



**ÖZNİTELİK SEÇME VE SINIFLANDIRMA ALGORİTMALARINI
KULLANARAK SÜRDÜRÜLMÜŞ FONASYON TESTLERİNDEN WEB
TABANLI PARKİNSON HASTALIĞININ TEŞHİSİ**

Bünyamin GÖKPINAR

**YÜKSEK LİSANS TEZİ
ELEKTRONİK VE BİLGİSAYAR EĞİTİMİ**

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

NİSAN 2014

Bünyamin GÖKPINAR tarafından hazırlanan “ÖZNİTELİK SEÇME VE SINIFLANDIRMA ALGORİTMALARINI KULLANARAK SÜRDÜRÜLMÜŞ FONASYON TESTLERİNDEN WEB TABANLI PARKİNSON HASTALIĞININ TEŞHİSİ” adlı tez çalışması Gazi Üniversitesi Elektronik-Bilgisayar Eğitimi Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Yrd. Doç. Dr. İsmail ATACAK

Teknoloji Fakültesi-Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Başkan: Prof. Dr. Ömer Faruk BAY

Elektronik ve Bilgisayar Eğitimi-Kontrol A.B.D., Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Üye: Doç. Dr. Murat YÜCEL

Teknoloji Fakültesi-Elektrik-Elektronik Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum

Tez Savunma Tarihi: / /

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartların yerine getirdiğini onaylıyorum.

.....
Doç. Dr. Nurettin TOPALOĞLU
Bilişim Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
 - Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
 - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Bünyamin GÖKPINAR

29.04.2014

ÖZNİTELİK SEÇME VE SINIFLANDIRMA ALGORİTMALARINI KULLANARAK
SÜRDÜRÜLMÜŞ FONASYON TESTLERİNDEN WEB TABANLI PARKİNSON
HASTALIĞININ TEŞHİSİ
(Yüksek Lisans Tezi)

Bünyamin GÖKPINAR

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

Nisan 2014

ÖZET

Nörolojik bir hastalık olan Parkinson'un 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 vermelerini kolaylaştırmak ve konulan teşhislerde doğruluk oranını yükselterek, Türkiye'de nöroloji ve bilgisayar bilimleri literatürüne katkı sağlanması hedeflenmiştir. Bu çalışmanın en önemli kısımlarından birisi, Parkinson hastalarına ait karmaşık ses verilerinden problemi temsil edebilecek en iyi veri kümesinin tespiti için öznitelik seçme algoritmalarının kullanılmasıdır. Ayrıca Parkinson Hastalığının teşhisinde etkili bir sınıflandırma yapılabilmesi için yeni bir öznitelik seçme yaklaşımı önerilmektedir. 5 popüler öznitelik seçme algoritmalarının uygulanması ile elde edilen öznitelikler 5 popüler sınıflandırma algoritmalarına tabi tutulmuştur. Böylelikle farklı sınıflandırma algoritmaları ile elde edilen sonuçların doğruluk oranlarının kıyaslanması sağlanarak en iyi performans ortaya konulmuştur. Yapılan çalışmada önerilen yaklaşım ile seçilen 6 öznitelik kümesinin sınıflandırma doğruluk oranı %94,36 olarak elde edilmiştir.

Bilim Kodu : 702.1.022
Anahtar Kelimeler : Parkinson hastalığı, hastalık teşhisi, öznitelik seçme algoritmaları, sınıflandırma algoritmaları
Sayfa Adedi : 66
Danışman : Yrd. Doç. Dr. İsmail ATACAK

WEB BASED DIAGNOSIS OF THE PARKINSON DISEASE FROM SUSTAINED
VOWEL PHONATION TESTS USING FEATURE SELECTION AND
CLASSIFICATION ALGORITHMS
(M. Sc. Thesis)

Bünyamin GÖKPINAR

GAZİ UNIVERSITY
INFORMATICS INSTITUTE

April 2014

ABSTRACT

The diagnosis of Parkinson which is neuro-disease is especially difficult in its initial phase. For this reason computer based solution researches have widely increased to support medical decision recently. With the help of this research experts who have been working in this field are aimed to make easier decisions on the way to Parkinson diagnosis and contribute to neurological and computer sciences Literature in Turkey by increasing the accuracy rate in diagnosis. One of the most important aims of this research is to get information which can be classified from voice data which belong to Parkinson patients and use feature selection algorithms in order find out the best datum unit that can represent the problem from these complicated voice data. Furthermore, a new feature selection approach is suggested to be able to make an effective classification at diagnosis of Parkinson disease. Features which are supplied with the application of five popular feature selection algorithms are applied to five popular classifications of algorithms. In this way, the best performance has been sorted out by comparing the calculation periods of results that are supplied with different classification algorithms and their accuracy rates. In this study the classification accuracy rate of six feature clusters has been 94.36% by the help of suggested approach.

Science Code : 702.1.022
Key Words : Parkinson's disease, disease diagnosis, classification,
feature selection
Page Number : 66
Supervisor : Yrd. Doç. Dr. İsmail ATACAK

TEŞEKKÜR

Tez çalışmalarım boyunca gösterdiği her türlü destek ve yardımlarından dolayı danışman hocam, Sayın Yrd. Doç. Dr. İsmail ATACAK'a, bu çalışmanın yürütülmesinde her türlü bilgi ve deneyimlerini bana aktaran ve tezin hazırlanmasında yardımlarını eksik etmeyen çok değerli arkadaşlarım, Musa PEKER ve Ali YAZICI'ya, ayrıca manevi desteğiyle beni hiçbir zaman yalnız bırakmayan çok değerli eşim, Büşra GÖKPINAR'a en içten dileklerimle teşekkür ederim.

