-EK) algoritması

k - En Yakın Komşu (k-EK) yöntemi uygulaması kolay ama etkili sınıflandırma yöntemlerinden birisidir. Sınıfı önceden belirlenmemiş özniteliği, Öklid mesafesi ile kendisine en yakın komşusunun sınıfına bakarak sınıflandırma işlemini gerçekleştirir. k-EK yöntemi ayırık değerlere karşı çok duyarlı olduğundan aynı algorithmadan türetilmiş benzer bir yöntemden faydalanır. (Gümüşçü ve ark., 2019)

Algoritma, sınıfları belli olan bir örnek kümesindeki verilerden yararlanılarak kullanılmaktadır. Örnek veri setine katılacak olan yeni verinin, mevcut verilere göre uzaklığı hesaplanıp, k sayıda yakın komşuluğuna bakılır. Uzaklık hesapları için genelde 3 tip uzaklık fonksiyonu kullanılmaktadır:

- “Euclidean” Uzaklık
- “Manhattan” Uzaklık
- “Minkowski” Uzaklığıdır.

k-EK; eski, basit ve gürültülü eğitim verilerine karşı dirençli olması sebebiyle en popüler makine öğrenme algoritmalarından biridir. Fakat bunun yanında dezavantajı da mevcuttur. Örneğin, uzaklık hesabı yaparken bütün durumları sakladığından, büyük veriler için kullanıldığında çok sayıda bellek alanına gereksinim duymaktadır.



Şekil 3.2. k-EK Algoritması Örnek Çizim

k-EK yönteminde en yakın k komşu belirlenirken seçilen eğitim kümesindeki noktaların mutlak uzaklığına bakılır. Bu uzaklığın en az olduğu k komşu belirlenir, diğer komşuların sınıflarına bakılır ve ağırlığı en fazla olan sınıf seçilir. Mesafe aynı olursa k noktadan rastgele birisi seçilir. Kazanan sınıfın oranı belirlenen eşik değerinden düşük olduğunda tahmin mümkün olmayacağından, bazen direk sınıfı belirlemek yerine sınıflandırma işleminde şüpheli seçeneği seçmek de faydalı olabilir.

k-EK algoritması, model tanımada sınıflandırma ve regresyon işlemlerinde kullanılan parametrik olmayan bir yöntemdir. Girdi her iki durumda da özellik

alanında k en yakın eğitim örneklerinden oluşur. k -EK algoritmasının sınıflandırma işleminde kullanılıp kullanılmayacağı çıktıyı belirler.

k -EK sınıflandırmasında sınıf üyeliği çıktıdır. Sınıflandırma işleminde nesne en yakın komşuları arasından en yaygın olan sınıfa verilir. $k=1$ olduğunda ise nesne en yakın komşusunun sınıfı olarak belirlenir.

k -EK regresyonunda çıktı, en yakın komşularının değerlerinin ortalaması olan cismin özellik değeridir.

Yakın komşuların ortalama ya da uzak komşulardan daha fazla katkıda bulunması için hem sınıflandırma hem de regresyon işleminde komşuların katkılarına ağırlık koyulabilir. Komşuya olan uzaklığı d olarak belirtirsek, her komşuya $1/d$ ağırlık oranı verilerek bu işlem gerçekleştirilebilir.

k -EK algoritmasında verilerin yerel yapısına duyarlı olması bir diğer özelliğidir. Algoritma yaygın olarak kullanılan k -means makine öğrenme yöntemi ile karıştırılmamalıdır. k -EK mevcut tüm durumları depolayan ve benzerlik ölçüsü dâhilinde yeni olayları sınıflandırabilen basit bir algoritmadır. İstatistiksel tahmin ve örüntü tanıma işlemlerinde parametrik olmayan bir teknik olarak 1970 yıllarında kullanılmıştır.

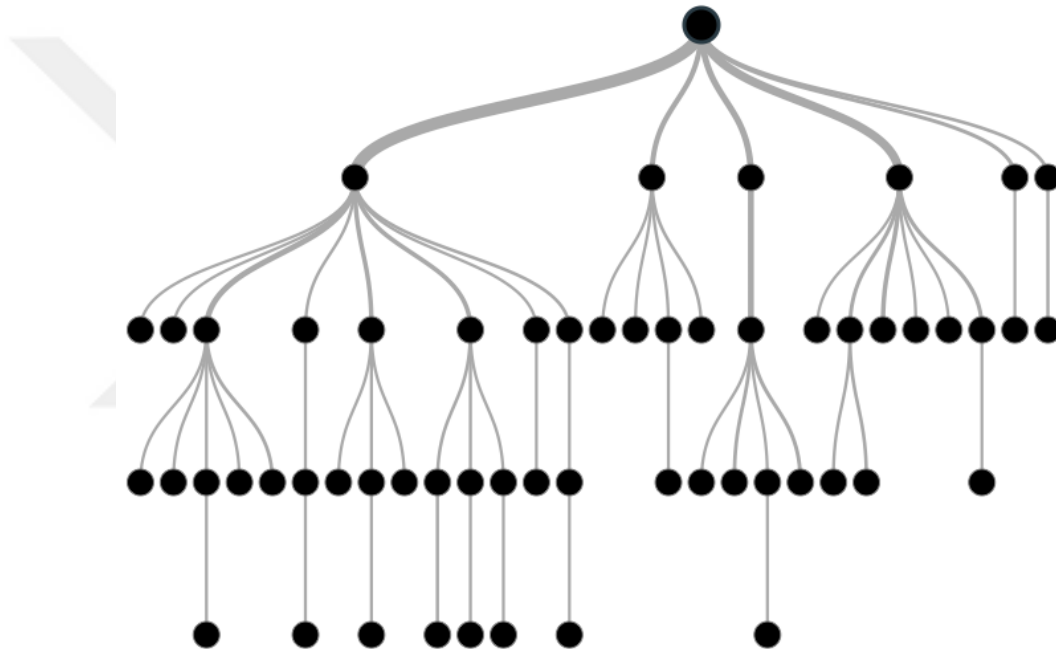
3.2.2. Karar ağacı sınıflandırma metodu

Denetimli Makine Öğrenmesi türü olarak belirli bir parametreye göre verilerin sürekli olarak bölünmesi yöntemine Karar Ağaçları sınıflandırma yöntemi denir. Karar düğümleri ve yapraklar olmak üzere iki bölümden oluşur. Karar düğümleri verilerin nereye bölündüğünü gösterirken yapraklar ise nihai sonuçtur yani karardır.

Sınıflandırma problemleri için yaygın olarak kullanılan Karar Ağaçları sınıflandırma yöntemi sınıflandırma doğruluğu bakımından diğer öğrenme

yöntemlerine göre çok etkindir. Sınıflandırma modeli ağaç şeklinde gösterilir ve Karar Ağacı (Decision Tree) olarak isimlendirilir.

Akış şemalarına benzeyen yapısı ile Karar Ağaçlarında her bir nitelik bir düğüm tarafından temsil edilir (Şekil 3.3.). Ağaç yapısının elemanları; dallar ve yapraklardır. En üst yapı kök, en son yapı yaprak olarak adlandırılırken bunlar arasında kalan yapılar dal olarak adlandırılır. En tepeden en aşağı inen bir yapıları vardır ve önceden belirlenmiş bir hedef değişkene sahiptirler.



Şekil 3.3. Karar Ağacı Sınıflandırma Yöntemi

Çok sayıda veri içeren bir veri kümesini daha küçük kümelere bölmek için karar ağaçları yapısı kullanılır. Büyük miktardaki verileri basit karar verme adımları uygulayarak daha küçük veri gruplarına bölerek kullanırlar.

Karar Ağaçlarının Avantajları:

- Kolay anlaşılır ve yorumlanır. Ağaç yapılar görsel bir şekilde ifade edilebilir.
- Bu model kayıp değerleri desteklemezken, veri hazırlığı işlemi azdır.
- Ağaç eğitiminde kullanılan veri noktalarının sayısı ile kullanılan ağacın maliyeti orantılıdır.
- Hem sayısal hem de kategorik veriler işlenebilir.
- Birden fazla sayıda çıktıya sahip problemlerde kullanılabilir.
- Bir modelin doğrulanması işlemi istatistiksel testler kullanılarak yapılabilir.
- Parametrik olmayan bir yöntem olan karar ağaçları, uzay dağılımı ve sınıflandırma yapısı hakkında bir yaklaşımı yoktur.

Karar Ağaçlarının Dezavantajları:

- Aşırı karmaşık ağaçların elde edilmesiyle veri doğru bir şekilde açıklanmayabilir. Ağaç dallanmasının takibi zorlaşabilir.
- Ezbere öğrenme problemi yaşanabilir. Parametrelerin azaltılması ve budama gibi işlemlerle bu problem çözülebilir. Az sayıda nesne barındıran yaprak düğümlerin grafikten atılması ile karar ağacında budama işlemi gerçekleştirilebilir.

Bölünmenin nasıl olacağı karar ağacı algoritmalarında doğruluğu etkileyen önemli faktörlerden biridir. Bölünme şartları sınıflandırma ve regresyon problemlerinde genellikle birbirinden farklıdır.

Bir düğümü daha fazla alt düğüme bölmeye karar vermek için karar ağaçları yöntemi, birden fazla algoritma kullanabilir. Alt düğümlerin oluşturulması, alt düğümlerin homojenliğini sağlar. Hedef değişkenlere göre düğümün saflığının arttığı ifade edilebilir.

Hedef değişkenin tipi algoritma seçimini etkiler. Entropi, Gini, Sınıflandırma Hatası yöntemleri kategorik değişkenlerde kullanılan karar ağacı algoritmalarıdır. Sürekli değişkenler için En Küçük Kareler yöntemi kullanılır.

Bu çalışmada Quinlann (1996)'ın ID3 algoritmasını geliştirerek elde ettiği C4.5 karar ağacı oluşturma algoritması kullanılmıştır. C4.5 algoritmasında ID3 algoritmasına ek olarak bölünme bilgisi, özelliklerin kayıp değerleriyle baş edilmesi ve sayısal özellik değerlerinin hesaba katılması özelliklerini barındırmaktadır.

3.2.3. Genetik algoritma ile öznitelik seçme

Örnek sayısının az, öznitelik sayısının çok fazla olduğu veri setlerinde makine öğrenme algoritmaları çok başarılı sonuçlar vermemektedir. Daha başarılı sonuçlar elde etmek amacı ile öznitelikler seçilerek daha değerli öznitelikler makine öğrenme algoritmaları modelleri oluştururken kullanılmaktadır.

Öznitelik seçme yöntemleri genellikle filtre, sarmal ve gömülü şekilde olmak üzere üç ana başlıkta incelenir. Sarmal modellenli öznitelik seçme yöntemleri genellikle meta-sezgisel algoritmalar ile en ideal öznitelikleri seçmeyi amaçlamaktadır.

Başarı oranının yüksek olduğu Sarmal modellenli öznitelik seçme yöntemlerinde öznitelik sayısının fazla olduğu veri setlerinde işlem süresi çok fazla olmaktadır. Genetik algoritma da bu amaçla yapılan birçok çalışmada sıklıkla kullanılmaktadır. Belirlenen uygunluk fonksiyonun istenilen değere ulaştırılması için seçim, mutasyon ve çaprazlama gibi operatörler kullanılır.

1975 yılında John Holland canlılardaki evrimden ve değişimden esinlenerek genetik evrim sürecini bilgisayar ortamına aktarmış ve makine öğrenmesi üzerine çalışmalar yapmıştır. Yaptığı çalışmalarla tek mekanik yapının öğrenme yeteneğini geliştirmek yerine canlılardaki çiftleşme, çoğalma ve değişim gibi genetik süreçlerin sonunda daha üstün yeni bireylerin elde edilebileceğini göstermiştir. Yayınladığı sonuçların ardından yöntemin adını “Genetik Algoritmalar” olarak ifade etmiştir.

Genetik Algoritmalar, bir problemin çözümünü olası çok çözümünden en iyisini, en uygununu bulmaya çalışan algoritmalarlardır. Popülasyon geliştikçe kötü sonuçlar nesilden nesile kaybolurken, iyi sonuçlar ise daha iyi sonuçlar oluşturmak için tekrar tekrar kullanılırlar.

Genetik Algoritmalar, çözüm uzayının tamamını değil sadece bir kısmını tararlar. Tarama kısmının az olması etkin aramanın gerçekleşmesine ve çözüme daha kısa sürede ulaşılmasını sağlar. Yerelde en iyi çözümlere takılmayarak popülasyonu eş zamanlı incelemesi önemli bir üstünlüğüdür. Problemi etkileyen etmenlerin çok olduğu durumlarda kullanılan Genetik Algoritmaların çözüm alanı genellikle geniştir. Olasılık kurallarına göre çalışan genetik algoritmaların ne kadar iyi çalışacağı önceden tahmin edilemez.

Bu çalışmada ise 3.1.’deki formüle göre sınıflandırma hata oranı genetik algoritmanın uygunluk fonksiyonu olarak belirlenmiştir.

$$\text{Uygunluk Fonksiyonu} = \text{Sınıflandırma Hata Oranı (Öznitelikler)} \quad (3.1.)$$

Genetik algoritma ile öznitelik seçebilmek için öznitelik sayısı kadar bitlik bir dizi oluşturulur. Bu dizinin değerleri lojik 1 veya lojik 0 olabilir. Bir dizi elemanın değerinin lojik 1 olması ilgili elemana denk gelen özniteliğin seçilmesi, lojik 0 olması ise ilgili elemana denk gelen özniteliğin seçilmemesi anlamına gelir.

3.2.3.1. Genetik algoritmalar temel kavramlar

Gen; genetik bilgileri taşıyan ve kendi başına anlamı olan en küçük genetik birim olarak ifade edilir. Bu küçük yapıların bir araya gelmesiyle kromozomlar meydana gelir ve kromozomlar tüm bilgileri içerir.

Kromozomlar bir ya da birden fazla genin bir araya gelmesiyle oluşurlar. Kromozomlar toplumdaki bireylere karşılık gelirken probleme ait tüm bilgileri içerirler. İşlem yapılan problemlerde alternatif çözüm adaydırlar.

Popülasyon ise kromozom topluluğudur. Üzerinde durulan problem için alternatif çözümler kümesidir. Bir popülasyondaki birey sayısı probleme göre uygulayıcı tarafından saptanır. Zayıf olan bireylerin yerini güçlü bireyler alır. Bu yenilemeler sayesinde probleme daha uygun bireyler bulunabilmektedir.

3.2.3.2. Genetik algoritmanın uygulanması

İlk jenerasyonun oluşturulup uyum değerinin hesaplanması genetik algoritmanın ilk adımıdır. İkinci adım mevcut popülasyona seçim, çaprazlama ve mutasyon gibi temel genetik operatörlerin uygulanmasıdır. Her bir popülasyonun uyum değeri hesaplanır. Bu adımlar durdurma kriterine ulaşana kadar tekrarlanır. Şekil 3.4.'te gösterildiği gibi adımlar şu şekilde açıklanabilir:

- Olası bütün çözümlerin yer aldığı bir çözüm grubu oluşturulur. Bu grup biyolojideki benzerliği nedeniyle popülasyon, çözümlerin kodları da kromozom olarak adlandırılır.
- Popülasyondaki bireylerin sayısı belirlenerek ilk adıma başlanır. Birey sayısı için özel bir hesaplama ya da standart yoktur. Genellikle 100 – 300 arasında bir değer verilir. Yapılacak işlemlerin karmaşıklığı ve arama işleminin derinliği birey sayısının büyüklük seçimini etkiler. Bu işlemten sonra ilk popülasyon rastgele oluşturulur.

- Uygunluk fonksiyonu yardımıyla her bir kromozomun ne kadar iyi olduğu tespit edilir. Hesaplama (evaluation) adı verilen bu işlem ile kromozomların uygunluk değerleri hesaplanır. Genetik algoritmanın temelini oluşturan bu fonksiyonun verimli ve hassas olması Genetik algoritmanın başarısını etkiler.
- Kromozomlar eşlenerek yeniden kopyalanır, değiştirme operatörleri uygulanır ve böylelikle yeni bir popülasyon elde edilir. Kromozomların uygunluk değerine göre kromozomların eşlenme işlemi gerçekleştirilir. Rulet tekerleği seçimi, turnuva seçimi gibi yöntemler kullanarak bu seçim yapılır.
- Eski kromozomlar silinerek yeni kromozomlara yer açılır. Böylelikle popülasyon büyüklüğünün sabit kalması sağlanır.
- Tüm kromozomların n uygunluk değerleri tekrar hesaplanarak yeni popülasyonun başarısı tespit edilir.
- Genetik algoritma tekrar tekrar çalıştırılarak çok sayıda popülasyon elde edilir. Durdurma kriterine ulaşıldığında o ana kadar bulunan en iyi kromozom sonuçtur. Çünkü en iyi bireyler popülasyon hesaplamalarında saklanmıştır.
- Bu çalışmada Genetik Algoritmanın kodlanmasında popülasyon sayısı 30, jenerasyon sayısı 100, mutasyon oranı 0.3, çaprazlama oranı 0.8 olarak belirtilmiştir. Bu parametrelerin farklı değerleri ile genetik algoritma çalıştırılmış ama sonuçları etkilemediği gözlemlenmiştir.