Bu çalışmamı, biricik oğlum Ali Asaf GÖKPINAR'a atfediyorum.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ	ix
ŞEKİLLERİN LİSTESİ	x
RESİMLERİN LİSTESİ	xi
SİMGELER VE KISALTMALAR	xii
1. GİRİŞ.....	1
2. GENEL KISIMLAR	5
2.1. Hastalığın Tanımı.....	5
2.2. Parkinson Hastalığının Klinik Özellikleri.....	6
2.2.1. Titreme (Tremor).....	7
2.2.2. Kaslarda sertleşme (Rijidite)	7
2.2.3. Hareketlerde yavaşlama (Akinezi/Bradikinezi)	8
2.2.4. Duruş ve denge bozuklukları (Postüral Instabilite).....	9
2.2.5. Parkinson hastalığında konuşma bozukluğu	9
2.3. Parkinson Hastalığında Bilgisayarlı Tabanlı Teşhis Çalışmaları	10
3. VERİ SETİ VE ALGORİTMALAR	13
3.1. Parkinson Hastalığı Veri Seti (PHV).....	13
3.1.1. Jitter sınıfı öznelilikler.....	13
3.1.2. Shimmer sınıfı öznelilikler	15
3.1.3. Diğer öznelilikler.....	16
3.2. Öznellik Seçme Algoritmaları	17
3.2.1. Hızlı korelasyon tabanlı filtre-Fast correlation based filter (FCBF).....	18
3.2.2. Minimum fazlalık maksimum ilişki-minimum Redundancy Maximum Relevance (mRMR)	21
3.2.3. Fisher Puanlama-Fisher Score (FS)	23
3.2.4. Relief algoritması	23
3.2.5. Gini index algoritması	24
3.3. Sınıflandırma Algoritmaları	25

	Sayfa
3.3.1. Yapay Sinir Ağları (YSA-ANN)	25
3.3.2. Radyal Tabanlı Fonksiyon Ağları (RBFN)	30
3.3.3. Karar Ağaçları (DT)	32
3.3.4. Destek Vektör Makineleri (SVM)	34
3.3.5. Rastgele Orman (RF).....	37
4. WEB ARAYÜZ YAZILIMI VE DENEYSEL SONUÇLAR	41
4.1. Web Arayüz Yazılımı.....	41
4.2. Matlab Builder NE ve Matlab Web Figure.....	42
4.3. Tasarlanan Web Arayüz Uygulaması	43
4.4. Yapılan Deneyler ve Sonuçların Değerlendirilmesi.....	47
5. SONUÇ VE ÖNERİLER	57
KAYNAKLAR	59
ÖZGEÇMİŞ	65

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Parkinson hastalığı veri seti öznitelikleri.....	13
Çizelge 3.2. Bazı aktivasyon fonksiyonları	27
Çizelge 4.1. Fisher Skor algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları	48
Çizelge 4.2. Gini indeks algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları	50
Çizelge 4.3. mRMR algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları	51
Çizelge 4.4. ReliefF algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları	52
Çizelge 4.5. FCBF algoritmasıyla seçilen öznitelik kümesine göre tüm yöntemlerde elde edilen en iyi başarı oranları.....	53
Çizelge 4.6. Yapılan deneylerde tüm yöntemlerde elde edilen en iyi başarı oranları.....	55
Çizelge 4.7. En iyi sonucun confusion matrix, accuracy, sensitivity ve specificity değerleri	56
Çizelge 4.8. En iyi sonucun diğer yapılan çalışmalara kıyaslanması.....	56

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. FCBF öznitelik seçme algoritması.....	18
Şekil 3.2. Proses elemanının temel elemanları	26
Şekil 3.3. İleri beslemeli tek katmanlı yapay sinir ağı	28
Şekil 3.4. İleri beslemeli çok katmanlı yapay sinir ağıları	29
Şekil 3.5. Geri beslemeli yapay sinir ağı	30
Şekil 3.6. Radyal tabanlı fonksiyon ağının yapısı.....	31
Şekil 3.7. Karar ağacı yapısı.....	33
Şekil 3.8. Destek vektör makineleri.....	35
Şekil 3.9. Destek vektörleri	36
Şekil 3.10. Destek vektör makinelerinde uzaklık hesaplama.....	36
Şekil 3.11. Rastgele orman yönteminde veri seçimi.....	38
Şekil 4.1. Parkinson hastalığı için uygulanan metot	41
Şekil 4.2. MATLAB Web Figure içeren bir uygulamanın çalışma yapısı	42
Şekil 4.3. Fisher skor algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler.....	48
Şekil 4.4. Gini indeks algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler.....	49
Şekil 4.5. mRMR algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler.....	50
Şekil 4.6. ReliefF algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler.....	52
Şekil 4.7. FCBF algoritması ile seçilen özniteliklerin sınıflandırma doğruluğundaki değişimler	53
Şekil 4.8. Öznitelik seçme algoritmaları uygulamasıyla en çok seçilen öznitelikler.....	54
Şekil 4.9. Hibrit yaklaşımla seçilen özniteliklerin sınıflandırma doğruluğundaki değişimler	55

RESİMLERİN LİSTESİ

Resim	Sayfa
Resim 3.1. Parkinson veri tablosu	17
Resim 3.2. Sinir hücresinin yapısı.....	26
Resim 4.1. Uygulama web arayüzü ve login ekranı	43
Resim 4.2. Veri setinin yüklenmesi.....	44
Resim 4.3. Öznitelik seçme algoritmasının seçilmesi.....	45
Resim 4.4. Sınıflandırma algoritmasının seçilmesi	46
Resim 4.5. Seçilen sınıflandırma algoritmasına göre elde edilen sonuçların gösterilmesi ..	47

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklama

$IG(A B)$	A'nın B değerlerine bağlı bilgi kazancı
$H(A)$	A değişkeninin entropisi
$H(B)$	B değişkeninin entropisi
$H(A B)$	A'nın B değerlerine bağlı entropisi
F_t	Fisher Ölçüt Skoru
W	Maksimum İlişki
V	Minimum İlişki

Kısaltmalar

Açıklama

PH	Parkinson hastalığı
apq3	3 noktalı genlik değişim katsayısı
apq5	5 noktalı genlik değişim katsayısı
apq11	11 noktalı genlik değişim katsayısı
dB	Desibel
F	Periyotlar veya genlikler arası fark
G	Genlik
Hz	Saniye başına düşen devir sayısı
HNR	Harmonik-gürültü oranı
FCBF	Hızlı korelasyon tabanlı filtre
SU	Simetri Belirsizlik
mRMR	Minimum fazlalık maksimum ilişki-minimum
maxMR	Maksimum ilişki
minMV	Minimum artıklık
MID	Ortak bilgi farkı
MIQ	Ortak bilgi oranı

Kısaltmalar**Açıklama****NHR**

Gürültü-harmonik oranı

P

Periyot

ppq5

5 noktalı periyot değişim katsayısı

rap

Bağıl ortalama değişimi

YSA

Yapay sinir ağı

SVM

Destek vektör makineleri

DT

Karar ağacı

RF

Rastgele ormanı

RBFN

Radyal tabanlı fonksiyon ağı

RNN

Radyal tabanlı yapay sinir ağı

UPDRS

Birleşik Parkinson Hastalığı Değerlendirme Ölçeği

FFNN

İleri beslemeli yapay sinir ağı

CART

Sınıflandırma ve regresyon ağaçları

1. GİRİŞ

İngiliz Doktor James Parkinson tarafından 1817 yılında ilk kez “shaking palsy” (titrek felç) şeklinde tanımlanan Parkinson hastalığı (PH), genelde hastanın motor becerileri, konuşması ve diğer vücut işlevlerinde zarara yol açan bir merkezi sinir sistemi bozukluğudur [1-3].