Şekil 3.4. Genetik Algoritma Genel Akış Diyagramı

3.2.3.3. Seçim (Çoğalma)

Dizilerin uyum fonksiyonunun değerine göre kopyalanma sürecine seçim işlemi denir. Dizilerin uyum değerinin yüksek olması bir sonraki kuşağa katkılarının daha fazla olacağı anlamına gelir. Literatürde çoğalma operatörü olarak da adlandırılmaktadır.

Seçim işlemi bir sonraki jenerasyonda yavru üretmede hangi ailelerin olması gerektiğine karar verir. Ortalama uyum değerinin üzerindeki dizilere çoğalma fırsatı tanınır. Rulet tekerleği (roulette wheel), turnuva (tournament), sabit durum (steady state) ve sıralama (permutation) yöntemleri başlıca çoğalma yöntemleri arasında yer alır.

3.2.3.4. Çaprazlama

Çaprazlama operatörü, mevcut gen havuzunun potansiyelini arttırmak ve bir önceki nesilden daha iyi niteliklere sahip yeni diziler oluşturmak için kullanılır. Seçim işlemi sonucunda rastgele seçilen iki dizi karşılıklı çaprazlama işlemine tabi tutulur. Bu işlem iki dizinin ortasından rastgele alınan karşılıklı noktalardan çaprazlama yapılmak suretiyle gerçekleştirilir. Çaprazlama noktasının uygun olması ebeveynlerin uyum değerinin daha yüksek olduğu diziler elde edilmesine olanak sağlar. Ama uygun çaprazlama noktası kolay bulunmadığından nokta seçme işlemi genellikle rastgele yapılmaktadır.

Tek noktalı ve çift noktalı çaprazlama yöntemleri temel çaprazlama yöntemleridir. Tek noktalı çaprazlama, iki noktalı çaprazlama, sıra temelli çaprazlama gibi farklı yöntemleri vardır.

3.2.3.5. Mutasyon

Çaprazlama işlemi ile oluşturulan gen havuzunu zenginleştirmek için uygulanır. Mevcut gen havuzundan yeni diziler elde üretmek için mutasyon işlemine ihtiyaç

duyulur. Mutasyon, ikili kodlamada bir bireyin geninde herhangi bir bit değerini diğeri ile değiştirmek suretiyle yapılır.

Bu işlemler sırayla uygulandıktan sonra değerlendirilmeye hazır yeni bir popülasyon oluşmuş olur. Bu adımlar ilk başta belirlenecek durdurma kriteri sağlanana kadar yenilenir. Değer değiştirme yöntemi, kaydırma yöntemi, yerleştirme yöntemi ve karşılıklı değişim yöntemi gibi çeşitli yöntemleri vardır.

3.2.4. Birini dışarıda bırak çapraz doğrulama

Bir makine öğrenmesi modelinde yapılan testlerin hatasını daha iyi tahmin etmek için model seçiminde çapraz doğrulama tekniği kullanılır. Eğitim verileri setinden doğrulama kümeleri olarak adlandırılan örnek gözlem bölümlerinin oluşturulması çapraz doğrulamanın ana fikridir. Eğitim verilerine yerleştirilen her modelin performansı eklenen her doğrulama kümesine karşı ölçülür. Daha sonra modelin nasıl performans göstereceğine dair daha iyi bir değerlendirme elde edilir.

Çapraz doğrulamanın en basit yöntemlerinden birisi Holdout yöntemidir. Veri seti, test ve eğitim seti olmak üzere ikiye ayrılır. İşlev yaklaşımıcısı, yalnızca eğitim setini kullanarak bir işleve uyar. Test setindeki verilerin sonuç değeri daha önce bu sonuç değerlerini hiç görmemiş olan fonksiyon yaklaşımı ile tahmin edilir. Yapılan hatalar, model değerlendirmede kullanılan mutlak hata test kümesinde biriktirilir. Bu yöntemin avantajı, artık yöntemle göre genellikle tercih edilmesi ve hesaplanması artık gerekmemesidir. Ancak, değerlendirmesi yüksek bir varyansa sahip olabilir. Değerlendirme, eğitim setinde hangi veri noktalarının sona erdiğine ve test setinde hangi sonuçların elde edildiğine büyük ölçüde bağlı olabilir ve bu nedenle değerlendirme, bölümün nasıl yapıldığına bağlı olarak önemli ölçüde farklı olabilir.

K-katlama çapraz doğrulama, Holdout yöntemine göre iyileştirmenin bir yoludur. Veri seti k alt gruba ayrılır ve bekleme yöntemi k kez tekrarlanır. Her seferinde, k altkümelerinden biri test seti olarak kullanılır ve diğer k-1 altkümeleri bir eğitim seti oluşturmak için bir araya getirilir. Sonra tüm k denemelerinde ortalama hata

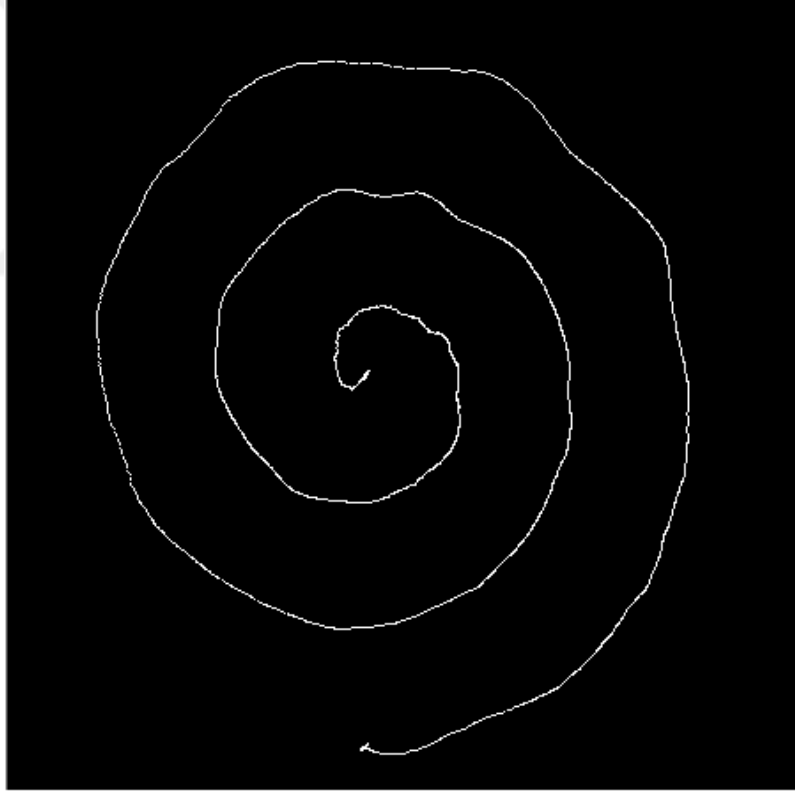
hesaplanır. Bu yöntemin avantajı, verinin nasıl bölündüğünün daha az önemli olması. Her veri noktası tam olarak bir kez bir test setine girer ve $k-1$ kez bir eğitim setine girer. k arttıkça elde edilen tahminin varyansı azalır. Eğitim algoritmasının sıfırdan k defa tekrarlanması, yani değerlendirme yapmak için k kat daha fazla hesaplamanın gerekmesi bu yöntemin dezavantajıdır. Bu yöntemin bir çeşidi, verileri rastgele bir test ve eğitim setine farklı zamanlarda bölmektir. Bu, bağımsız olarak her test setinin ne kadar büyük olduğunu ve ortalama kaç deneme yapacağınızı seçebilmenizi sağlar.

Birini Dışarıda Bırak Çapraz Doğrulama (Leave-One-Out Cross Validation), kümedeki veri noktalarının sayısı N 'ye eşit olan, K mantıksal sınırına alınmış K katlamalı çapraz doğrulamadır. Setteki veri noktalarının sayısı K , N 'ye eşittir. Bunun anlamı N 'nin ayrı zamanlarda, fonksiyon yaklaşımının bir nokta dışındaki tüm veriler üzerinde eğitilmesi ve bu nokta için bir tahmin yapılmasıdır. Ortalama hata hesaplanarak modelin değerlendirilmesi sağlanır. Tek çıkışlı çapraz doğrulama hatası (leave one out cross validation error) ile yapılan değerlendirme başarılıdır, ancak ilk hesaplanması maliyetli olmaktadır. Birini dışarıda bırak çapraz doğrulama hatasının hesaplanması, artık hatanın hesaplanmasından daha fazla zaman alması anlamına gelmez ve modelleri değerlendirmek için çok daha iyi bir yoldur.

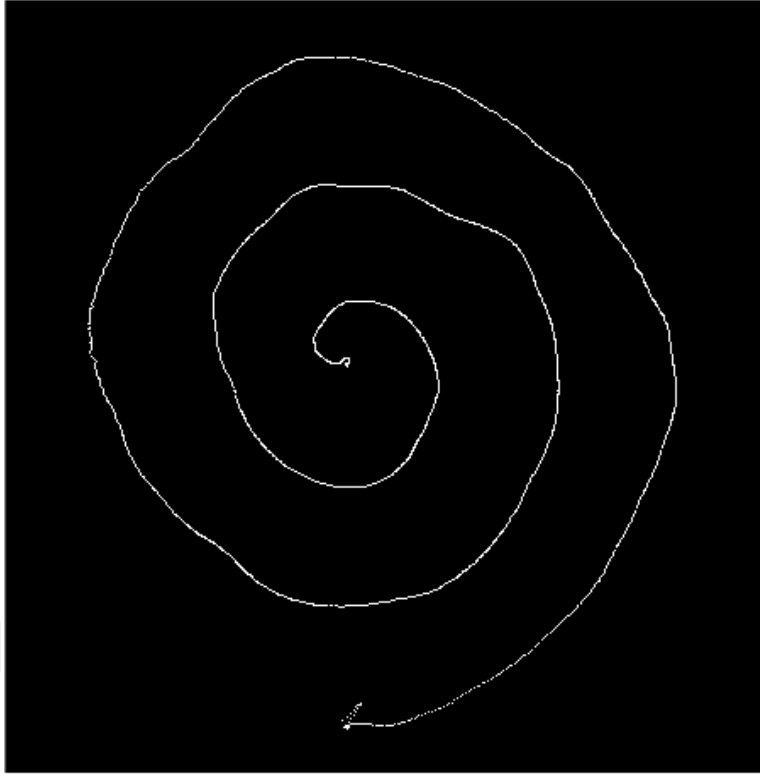
4. ARASTIRMA BULGULARI ve TARTISMA

4.1. Deneyisel Calismalar

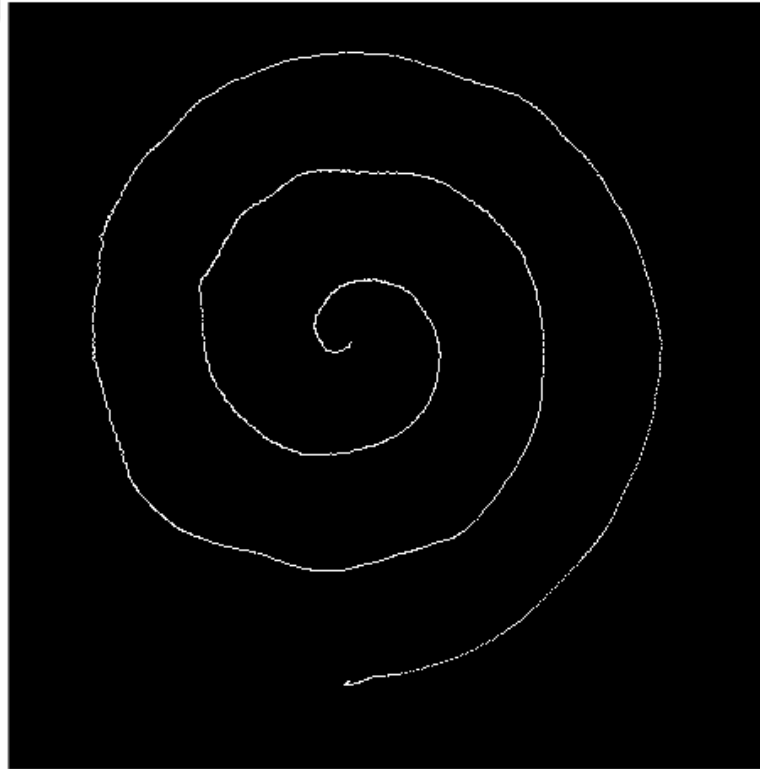
Isenkul ve ark. (2014) tarafından hazirlanan ve koordinat noktalarından oluřan veriler, dönüřtürölerek statik ve dinamik çizim řeklinde ikilik görüntü diziler elde edilmiřtir. řekil 4.1.'de bir Parkinson hastasının Statik Spiral Test (SST) çizimi, řekil 4.2.'de bir Parkinson hastasının Dinamik Spiral Test (DST) çizimi, řekil 4.3.'te bir sađlıklı bir insanın Statik Spiral Test (SST) çizimi, řekil 4.4.'te ise sađlıklı bir insanın Dinamik Spiral Test (DST) çizimi verilmiřtir.



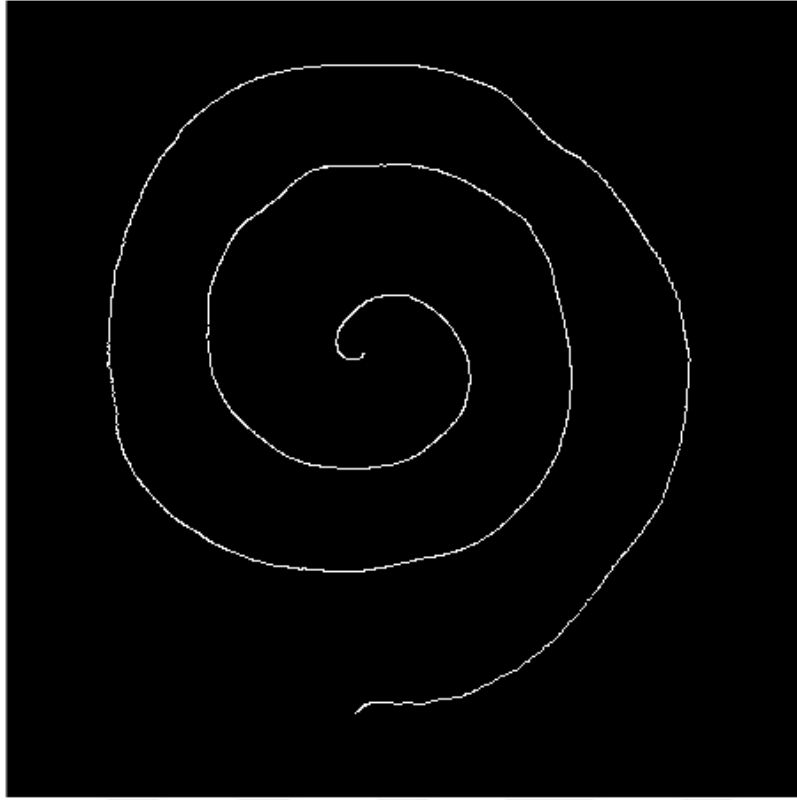
řekil 4.1. Parkinson Hastasının Statik Spiral Test Çizimi



Şekil 4.2. Parkinson Hastasının Dinamik Spiral Test Çizimi



Şekil 4.3. Sağlıklı İnsanın Statik Spiral Test Çizimi



Şekil 4.4. Sağlıklı İnsanın Dinamik Spiral Test Çizimi

Çizim verilerinden oluşan ve ikilik görüntü dizi olan bu verilere Şekil 4.5.’te görüldüğü gibi statik çizimin tamamlanma süresi, statik çizim boyunca kalem ile tablete uygulanan basınç ortalaması, dinamik çizimin tamamlanma süresi, dinamik çizim boyunca kalem ile tablete uygulanan basınç ortalaması ve kişinin Parkinson hastası ya da sağlıklı bir insan olduğunu belirten hastalık etiketi verileri de eklenmiştir. Her bir birey için veri uzunluğu 1x405.005’lik bir boyuta ulaşırken, toplam veri 71 birey için 71x405.005’lik bir boyuta ulaşmıştır.