Hastalık 40-75 yaş civarında, sıklıkla da 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 [3].

Dünyada 4 milyondan fazla kişiyi etkileyen yaygın bir nörolojik sağlık sorunu olan Parkinson, ülkemizde belli bir yaş grubunda Alzheimer’dan sonra en sık görülen hastalıktır. Türkiye’de 130 bin Parkinson hastası 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 [4]. 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.

PH, tremor (titreme), rijidite (kaslarda sertleşme), bradikinezi (hareketlerde yavaşlama) 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 duyuşsal yakınmalar ve depresyon ile kişiyi etkilemektedir [1,5, 6-8].

Hastalıkta görülen bu belirtiler bireylerin işlerini ve günlük aktivitelerini tek başına sağlıklı bir biçimde yapmalarına engel olmaktadır. Bu sebeple hastalığın ilk evrelerinde hastanın iş hayatından ve sosyal hayattan soyutlanması, ileri evrelerinde ise yaşamlarını başkalarından yardım alarak sürdürmek zorunda kalmaları hastalığın psikolojik olarak etkilenmelerine neden olmaktadır [9].

Son yıllarda yapılan çalışmalar incelendiğinde, hastalığın teşhisi ve takibinde özellikle ses kayıtlarından çıkarılan disfoninin (hastalıktan kaynaklanan ses tellerindeki bozulma) önemli rol oynadığını göstermektedir. Bu çalışmada konuşma sinyallerini kaydeden Ulusal Ses ve Konuşma Merkezi, Denver, Colorado işbirliği ile Oxford Üniversitesi'nden Max Little tarafından oluşturulan insanın biyomedikal konuşma sesinden (sürdürülmüş fonasyon test kayıtları) elde edilmiş özelliklerini içeren Parkinson veri tabanı kullanılmıştır.

Nörolojik bir hastalık olan Parkinson'un teşhisi özellikle hastalığın ilk evrelerinde zordur. Bu sebeple son yıllarda tıbbi karar verebilmeyi desteklemek amaçlı bilgisayar tabanlı çözüm araştırmaları oldukça artmıştır. Bu çalışmada, sağlıklı kişiler ile Parkinson hastalığı olan kişileri birbirinden ayırt etmek amacıyla sınıflandırmada tahmin doğruluğunu artırmak ve hesaplama yükünü azaltmak için ilgisiz ve gereksiz öznitelikleri ortadan kaldıran öznitelik seçme algoritmaları kullanılmıştır. Bu algoritmaların seçtiği öznitelikler sınıflandırma algoritmalarında kullanılarak hastalığın teşhisi için en iyi sonuca ulaşılması gerçekleştirilmiştir.

23 Parkinson hastasının ve 8 sağlıklı insanın 195 adet biyomedikal ses ölçümlerinden oluşan veri tabanında 22 adet öznitelik kullanılmıştır. Tüm öznitelikler, öznitelik seçme algoritmasının belirlediği öznitelikler ve ortaya konulan yeni yaklaşımla seçilen öznitelikler popüler sınıflandırma algoritmalarına (FFNN, RBNF, RF, DT ve SVM) ayrı ayrı tabi tutularak elde edilen sonuçların hesaplama sürelerinin ve doğruluk oranlarının kıyaslanması sağlanarak en iyi sonuçlar karşılaştırmalı analizlerle verilmiştir.

Bölüm-2'de tezde adı geçen Parkinson Hastalığının genel tanımı, erken teşhiste ses ve el yazısı gibi semptomların tedavi sürecindeki etkileri ve bu konuda bugüne kadar yapılmış olan çalışmalardan bahsedilmektedir.

Bölüm-3'te, bu tezde kullanılan, Parkinson veri setindeki özniteliklerin incelenmesi ve bu bağlamda kullanılan öznitelik seçme ve sınıflandırma algoritmaları anlatılmaktadır.

Bölüm-4'te, geliştirilen web arayüz programının tanıtılması ve bu uygulamayla yapılan çalışmalar anlatılmıştır. Parkinson hastalığına sahip olan hastalar ile kontrol gruplarının seçilen özniteliklerin farklı sınıflandırma algoritmalarıyla karşılaştırılmaları ve bu karşılaştırmalar sonunda elde edilen bulgular ile bu hastalıkla ilgili yapılan çalışmaların sonuçları incelenmiştir.

Son olarak Bölüm-5'te, elde edilen sonuçların, literatürdeki veriler ile olan uyumu ve tez paralelinde ileride yapılabilecek çalışmalardan bahsedilmektedir.

2. GENEL KISIMLAR

2.1. Hastalığın Tanımı

İlk kez 1817 yılında İngiliz hekim James Parkinson tarafından, “titrek felç” adıyla tanımlanmıştır. Bu hekimin adıyla anılan ve bugünkü anlamda felç özelliği taşımayan hastalık, iyi tanınması gereken en önemli parkinsonizm tipidir. Hastalığın temel belirtileri hareketlerin yavaşlığı, titreme, kasların sertliği ile duruş ve denge bozukluklarıdır. Klinik tanı için hareket yavaşlığına diğer üç ana belirtiden birinin eklenmiş olması yeterlidir [10, 14].

Hastalık 40-75 yaşları arasında, sıklıkla da 60 yaşın üzerinde başlar. Tüm Parkinson hastalarının sadece % 5 ila 10’unda hastalık başlangıç yaşı 20 ila 40 yaşları arasındadır. Genç yaş grubundaki hastalarda genetik nedenler araştırılmalıdır [10].