405.005							
0000.....	0	0		0000	☐	☐	☐
0000.....	0	1		0000	☐	☐	☐
0001.....	0	0		0010	☐	☐	☐
0010.....	0	1		1000	☐	☐	☐
.....					☐	☐	☐
.....					☐	☐	☐
.....					☐	☐	☐
.....					☐	☐	☐
.....					☐	☐	☐
1000.....	0	0		0100	☐	☐	☐
0100.....	0	0		0001	☐	☐	☐
0001.....	0	1		1000	☐	☐	☐
0010.....	0	0		0000	☐	☐	☐
X ✓							

Şekil 4.5. Çizim Verilerinden Elde Edilen Ham Veri Seti

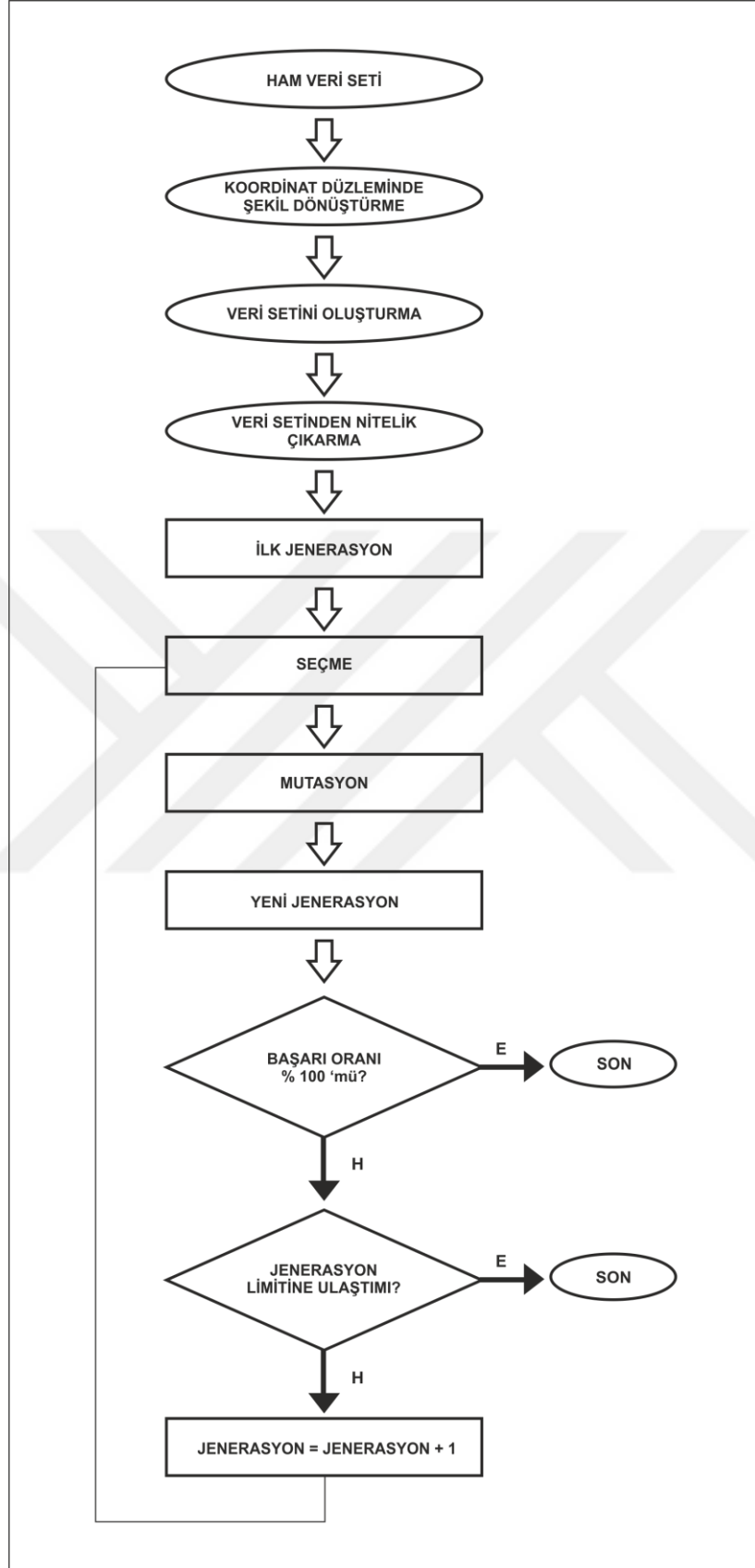
Elde edilen bu veri seti incelendiğinde ve spiral çizimlerinin birçok noktasında herhangi bir çizimin bulunmaması göz önünde bulundurulduğunda öznitelik sayısını azaltmak ve veri setinin gereksiz verilerden temizlemek amacıyla tüm örneklerde aynı değere sahip nitelikler çıkarılmıştır (Şekil 4.5.). Böylelikle veri setinde standart olan ve tüm çizimlerde bulunan veriler çıkarılmış olup elde edilen nitelikler farklılık içeren bilgiler olarak ele alınmıştır.

Sonuç olarak; Şekil 4.6.'da gösterildiği gibi 71x123066 boyutunda bir veri elde edilmiştir. Örnek sayısının 71, öznitelik sayısının 123066 olduğu düşünüldüğünde elde edilen veri setinde öznitelik seçme algoritmalarının kullanılması gerektiği anlaşılmaktadır.

		123.066	
	0000.....0.....0000		☐ Hata Etiket
	0000.....0.....0000		☐ Dinamik Basınç Ortalaması
	0001.....1.....0010		☐ Dinamik Çözüm Süresi
	0010.....0.....1000		☐ Statik Basınç Ortalaması
			☐ Statik Çözüm Süresi
			
71			
			
			
	1000.....1.....0100		
	0100.....0.....0001		
	0001.....0.....1000		
	0010.....0.....0000		

Şekil 4.6. Gereksiz Verilerden Temizlenmiş Veri Seti

Gereksiz özellikler veri setinden temizlendikten sonra veri seti sınıflandırma için makine öğrenme algoritmasına gönderilmiştir. Ek olarak, öznelilik seçme yöntemlerinin etkisini test etmek için Genetik Algoritma kullanılmıştır. Genetik algoritma temelli öznelilik seçme yönteminin işlem adımları Şekil 4.7.'de gösterilmiştir.



Şekil 4.7. Genetik Algoritma Tabanlı Öznitelik Seçme İşlemi Süreci

Ham veri setindeki koordinat verileri kullanılarak koordinat düzleminde şekil dönüştürme işlemi yapıldıktan sonra yukarıda bahsedildiği gibi çizim verileri, zaman ve basınç bilgileri ve hastalık etiketi gibi bilgiler eklenerek kendi veri setimiz oluşturulmuştur.

Gereksiz özellikler veri setinden temizlendikten sonra öznitelik seçmeden k-EK sınıflandırma ve Karar Ağacı yöntemleri uygulanmıştır. Ayrıca Şekil 4.7’de gösterildiği gibi Genetik Algoritma öznitelik seçme yöntemi kullanılarak k-EK sınıflandırma ve Karar Ağacı yöntemleri de uygulanmıştır. Burada ilk jenerasyon oluşturulduktan sonra seçme ve mutasyon işlemleri uygulanarak yeni jenerasyon elde edilmiştir. Başarı oranına ya da jenerasyon sınırına ulaşılmamışsa jenerasyon sayısı arttırılarak tekrar jenerasyon seçme adımına dönülür.

Öznitelik seçme işleminin sona ermesi iki durumda gerçekleşir: %100 başarı oranına ulaşıldığında ya da üretim sınırı aşıldığında.

4.2. Araştırma Bulguları

Bu çalışmada Parkinson hastalarına çizdirilen spiral çizimlerinden Parkinson hastalığının sınıflandırılması konusu ele alınmıştır. Çalışmada, kişilere çizdirilen spiral görüntülerinden kişilerin Parkinson hastası olup olmadığına karar verilmektedir.

Parkinson hastalığının teşhisinin başarısını değerlendirmek için Eşitlik 4.1.’de ifade edilen Duyarlılık, Eşitlik 4.2.’de ifade edilen Özgüllük, Eşitlik 4.3.’te ifade edilen Doğruluk, Eşitlik 4.4.’te ifade edilen Hassasiyet ve Eşitlik 4.5.’te ifade edilen Matthews Korelasyon Katsayısı (MKK) parametreleri kullanılmaktadır. Bu ölçütlerin hesaplanabilmesi için Doğru Pozitif (DP), doğru olarak tahmin edilen hasta sayısı, yanlış bir şekilde hasta olarak tahmin edilen hasta olmayan kişiler Yanlış Pozitif (YP), Hasta olarak tahmin edilmeyen hastaların sayısı ise Yanlış Negatif (YN) ve hasta olmadığı tahmin edilen hasta olmayan kişilerin sayısı ise Doğru Negatif (DN) olmak üzere 4 adet parametrede hesaplanmıştır.

$$Duyarlılık = \frac{DP}{DP+YN} \quad (4.1.)$$

$$Özgüllük = \frac{DN}{YP+DN} \quad (4.2.)$$

$$Doğruluk = \frac{DP+DN}{DP+DN+YP+YN} \quad (4.3.)$$

$$Hassasiyet = \frac{DP}{DP+YP} \quad (4.4.)$$

$$MKK = \frac{(DP \times DN) - (YP \times YN)}{\sqrt{(DP+YP) \times (DP+YN) \times (DN+YP) \times (DN+YN)}} \quad (4.5.)$$

Elde edilen sonuçlar, 2.8 GHz frekansa sahip Intel Core i7 işlemci, 8 GB RAM hafıza bulunan bir bilgisayarda gerçekleştirilmiş olup, MATLAB yazılım ortamında algoritmalar kodlanmıştır. Bu şartlarda elde edilen sonuçlar Çizelge 4.1.'de verilmiştir. Çizelge 4.1.'de elde edilen sonuçlar Birini Dışarda Bırak Çapraz doğrulama yöntemi ile hesaplanmıştır. k-EK sınıflandırma yöntemi için k değeri 1 olarak seçilmiştir. Çalışma kapsamında öznitelik seçme yapılmadan k-EK (k-EK), öznitelik seçme yapılmadan karar ağaçları (KA), Genetik algoritma ile öznitelik seçme yapılarak k-EK (GA-k-EK), Genetik algoritma ile öznitelik seçme yapılarak karar ağaçları (GA-KA) olmak üzere 4 farklı sınıflandırma yapılmış ve elde edilen sonuçlar verilmiştir.