Parkinson hastalığı erkeklerde kadınlara oranla biraz daha sık görülür. Dünyanın her yerinde ve her türlü sosyoekonomik koşulda rastlanılan hastalığın görülme sıklığı çeşitli ülkelerde farklıdır. Libya’da Bingazi şehrinde yapılan bir araştırmada 100 000’de 31 kişide rastlanan hastalık, Hindistan’da Bombay şehrinde kapı-kapı dolaşarak yapılan ve tüm yaş gruplarını kapsayan bir çalışmada 100 000’de 328, İspanya’da yapılan benzer bir çalışmada ise 100 000’de 270 oranında saptanmıştır. Kapı-kapı dolaşarak 50 yaşın üzerindeki toplum kesimlerinde yapılan çalışmalarda bu oran yaklaşık 100 000’de 15-170 arasında bulunmuştur. Toplumda 65 yaş üzerinde her 100 kişiden birinin Parkinson hastası olduğu kabul edilmektedir [10].

Hastalık genellikle sinsi başlar ve belirtileri yıllar içinde, son derece yavaş ama giderek artan biçimde ilerler, öyle ki hastalar çoğu zaman hastalığın başlangıç tarihini kesin olarak söyleyemezler. Hastalar ilk belirtinin farkına vardıkları zaman, hastalığın bazı belirtileri uzun zaman önce başlamış olabilir. Bir elinde titreme yakınmasıyla başvuran bir hastanın 5-6 yıl öncesine ait çekilmiş video filmlerinde yürürken bir kolunu sallamadığı fark edilebilir veya bazen hastanın eski fotoğraflarında öne eğik duruş özelliği dikkati çekebilir [10].

Parkinson hastalarının büyük çoğunluğunda sıklıkla ilk belirti bir el parmağında veya elde titremedir. Daha seyrek olarak ayak titremesiyle başlayabilir. Kimi hastada ise yazı yazarken harflerde küçülme veya yüzünde donuk ifade ilk belirtiyi oluşturur. Özellikle genç yaştaki hastalarda ilk belirti bir ayağın içe doğru dönük halde kasılmasıdır (distoni). Hastaların hemen tümünde belirtiler tek bir beden yarısında ortaya çıkar ve zamanla karşı beden yarısında daha hafif olmak üzere kendini gösterir [10].

Hastaların başvuru yakınmaları arasında yürüyüş, rijidite ve postür ile ilgili refleks mekanizmalarını etkilenmesine bağlı olarak sandalyeden zor kalkma, yatakta dönme veya dengeyi sağlama güçlüğü yer alabilir. Bu belirtilerle beraber bu tezde de yer alan fonasyondaki değişimler, volüm kayıpları ve disfoni yani sesin yapısının bozulması gibi belirtilerde mevcuttur [10-12].

Hastalığın ilerleme hızı ile belirtilerin türü ve şiddeti hastadan hastaya değişiklik gösterecek şekilde farklıdır [10].

2.2. Parkinson Hastalığının Klinik Özellikleri

Parkinson hastalığının temel belirtileri olan titreme (Tremor), uzuvların sertliği (Rijidite), hareketlerin yavaşlığı (Akinezi-Bradikinezi) ve duruş bozuklukları (Postural İnstabilite) yanı sıra, bu belirtilerin birkaçının bir arada sorumlu olduğu el yazısının bozulması, yürüme bozukluğu, yürürken kilitlenme, düşmeler vücuttaki eğiklik ve motor fonksiyonlarında donma gibi özellikler Parkinson hastalığında sıklıkla karşılaşılan belirtilerdendir [10].

Hastalık seyri boyunca hastalığın ilerleme veya gerileme gibi durumlarını kontrol etmek için bazı yöntemler kullanılmaktadır. İlk olarak hastalar Hoehn-Yavr adı verilen ve semptomların tek veya çift oluşu, düşmelerin ortaya çıkması, denge sorunları, destek gereksinimi veya yatağa bağımlı hale gelme gibi aşamalara göre değerlendirilirler. Ayrıca hastalık semptomlarının artışı veya azaldığını değerlendirmek ve hastaların tedaviye verdikleri tepkiyi izlemek için mental, ses ve motor fonksiyon testi yani “Birleştirilmiş Parkinson Hastalığı Değerlendirme Ölçeği (UPDRS)” kullanılır.

2.2.1. Titreme (Tremor)

Parkinson hastalığının temel belirtilerinden en belirginini ve en erken ortaya çıkan olan genellikle hastanın doktora en sık başvurma nedenidir. Parkinson hastalarının yaklaşık % 80’inde titreme ortaya çıkmaktadır. Titreme sıklıkla bir taraftaki elde, bazen de bir ayakta ortaya çıkar. Titreme tek bir parmağa sınırlı kalabildiği gibi bazen dili, dudakları veya çeneyi de etkileyebilir. Ancak Parkinson hastalığı baş veya ses titremesine yol açmaz. Bazı hastalarda, bir elde başlayan titreme kola ve aynı taraftaki bacağı daha sonra da karşı kola yayılabilir [10].

Titreme başparmak ve işaret parmakların ileri-geri hareketleri ve elin bozuk para sayma ya da bir çakıl taşını başparmak ve işaret parmak arasında yuvarlama hareketi şeklinde olabilir. Titreme ayakta ortaya çıktığı zaman ayağı pedala basma-kaldırma hareketini andırır. Tremor diğer kardinal belirtilere göre daha yavaş bir ilerleme gösterir [10].

Düzenli ve belli bir hızda olan titreme saniyede 5-6 vurumludur. Diğer hastalıklarda görülebilen titremelere benzemeksizin, Parkinson hastalığında etkilenmiş olan el veya ayak dinlenme sırasında titrer. Titreme uyku sırasında ve o uzvun harekete başlamasıyla kaybolur [10].

Bazı hastalarda bu titreşimler vücudun içerisinde yani gözle görülmeyen biçimde gerçekleşir. Bu durum “internal tremor” olarak adlandırılır. Hastalarda ayrıca postural ve kinetik tipte aksiyon tremorları gözlenebilir, ancak hastalıkta ön planda değildirler [11].