Çizelge 4.1. Karışıklık Matrisleri

		Tahmin Değerleri	
		P	N
Gerçek Değerler	P	45	11
	N	10	5

A. k-EK

		Tahmin Değerleri	
		P	N
Gerçek Değerler	P	54	2
	N	12	3

B. Karar Ağacı

		Tahmin Değerleri	
		P	N
Gerçek Değerler	P	53	3
	N	5	10

C. GA - k-EK

		Tahmin Değerleri	
		P	N
Gerçek Değerler	P	56	0
	N	0	15

D. GA – Karar Ağacı

Çizelge 4.1.'de sınıflandırma sonucunda elde edilen karışıklık matrisleri gösterilmektedir. Görüldüğü üzere öznitelik seçme yöntemleri daha başarılı bir sınıflandırma işlemi yapılmasını sağlamıştır. Eşitlik (4.1. - 4.4.)'e göre hesaplanan başarı parametreleri Çizelge 4.2.'de verilmiştir.

Çizelge 4.2. Sınıflandırma Başarı Oranları

Sınıflandırma Yöntemleri	Öznitelik Seçme Yöntemi	Doğruluk (Accuracy)	Duyarlılık (Sensitivity)	Özgüllük (Specificity)	Hassasiyet (Precision)	MKK
k-EK	-	0,7042	0,8036	0,3333	0,8182	0,1338
Karar Ağacı	-	0,8028	0,9643	0,2000	0,8182	0,2621
GA – k-EK	GA	0,8873	0,9464	0,6667	0,9138	0,6471
GA- Karar Ağacı	GA	1	1	1	1	1

Çizelge 4.2.'deki verilere göre Parkinson hastalarını spiral çizimleri ile sınıflandırma da en başarılı durum genetik algoritma ile öznitelik seçme yapılarak karar ağaçları yöntemi ile elde edilmiştir. Özellikle oluşturulan veri setinde örnek sayısının az, nitelik sayısının fazla olduğu düşünüldüğünde öznitelik seçme işleminin sınıflandırma başarı oranına önemli iyileştirmeler sunduğu söylenebilir. Çizelge 4.3.'te ise yapılan sınıflandırma işlemleri işlem süreleri bakımından karşılaştırılmaktadır.

Çizelge 4.3. İşlem Süreleri

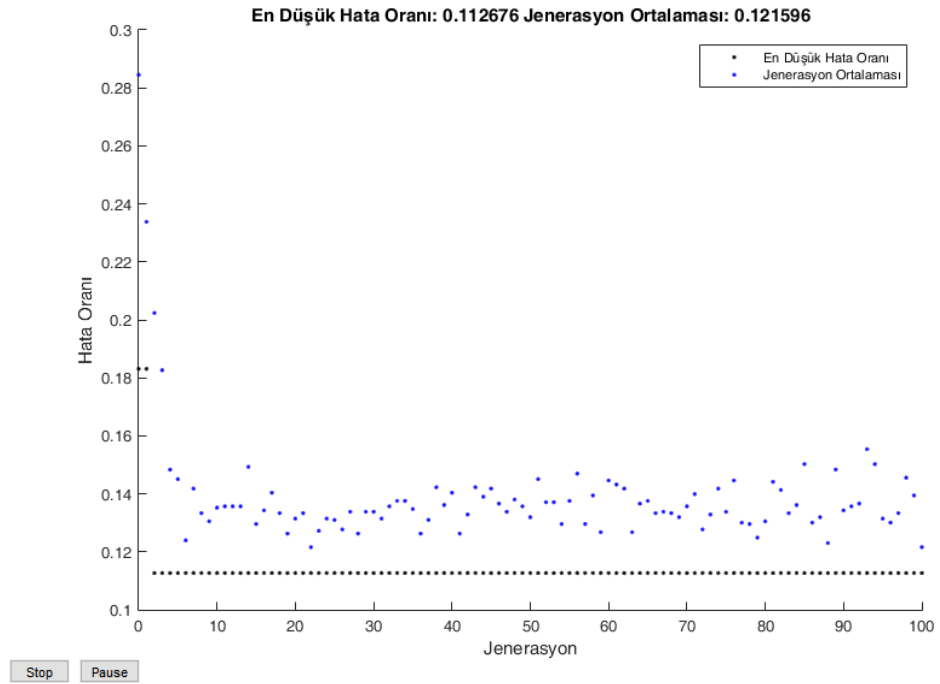
Sınıflandırma Yöntemi	Öznitelik Seçme Yöntemi	Öznitelik Seçme Zamanı (sn.)	Sınıflandırma Zamanı (sn.)	Toplam süre (sn.)
k-EK	-	-	7,3867	7,3867
Karar Ağacı	-	-	42,5079	42,5079
k-EK	GA	14340,6603	1,9405	14342,6008
Karar Ağacı	GA	26497,9599	16,3686	26514,3285

Toplam Süre bakımından GA ile öznitelik seçmenin işlem süresini oldukça fazla olumsuz etkilediğini fakat nitelik sayılarını azaltması sebebiyle öznitelik seçme işleminden sonra yapılan sınıflandırmanın daha kısa sürede tamamlandığı gözlemlenmektedir.

Çizelge 4.4. GA ile Seçilen Öznitelik Sayısı

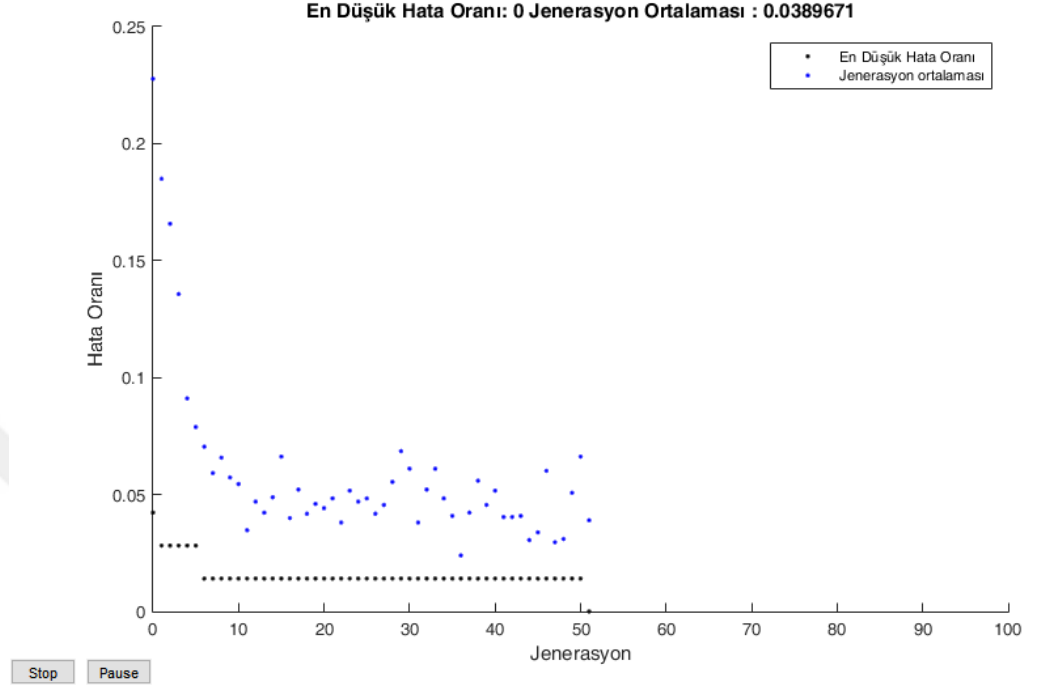
Öznitelik Seçme Yöntemi	Sınıflandırma Yöntemi	Seçilen Öznitelik Sayısı
Genetik Algoritma	k-EK	60954
Genetik Algoritma	Karar Ağacı	60768

Ham veride 405.005 olan nitelik sayısı tekrarlanan ve gereksiz olan verilerin çıkarılması ile başlangıçta 123.066'ya düşürülmüştü. Çizelge 4.4.'te görüldüğü gibi Genetik algoritma öznitelik seçme ve k-EK sınıflandırma metodu ile yapılan uygulamada seçilen öznitelik sayısı 60.954, Genetik algoritma öznitelik seçme ve Karar Ağacı sınıflandırma metodu ile yapılan uygulamada ise seçilen öznitelik sayısı 60.768 olarak bulunmuştur.