2.2.2. Kaslarda sertleşme (Rijidite)

Rijidite, hastanın boyun veya gövdesinin pasif hareketi esnasında hekimin her yöne ve tüm hareket genliği boyunca hissettiği dirence verilen addır [12]. Bazı hastalar uzuvlarda sertlik hissinden yakınır. Bununla birlikte kas sertliği çoğu kez hastanın bir yakınması olmayıp hekimin fizik muayenede pasif harekete karşı olan bir direncin varlığını saptaması ile tanınır. Hekim hastaya gevşemesini söyleyerek, hastanın uzuvlarını eklem yerlerinden bir çok kez nazıkçe gerer ve bükerek ve bu pasif harekete karşı eklem çevresinde direnç arar. Böyle pasif harekete karşı sürekli bir direnç bulunması kaslardaki sertleşmeyi (rijidite) yansıtır. Normalde kasların dinlenme halinde yumuşak ve gevşek olmaları gerekirken,

rijidite varlığında dinlenme halinde bile sabit biçimde gergin ve elle hissedilebilen belli bir sertlikte oldukları görülür. Parkinson hastalığında kaslarda sertleşme en sık el, ayak bileği, dirsek veya diz gibi eklemlerin muayenesinde saptanır [10].

Bazen kas sertliği hekim tarafından eklemden sanki “dişli çark” takılması varmış gibi hissedilir. Hastalar kas sertliğini yorgunluk, batma hissi, ağrı veya kramp şeklinde hissedebilirler [10].

Rijidite istemli hareket hızını sınırladığı halde, rijiditesi olan bazı hastalar motor işlevlerini rahat sürdürebilirler ve bradikinezi, hastanın özürüllüğünde rijiditeden daha belirleyici bir rol oynar [11].

2.2.3. Hareketlerde yavaşlama (Akinezi/Bradikinezi)

“Hipokinezi” yapılan hareketlerin azalması yani hareketlerde fakirleşme anlamındadır. “Akinezi” ise hareketsizlik anlamı taşır ve genellikle hastalığın ilerlemiş olduğu dönemlerde ortaya çıkar. Parkinson hastalığının belki de özürüllük yaratan en temel belirtisi olan hareketlerdeki yavaşlama yani “bradikinezi”, erken veya geç olarak her hastada gelişir. Hareket yavaşlığı günlük yaşamdaki faaliyetlerin tümünün belli bir yavaşlıkta olmasına yol açar. Hareketlerin ardı sıra tekrarı ve eklemlerin hareket açıklığı azalmıştır. Hastaların basit günlük işlerini yapma sırasında, örneğin düğme ilikleme, kravat ve ayakkabı bağlama, yazı yazma ve çatal-bıçak kullanma gibi incelik isteyen işlerde başlangıçta hafif derecede hissettikleri güçlük giderek artar. Zamanla istemli hareketlerin çoğunun yapılmasında, örneğin yemek yerken ve çiğnerken, alçak bir koltuktan doğrulurken, otomobile binerken ve inerken, yatakta bir taraftan diğer tarafa dönerken zorlanmalar dikkati çeker. Yukarıda sözü edilen istemli hareketlerin yavaşlamasının yanı sıra, gözleri kırpmak ve yürürken kolları sallamak gibi farkında olmadan otomatik olarak yaptığımız hareketler de azalır veya kaybolur [10].

Hareket yavaşlığı belirgin olsa da hastaların kas kuvveti normaldir. Hastanın bu yöndeki yakınması genel bir yorgunluk hali, örneğin yürürken ya da dış fırçalarken yapılması gereken ardı sıra hareketler sırasında uzuvlarda hissettiği tutukluktur. Hareketlerdeki bu tür yavaşlık zamanla hastaları başkalarına bağımlı hale getirebilir. Hareketleri belirgin ölçüde yavaşlamış olan bir hastada titreme veya rijidite bulunmayabilir [10].

Bu durumdaki Parkinson hastaları uzun süre izlendiğinde, hareket yapma yeteneğini gözle görülür derecede yitirdikleri görülür. Hareketlerdeki yavaşlama sonucunda, hastaların yüz ifadesinde donukluk ve el yazılarında küçülme dikkati çeker. Hastaların konuşmaları monoton ve alçak seste yani hipofoniktir. Bazı basit kelimelerin tam olarak ifade edilemediği, kelimeler arasında duraklama yapılmadan, bütün kelimelerin bir arada okunduğu taşifemi adı verilen bir konuşma görülür [10,13].

2.2.4. Duruş ve denge bozuklukları (Postüral Instabilite)

Parkinson hastalarında duruş ve dengeyi koruma yeteneği zamanla bozulabilir. Birçok Parkinson hastası öne doğru hafifçe eğik durma eğilimindedir. Bu duruş biçimi ayakta dururken ve yürürken daha belirgin olur, bazı hastalarda ise çok çarpıcıdır. Çok ender olarak, belden yaklaşık 90° öne eğilme hali gelişebilir (kamptokormi), ancak böyle hastalar sırt üstü uzanabilirler, yani omurgada kalıcı bir deformiteye bağlı değildir. Kimi hastanın bir yana doğru hafifçe yamuk durduğu görülür ve bu durum özellikle oturma halinde dikkati çeker. Kimi hasta otururken bir ayağını içe doğru dönük tutar. Duruştaki diğer bir özellik hastanın yürürken ya da ayakta dururken bir veya her iki kolunu dirsekten, ve bacaklarını dizlerden ve kalçadan bükme eğilimidir. Sürekli olarak öne eğik durmak sırt kaslarının mekanik olarak gerilmesine yol açacağından hastalarda sırt ve bel ağrılarına yol açabilir. Duruş bozukluğunu fark eden hasta bunu en aza indirmek üzere egzersizler yapmaya başlarsa ve yürürken kollarını sallamak için gayret sarf ederse çok yarar görecektir. Duruş bozukluğu zamanla ayakta dururken, yürürken ve dönüş yaparken veya sandalyeden kalkarken ve ileri doğru uzanırken emniyetsiz hissetmeye, denge kaybına ve düşmelere yol açabilir [10].

Öne eğik gövdenin denge merkezini koruyabilmek ve düşmemek için adımlarını öne atan hasta arkadan itiliyormuşçasına gitgide daha hızlı yürür. Yürüyüşte meydana gelen bu değişikliğe “festination” adı verilir [11].

2.2.5. Parkinson hastalığında konuşma bozukluğu

Parkinson hastalarının ortalama %80’inde hastalığın ilerleyen safhalarında konuşma ile ilgili sorun yaşayacakları bilinmektedir. Konuşma sorunları nadiren hastalığın başlangıcında ortaya çıkar. Hastalıkları uzun süreden beri mevcut olsa bile, Parkinson hastalarının yaklaşık yarısının konuşmalarında bir değişiklik görülmez. Bazı hastalarda,

zaman içinde konuşma hastalık için tipik olan bir değişikliğe uğrar. İlk değişiklik genellikle yumuşak tonda bir konuşma şeklidir. Hastaların konuşmalarında her heceye aynı tonda vurgulama söz konusudur. Bazı hastalar hem monoton hem de yumuşak ve hızlı tonda konuşurlar. Kelimeler arasında alışılmış aralar yoktur ve kelimeler iç içe girermişçesine konuşma anlaşılmaz hale gelebilir. Tanımlanan bu konuşma özelliği Parkinson hastalığı için tipiktir. Bunun aksine bazı hastalar hızlı değil, yavaş konuşurlar, ses tonları giderek zayıflar. Ender görülen bir durum da belli bir hecenin kelimenin tam ortasında veya sonunda bir çok kez tekrar edilmesidir. Bunların dışında konuşurken yüz ifadesini veren hareketlerin ve yardımcı açıklayıcı el hareketlerinin yapılamaması da iletişim yeteneğini zorlaştırır. Hastanın anlatılana ilgisiz kaldığı veya anlatılanı tam anlamadığı izlenimi ortaya çıkar [10].

Bazı hastaların telefondaki konuşmaları anlaşılabilir. Böyle hastalar konuşmalarının yumuşak ve başkalarının duyamayacakları kadar belirsiz olduğunun farkında olmayabilirler. Konuşmanın normal tonunun kaybı bazıları için rahatsızlık verici olmazken, diğerleri için büyük önem taşır. Bu durum hastaların yaşamlarına ve işlerine göre değişiklik gösterir. Örneğin öğretmen, avukat, tiyatro oyuncusu olan ya da konuşma veya okumaya ilişkin görevi olan hastalar, konuşmalarındaki değişikliği birçok hastadan daha erken fark ederler [10].

Dizartri Parkinson hastalığı ile ilişkili başka bir konuşma sorunudur. Bu konuşma sorunu, yavaş, zayıf ya da tutarsız konuşma gibi şekillerde kendini göstermektedir. Aynı zamanda sesin tonu ve şiddeti de dizartriden etkilenir ve sonunda konuşma anlaşılmaz hale gelir. Konuşma kaslarının zayıflaması ve Parkinson rahatsızlığı nedeniyle bu kaslardaki koordinasyon problemleri hastada dizartrinin ortaya çıkmasına sebep olmaktadır. Dizartrinin şiddeti hastadan hastaya değişebilir. Bazı hastalarda bu sorunun çok hafif etkileri olsa da, bazı hastalarda problem çok şiddetli biçimde olabilmektedir [15].

2.3. Parkinson Hastalığında Bilgisayarlı Tabanlı Teşhis Çalışmaları

Bugüne kadar Parkinson hastalığının teşhis ve tedavisinde tıbbi muayeneler yapılarak tanı konulmuştur. Hastalığın belirtileri fiziksel olarak hastayı etkilediğinden teşhisin bilgisayar ile gerçekleştirip gerçekleştirilmeyeceği birçok çalışmada araştırma konusu olmuştur. Teknolojik gelişmelerin arttığı günümüzde Parkinson hastalığını uzaktan teşhis için

araştırmacılar bilimsel çalışmalar yapmaktadır. Bu çalışmalar özellikle konuşma üzerinde gerçekleştirilmiştir. Aşağıda bu konularla ilgili yapılan örnek çalışmalardan anlatılmaktadır [14].

Çalışmalar da hastalığın en önemli belirtileri olan konuşma bozukluğu, kısık ses ve donuk konuşma kullanılarak sonuçlar elde edilmiştir. Bu çalışmalardan Little ve arkadaşlarının [16], Parkinson hastalığıyla ortaya çıkan disfoniye ölçerek hastalığın şiddetini derecelendirmeyi hedefleyen çalışmalarıdır. Yapmış oldukları çalışmada 23 tanesi Parkinson hastası olmak 31 denek kullanılmıştır. Bu kişilere mikrofon aracılığıyla konuşma testlerini uygulayıp, sesteki titreşim, ton, ses düzeyi kişiye has özellikleri çıkartıp, hangisinin bu hastalıkta kullanılabileceğini belirlemişlerdir. Bu özellikler kullanılarak kişinin hastalığının hangi aşamada olduğu hedeflemişler ve %94 gibi bir başarılı sonucuna ulaşmışlardır.

Tsanas ve arkadaşları kişinin hastalığının ilerleme ya da gerileme değişim oranı tahmin etmeyi hedefleyen çalışmalarıyla [17], 42 Parkinson hastasından mikrofon aracılığıyla topladıkları yaklaşık 6000 ses kaydıyla konuşmadaki ses özelliklerini sinyal işleme algoritmalarıyla karakterize ederek, kullanılabilir özelliklerini ayırt etmişlerdir.

Revett ve arkadaşları [18], Little ve arkadaşlarının çıkardıkları ses örneklerinden elde ettikleri çeşitli frekansları kullanarak kurallar çıkarmışlardır. Daha sonra bu kurallar, filtreleme algoritmalarıyla elemiş ve seçilen kurallar sınıflandırma algoritmasına verilerek %90 başarı sağlanmıştır. Bu çalışma aslında yöntemi denemek için yapılmıştır. Parkinson benzeri hastalıklarda elde edilen verim, çalışmanın bir sonraki aşamasını oluşturmuştur.

Tsanas ve arkadaşları [19], yaptıkları diğer çalışmada ses sinyallerini logaritmik olarak dönüştürmüşlerdir. Bütün testler hastanın evinde kendi yönetiminde gerçekleştirilerek elde edilecek sonuçlarla doktorların hastalığın tespitinde sıklıkla kullandığı Birleşik Parkinson Hastalığı Değerlendirme Ölçeğiyle eşleştirmişlerdir. Ham ses sinyallerine kıyasla logaritmik olarak dönüştürülmüş ses sinyallerinin çok iyi klinik bilgiler sağladığı görülmüştür [14].

Bu tez çalışmasında Parkinson hastalığının teşhisi için öznel ve sınıflandırma algoritmalarıyla elde edilen sonuçlar, çalışmalardaki sonuçlarla karşılaştırılmıştır.