Şekil 4.8. GA Öznitelik Seçme ile k-EK Sınıflandırma Hata Oranı

Şekil 4.8.'de görüldüğü gibi Genetik algoritma öznitelik seçme ve k-EK sınıflandırma metodu ile yapılan uygulamada en düşük hata oranı 0.112676, jenerasyon ortalaması 0.121596, jenerasyon sayısı da 100 olmuştur.



Şekil 4.9. GA Öznitelik Seçme ile Karar Ağacı Sınıflandırma Hata Oranı

Şekil 4.9.'da görüldüğü gibi Genetik algoritma öznitelik seçme ve Karar Ağacı sınıflandırma metodu ile yapılan uygulamada en düşük hata oranı 0, jenerasyon ortalaması 0.0389671, jenerasyon sayısı da 51 olmuştur.

5. SONUÇLAR ve ÖNERİLER

Günümüzde hastalık teşhislerinin erken yapılması büyük önem arz etmektedir. Parkinson hastalığının teşhisinin de zor olması, teşhis için yardımcı araçlara olan ihtiyacı arttırmaktadır. Bilgisayar yardımıyla Parkinson hastalığının teşhisine yardımcı olmak için günümüzde çeşitli çalışmalar yapılmakta olup bu çalışma ile Parkinson hastalığının teşhisinde spiral test çiziminin katkısının olabilirliği araştırılmıştır. Ayrıca hasta ve kontrol grubu kişilerin çizim verilerinin öznitelik seçme işlemi ile belirlenip sınıflandırılması ile yönüyle de özgün bir çalışma olmuştur.

Günümüzde hastaların bilgileri yaş, cinsiyet, tedavi durumu gibi veriler ile tutulmaktadır. Bu verilere ek olarak, el yazısı, ses kaydı ve bu tezde çalışması yapılan çizim verileri gibi hastalığın gelişimi ve tedavi sürecine büyük ölçüde katkı sağlayabilecek bilgilerin de saklanması tedavi sürecinde büyük önem arz etmektedir. Bu yüzden hastaların iyileşme ya da gerileme gibi değişimlerini takip etmek ve bu verilerden istatistiki çıkarımlar/öngörüler yapmak zor olmaktadır. Bu çalışma, hastalık teşhisinin yanında hasta izlemenin de daha objektif olması bakımından özgün bir çalışmadır.

Bu tez çalışmasında Parkinson spiral test analizi ile elde edilen veriler doğrultusunda Parkinson hastalığının teşhisi yapılmıştır. Parkinson spiral testi, literatürde yapılan çalışmalardan farklı olarak, karakter tanıma problemi olarak ele alınmıştır. İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Nöroloji Anabilim Dalı'nda oluşturulan veri seti kullanılarak önerilen yöntem test edilmiştir. Elde edilen sonuçlar ele alındığında sonuçların literatürdeki çalışmalara oranla başarılı olduğu görülmektedir.

Karışıklık matrisleri sonuçları incelendiğinde öznitelik seçme işlemi yapılmadan yapılan Karar Ağacı ile sınıflandırma yönteminin k-EK sınıflandırma yöntemine göre daha başarılı olduğu gözlemlenmiştir. Genetik Algoritma ile öznitelik seçme işlemi yapıldıktan sonra Karar Ağacı ile sınıflandırma yönteminin k-EK sınıflandırma

yöntemine göre yine başarılı olduğu tespit edilmiştir. Genetik algoritma ile öznitelik seçme işlemi yapıp Karar Ağacı ile sınıflandırma yöntemi tüm işlemler göz önüne alındığında daha başarılı olmuş ve % 100 doğru tahmin oranına ulaşmıştır.

Sınıflandırma başarı oranları incelendiğinde Genetik algoritma ile Karar Ağacı sınıflandırma işleminin doğruluk, duyarlılık, özgüllük, hassasiyet ve MKK parametrelerinde en iyi sonucu verdiği görülmektedir. Öznitelik seçme işlemi yapılmadan Karar Ağacı ile sınıflandırma yönteminin doğruluk, duyarlılık ve MKK parametrelerinde daha iyi sonuçlar elde edilirken, özgüllük parametresi daha düşük olmuş ve hassasiyet parametresinin de değişmediği gözlemlenmiştir.

İşlem süreleri incelendiğinde öznitelik seçme yöntemlerinin toplam işlem sürelerini olumsuz yönde etkilediği görülmüştür. Genetik algoritma ile öznitelik seçme işleminin süresi oldukça yüksek olurken sınıflandırma zamanını düşürdüğü tespit edilmiştir. Genetik algoritma ile Öznitelik seçme işleminin bir kez yapılacağı göz önüne alındığında işlem süresinin fazla olması daha sonraki çalışmaları etkilemeyecektir. Yeni bir veri seti geldiğinde sadece öznitelikleri seçilip bakılacak, yani tekrardan öznitelik belirleme işlemi yapılmayacaktır. Belirlediğimiz öznitelikler üzerinden testler yapılacaktır. Nitelik sayılarını azaltması sebebiyle sınıflandırma zamanını daha da düşürdüğü dikkate alındığında olumlu bir etki yapacaktır.

Başlangıçta 123.066 olan nitelik sayısı Genetik algoritma öznitelik seçme ve k-EK sınıflandırma metodu ile yapılan uygulama ile 60.954'e, Genetik algoritma öznitelik seçme ve Karar Ağacı sınıflandırma metodu ile yapılan uygulamada ise 60.768'e düşmüştür. Sonucu etkileyen, önemli olan nitelik sayıları öznitelik seçme yöntemleri ile %50'den fazla oranda eksiltilmiştir. Öznitelik sayısının düşürülmesi işlem sürelerini azaltırken, sınıflandırma başarı oranlarının artmasına neden olmuştur.

Genetik algoritma ile k-EK sınıflandırma uygulamamızda üçüncü jenerasyondan itibaren en düşük hata oranına ulaşılmış, onuncu jenerasyondan sonra jenerasyon ortalaması dengeye ulaşmıştır. Yeterince iterasyon sayısı verilmesine rağmen hata oranı sıfıra ulaşmadığı için jenerasyon limiti 100'e ulaştığında işlem sonlanmıştır.

Genetik algoritma ile Karar Ağacı sınıflandırma uygulamamızda en düşük hata oranı olan sıfıra 51. jenerasyonda ulaşılmıştır. Jenerasyon ortalaması 10. adımdan sonra dengeyi sağlamıştır. Hata oranı 51. jenerasyonda sıfıra ulaştığı için işlem bu adımda sonlanmıştır.

Bu çalışma sonucunda Parkinson hastalığının teşhisinde spiral testinde örüntü tanıma yöntemleri kullanılarak başarılı sonuçlar elde edilebileceği görülmüştür. Erken dönemde teşhisi zor olan Parkinson hastalığı önerilen bu yöntemle daha etkili bir şekilde teşhis edilebilmektedir. Yapılan uygulama Parkinson hastalığının teşhisinin yanı sıra Parkinson hastalarının durum takibinde de kullanılabileceği öngörülmektedir. Hastalığın ilerleyip ilerlemediği, hangi aşamada olduğu gözlemlenebilir.