3. VERİ SETİ VE ALGORİTMALAR

Bu çalışmada kullanılan veri seti, konuşma sinyallerini kaydeden Ulusal Ses ve Konuşma Merkezi, Denver, Colorado işbirliği ile Oxford Üniversitesi'nden Max Little tarafından oluşturulan ve insanın biyomedikal konuşma sesinden (sürdürülmüş fonasyon test kayıtları) elde edilmiş bilgilerdir. UCI'den (Machine Learning Repository) alınmıştır [20]. 23 Parkinson hastasının ve 8 sağlıklı insanın 195 adet biyomedikal ses ölçümlerinden oluşmaktadır [21].

3.1. Parkinson Hastalığı Veri Seti (PHV)

PHV öznitelikleri şunlardır; ortalama, maksimum ve minimum ses temel frekansı, temel frekanstaki düzensizlik ölçümleri (Jitter(%), Jitter (mutlak), Jitter :RAP, Jitter: PPQ ve Jitter: DDP), genlik düzensizlik ölçümleri (Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer: APQ, Shimmer: DDA), sesteki ton bileşenlerinin gürültüye oranının ölçümleri (HNR ve NHR), iki adet doğrusal olmayan dinamik komplekslik ölçümleri (RPDE ve D2), üç adet doğrusal olmayan temel frekans değişimi ölçümleri (PPE, Spread1 ve Spread2) ve üstel kesirli sinyal (DFA) ölçümü içermektedir. Ayrıca PHV, Parkinson hastaları için 1, sağlıklı kişiler için 0 olarak tanımlı bir durum sütunu içermektedir. Çalışmada kullanılan Parkinson veri setindeki öznitelikler Çizelge 3.1'de verilmiştir.

Çizelge 3.1. Parkinson hastalığı veri seti öznitelikleri

No	Öznitelik	No	Öznitelik	No	Öznitelik
1	MDVP:F0(Hz)	9	MDVP:Shimmer	17	RPDE
2	MDVP:F1(Hz)	10	MDVP:Shimmer (dB)	18	DFA
3	MDVP:F2(Hz)	11	Shimmer:APQ3	19	spread1
4	MDVP:Jitter(%)	12	Shimmer:APQ5	20	spread2
5	MDVP:Jitter(Abs)	13	MDVP:APQ	21	D2
6	MDVP:RAP	14	Shimmer:DDA	22	PPE
7	MDVP:PPQ	15	NHR	23	Status
8	Jitter:DDP	16	HNR		

3.1.1. Jitter sınıfı öznitelikler

İdeal olarak, düz fonasyon sırasında, temel frekansın hiç değişmemesi gerekir. Fakat pratikte fonasyonla ilgili organlar bunu sağlayacak kadar mükemmel çalışmazlar ve peş

peşe gelen periyotlar arasında küçük farklar ortaya çıkar [22]. Temel frekanstaki istem dışı ortaya çıkan bu düzensizliğe frekans pertürbasyonu veya jitter adı verilir. Jitter grubundan 5 farklı öznelilik çıkarılabilmektedir.

- Jitter (local): Mutlak jitterin temel frekansa bağlı olarak değişiklik göstermesi sakıncasını ortadan kaldırmak için, mutlak jitterin ortalama periyoda bölünmesi ile elde edilir.

$$Jitter(local) = \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} |P_i - P_{i+1}|}{\frac{1}{n} \sum_{i=1}^n P_i} \quad (3.1)$$

- Jitter (local, absolute): Analiz edilen ses örneğindeki her periyodun, kendinden sonraki periyotla farkının mutlak değerinin ortalamasına eşittir. Birimi μs 'dir.

$$Jitter(local, absolute) = \frac{1}{n-1} \sum_{i=1}^{n-1} |P_i - P_{i+1}| \quad (3.2)$$

- Jitter (rap): İsteğe bağlı ya da ses tremoru kişinin sesini aynı perde de tutamaması gibi isteğe bağlı olmayan temel frekans değişikliklerinin jitter değerlerini etkilememesi için kullanılan bir jitter hesaplama yöntemidir. Burada üç periyotluk bir düzeltme faktörü uygulanır. Bir periyotla kendinden sonra gelen periyot arasındaki fark yerine ardışık üç periyodun ortalaması ile bu üç periyodun ortasında yer alan periyot arasındaki fark dikkate alınır.

$$Jitter(rap) = \frac{\frac{1}{n-2} \left[\sum_{i=1}^{n-1} \left| \frac{P_{i-1} + P_i + P_{i+1}}{3} - P_i \right| \right]}{\frac{1}{n} \sum_{i=1}^n P_i} \quad (3.3)$$

- Jitter (ppq5): RAP'dan farklı olarak 3 yerine 5 periyotluk düzeltme faktörü uygulanır.

$$Jitter(ppq5) = \frac{\frac{1}{n-2} \left[\sum_{i=1}^{n-1} \left| \frac{P_{i-1} + P_i + P_{i+1} + P_{i+2} + P_{i+3}}{5} - P_i \right| \right]}{\frac{1}{n} \sum_{i=1}^n P_i} \quad (3.4)$$

- Jitter (ddp): Ardışık periyotlar arasındaki ardışık farklar arasındaki mutlak farkın ortalaması bulunur. Bu değer ortalama periyoda bölünür.

$$Jitter(ddp) = \frac{\frac{1}{n-2} \left[\sum_{i=1}^{n-1} |F(P_{i-1}, P_i) - F(P_{i+1}, P_{i+2})| \right]}{\frac{1}{n} \sum_{i=1}^n P_i} \quad (3.5)$$

3.1.2. Shimmer sınıfı öznitelikler

Temel frekans pertürbasyonunda olduğu gibi, bunda da ses işaretlerindeki çok kısa süreli genlik değişiklikleri ölçülür. Shimmer adı verilen genlik pertürbasyonu dB veya % olarak ifade edilir.