KAYNAKLAR

- A. AGARWAL, S. CHANDRAYAN and S. S. SAHU, 2016. Prediction of Parkinson's disease using speech signal with Extreme Learning Machine, 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai. 3776-3779.
- ALY, N.M., PLAYFER, J.R., SMITH, S.L. and HALLIDAY, D.M., 2007. A Novel Computer-Based Technique for The Assessment of Tremor in Parkinson's Disease, Age and Ageing, 36; 395–399.
- APAYDIN, H., 2001. Parkinson hastalığının klinik özellikleri. Parkinson Hastalığı ve Hareket Bozuklukları Dergisi, 4(2); 114-24.
- BENBA, A., JİLBAB, A., and HAMMOUCH, A., 2015. Detecting Patients with Parkinson's disease using Mel Frequency Cepstral Coefficients and Support Vector Machines, International Journal on Electrical Engineering and Informatics, 7(2).
- BRAGA, D., MADUREIRA, A. M., COELHO, L. and AJITH, R., 2019. Automatic detection of Parkinson's disease based on acoustic analysis of speech. Engineering Applications of Artificial Intelligence, 77; 148–158.
- CHEN, S.-W., LIN, S.-H., LIAO, L.-D., LAI, H.-Y., PEI, Y.-C., KUO, T.-S., LIN, C.-T., CHANG, J.-Y. and CHEN, Y.-Y., 2011. Quantification and recognition of parkinsonian gait from monocular video imaging using kernel-based principal component analysis, Biomedical Engineering Online, 10(1);1–21.
- DANNA, J., VELAY, J., EUSEBIO, A., VÉRON-DELOR, L., WITJAS, T., AZULAY J. and PINTO S., 2018. Digitalized spiral drawing in Parkinson's disease: A tool for evaluating beyond the written trace. Human Movement Science, ISSN 0167-9457.
- DELİL, Ş., 2001, Parkinson Hastalığında Levodopa'ya Bağlı Motor Komplikasyonların Gelişmesinde Rol Oynayan Etkenler, Uzmanlık Tezi. İstanbul Üniversitesi.
- ERTAN, S., 2005. Parkinson Hastalığının Klinik Özellikleri, Nörolog Olmayanlar için Nöroloji. İstanbul, 42:249-254.
- FAHN, S. and PRZEDBORSKI, S., 2000. Parkinsonizm. In: Rowland LP ed. Merritt's Textbook of neurology. 10th ed., Lippincott Williams & Wilkins, Philadelphia, 9789752771819, s679-s693.
- GÜLGEZEN, G., 2009. Kararlı ve başarımlı yüksek öznelilik seçimi. Yüksek Lisans Tezi, İstanbul Teknik Fen Bilimleri Enstitüsü, İstanbul, 15-17.
- GÜMÜŞÇÜ, A., KARADAĞ, K., TENKECİ, M.E. ve AYDİLEK, İ.B., 2017. Ses Analizi ile Parkinson Hastalığının Teşhisinde Genetik Algoritma ile Öznelilik Seçilmesi, 25. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Antalya, Türkiye, 4s.
- GÜMÜŞÇÜ, A., TENKECİ, M.E. ve BİLGİLİ, A. V., 2019. Estimation of wheat planting date using machine learning algorithms based on available climate data. Sustainable Computing, Informatics and Systems.
- HAHNE, U., SCHILD, J., ELSTNER, S. and ALEXA, M., 2009. Multi-touch focus+context sketch-based interaction, 77.

- ISENKUL, M.E., SAKAR, B.E. ve KURSUN, O. , 2014. Improved spiral test using digitized graphics tablet for monitoring Parkinson's disease. The 2nd International Conference on e-Health and Telemedicine (ICEHTM-2014), 171-175.
- İSENKUL, M. E., 2011. Parkinson Hastalığı'nın Teşhisi İçin Veri Toplama ve Örüntü Tanıma Sistemi, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 53s.
- JANKOVIĆ, J., 2008. Parkinson's disease: clinical features and diagnosis, J. Neurol. Neurosurg. Psychiatry, 79(4):368–376, 2008.
- LITTLE, M.A., MCSHARRY, P.E., HUNTER, E.J., SPIELMAN, J., RAMIG, L.O., 2009. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. IEEE Transactions on Biomedical Engineering, 56(4); 1015-1022.
- PEREIRA, C. R., PEREIRA, D. R., Rosa, G. H., ALBUQUERQUE, V. H. C., WEBER, S. A. T., HOOK, C. and Papa, J. P. (2018). Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. Artificial Intelligence in Medicine, 87, 67–77.
- PULLMAN, SL., 1998. Spiral analysis: a new technique for measuring tremor with a digitizing tablet. Movement Disorders, 13 (3); 85-89.
- REVERT, K., GORUNESCU, F. and SALEM, A.B.M., (2009). Feature Selection in Parkinson's Disease. A Rough Sets Approach. Proceedings of the International Multiconference on Computer Science and Information Technology, Poland, 425-428.
- QUINLAN J.R. , 1996. Improved use of continuous attributes in C4.5 for Journal of Artificial Intelligence Research, 4; 77-90
- SAKAR, B. E., ISENKUL, M. E., SAKAR, C. O., SERTBAS, A., GURGEN, F., DELİL, S. and KURSUN, O., 2013. Collection and Analysis of a Parkinson Speech Dataset With Multiple Types of Sound Recordings. IEEE Journal of Biomedical and Health Informatics, 17(4); 828–834.
- SAN LUCIANO, M., WANG, C., ORTEGA, R. A., YU, Q., BOSCHUNG, S., SOTO-VALENCIA, J. and SAUNDERS-PULLMAN, R., 2016. Digitized spiral drawing: A possible biomarker for early Parkinson's disease. PLoS ONE, 11(10), 1–11.
- SAUNDERS-PULLMAN, R., DERBY, C., STANLEY, K., FLOYD, A., BRESSMAN, S., LIPTON, R. B., DELIGTISCH, A., SEVERT, L., YU, Q. and KURTISI M., 2008. Validity of spiral analysis in early Parkinson's disease. Movement Disorders, 23 (4); 531-537.
- SHARMA, P., SUNDARAM, S., SHARMA, M., SHARMA, A., GUPTA D., 2019. Diagnosis of Parkinson's disease using modified grey wolf optimization. Cognitive Systems Research, 54; 100-115.
- STEPHEN, K., VAN DEN EEDEN, S. K., TANNER, C. M., BERNSTEIN, A. L., FROSS, R. D., LEIMPETER, A., BLOCH, D. A. and NELSON, L. M., 2003. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. Am J Epidemiol, 157; 1015-1022.
- ŞEN, B. and PEKER, M., 2013. Novel approaches for automated epileptic diagnosis using FCBF feature selection and classification algorithms. Turkish Journal of Electrical Engineering & Computer Sciences, 21; 2092-2109.

- TSANAS, A., LITTLE, M. A., MCSHARRY, P. E., SPIELMAN, J. and RAMIG, L.O., 2012. Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. *IEEE transactions on biomedical engineering*, 59(5).
- TSANAS, A., LITTLE, M.A., MCSHARRY, P.E. and RAMIG, L.O., 2009. Accurate Telemonitoring of Parkinson's Disease Progression by Non-Invasive Speech Tests. *Nature Precedings*.
- TSANAS, A., LITTLE, M.A., MCSHARRY, P.E. and RAMIG, L.O., 2010. Enhanced Classical Dysphonia Measures and Sparse Regression for Telemonitoring of Parkinson's Disease Progression. *IEEE International Conference on Acoustics Speech and Signal*, 594-597.
- TUGWELL, C., 2008. *Parkinson's Disease on Focus*. Chicago: Pharmaceutical Press, 7-8.
- ÜNLÜ, A., BRAUSE, R., and KRAKOW, K., 2006. Handwriting Analysis for Diagnosis and Prognosis of Parkinson's Disease. *Proc. Int. Symp. Biological and Medical Data Analysis*, 4345; 441-450.
- VAN GEMMERT, A.W.A., ADLER, CH. and STELMACH, G.E., 2003. Parkinson's disease patients undershoot target size in handwriting and similar tasks. *J Neurol Neurosurg Psychiatry*, 74;1502-1508.
- WANG, H., YU, Q., KURTIS, M., FLOYD, A., SMITH, W. and PULLMAN, S., 2008. Spiral analysis—Improved clinical utility with center detection. *Journal of Neuroscience Methods*, 171-2; 264- 270.
- WATERS, CH., BÜYÜKKAL B., 2000. *Parkinson Hastalığının Tanı ve Tedavisi* , Turgut Yayıncılık ve Tic. A.Ş., İstanbul.
- YU, L. and LIU, H., 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 21-24 August 2003, Washington DC.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Mustafa GERGER
Uyruğu: T.C.
Doğum Yeri ve Tarihi: Şanlıurfa / 03.07.1987
Telefon: 04143132883 - 376
Faks: 04143132882
E-Posta: mgerger@harran.edu.tr

EĞİTİM BİLGİLERİ

Derece **Adı, İlçe, İl, Bitirme Yılı**
Lise: Şanlıurfa Anadolu Lisesi, Merkez, Şanlıurfa, 2005
Üniversite: Karadeniz Teknik Üniversitesi Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü, Merkez, Trabzon, 2010

İŞ DENEYİMİ

Yıl	Kurum	Görevi
2012 -	GAP Tarımsal Araştırma Enstitüsü Müdürlüğü	Araştırmacı, Bilgi İşlem Sorumlusu
2010-2012	Harran Üniversitesi Yüksek Başarımlı Hesaplama Merkezi	Proje Mühendisi