- Shimmer (local): Her periyodun kendinden sonraki periyotla arasındaki genlik farkının mutlak değerinin ortalaması ortalama periyot şiddetine bölünerek elde edilir.

$$Shimmer(local) = \frac{\frac{1}{n-1} \sum_{i=1}^{n-1} |A_i - A_{i+1}|}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (3.6)$$

- Shimmer (local, dB): Her periyodun tepe genliği bir sonraki periyodun tepe genliği ile karşılaştırılır. Bu şekilde dB cinsinden shimmer hesaplanır. Matematiksel olarak denklem 3.7'deki şekilde ifade edilir.

$$Shimmer(local, dB) = \frac{1}{n-1} \sum_{i=1}^{n-1} \left| 20 \log_{10} \left(\frac{A_i}{A_{i+1}} \right) \right| \quad (3.7)$$

- Shimmer (apq3): Three-point Amplitude Perturbation Quotient (3 Noktalı Genlik Değişim Katsayısı) olarak bilinmektedir. Bir periyodun genliği ile o periyodun kendisinin ve komşu iki periyodunun genliklerinin birleşiminden elde edilen ortalama genlik arasındaki mutlak farkların ortalaması bulunur. Bu değer ortalama genliğe bölünür.

$$Shimmer(apq3) = \frac{\frac{1}{n-2} \left[\sum_{i=2}^{n-1} \left| \frac{A_{i-1} + A_i + A_{i+1}}{3} - A_i \right| \right]}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (3.8)$$

- Shimmer (apq5): Five-point Amplitude Perturbation Quotient (5 Noktalı Genlik Değişim Katsayısı) olarak bilinmektedir. Bir periyodun genliği ile o periyodun kendisinin ve komşu dört periyodun genliklerinin (sağdan ve soldan 2'şer adet) birleşiminden elde edilen ortalama genlik arasındaki mutlak farkların ortalaması bulunur. Bu fark ortalama genliğe bölünür.

$$Shimmer(apq5) = \frac{\frac{1}{n-4} \left[\sum_{i=3}^{n-2} \left| \frac{A_{i-1} + A_i + A_{i+1} + A_{i+2} + A_{i+3}}{5} - A_i \right| \right]}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (3.9)$$

- Shimmer (apq11): Eleven-point Amplitude Perturbation Quotient (11 Noktalı Genlik Değişim Katsayısı) olarak bilinmektedir. Bir periyodun genliği ile o periyodun kendisinin ve komşu on periyodun genliklerinin (sağdan ve soldan 5'er adet) birleşiminden elde edilen ortalama genlik arasındaki mutlak farkların ortalaması bulunur. Bu fark ortalama genliğe bölünür.

$$Shimmer(apq11) = \frac{\frac{1}{n-10} \left[\sum_{i=6}^{n-5} \left| \frac{A_{i-5} + A_{i-4} + \dots + A_{i+4} + A_{i+5}}{11} - A_i \right| \right]}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (3.10)$$

- Shimmer (ddp): Ardışık periyotların genlikleri arasındaki ardışık farklar arasındaki mutlak farkların ortalaması bulunur.

$$Shimmer(ddp) = \frac{1}{n-2} \left[\sum_{i=1}^{n-1} F(|P_{i-1}, P_i) - F(P_{i+1}, P_{i+2})| \right] \quad (3.11)$$

3.1.3. Diğer öznitelikler

Ses izgesinde yer alan, temel frekansın katlarından oluşan harmonikler ve harmonikler dışındaki gürültü sesleri ile ilgili parametrelerdir.

- Harmonik-Gürültü Oranı(HNR): Harmonik enerjisinin gürültü enerjisine oranıdır.
- Normalize Gürültü Enerjisi (NNE): Harmonik enerjinin toplam ses enerjisinden çıkarılması ile elde edilir.

	C	D	E	F	G	H	I	J	K	L	M	N
	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP
1	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.1
2	148.65	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	0.626	0.03134	0.04518	0.1
3	131.111	111.555	0.0105	0.00009	0.00544	0.00781	0.01633	0.05233	0.482	0.02757	0.03858	0.1
4	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	0.517	0.02924	0.04005	0.1
5	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	0.584	0.0349	0.04825	0.1
6	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.1
7	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.1
8	113.84	104.315	0.0029	0.00003	0.00144	0.00182	0.00431	0.01567	0.134	0.00829	0.00946	0.1
9	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093	0.191	0.01073	0.01277	0.1
10	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.1
11	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.197	0.01079	0.01342	0.1
12	115.871	86.292	0.0054	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01424	0.01641	0.1
13	159.866	131.276	0.00293	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.1
14	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.1
15	163.305	75.836	0.00294	0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.1
16	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00772	0.00888	0.1
17	349.259	82.764	0.00544	0.00004	0.00211	0.00292	0.00632	0.02047	0.192	0.00969	0.012	0.1
18	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.1
19	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.1
20	189.398	142.822	0.00768	0.00005	0.00372	0.00399	0.01116	0.03995	0.348	0.01721	0.02374	0.1
21	165.738	65.782	0.0084	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.1
22	172.86	78.128	0.0048	0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.1
23	193.221	79.068	0.00442	0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.1

Resim 3.1. Parkinson veri tablosu

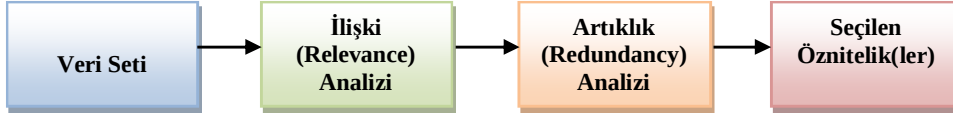
3.2. Öznitelik Seçme Algoritmaları

Sinyal işleme uygulamalarında kullanılacak olan sınıflandırma yöntemine girdi olarak genellikle veri setinden çıkarılmış olan öznitelikler kullanılır. Bazı uygulamalarda öznitelik sayısı birkaç tane ile sınırlı kalırken, bazılarında da çok fazla miktarda olabilir. Her sinyal için çıkarılan öznitelikler bir öznitelik matrisinde saklanır. Dolayısı ile hem sinyal büyüklüğü hem de öznitelik sayısı matrisin boyutunu yani işlem süresini etkileyecektir. Çıkarılmış olan özniteliklerin sinyalleri ve sinyallere karşılık gelen sınıfları ne kadar iyi temsil ettiği önemlidir. Sınıflar arasında daha iyi ayrım yapabilen öznitelikler, yüksek seviyeli öznitelikler olarak adlandırılır ve sınıflandırıcının performansı açısından daha önemlidir. Tüm öznitelikleri kullanmak yerine sadece bazı yüksek seviyeli özniteliklerin kullanılması yapılan işlem sayısını azaltabilir, aynı zamanda gürültülü özniteliklerin temizlenmesi ile performans da artışları da sağlanabilir. Öznitelik seçme algoritmaları ilgisiz ve gereksiz öznitelikleri elemek için kullanılır [23].

Bulgular bölümünde aşağıda anlatılan öznitelik seçme algoritmalarının veri setine uygulanma sonuçları verilmiştir.

3.2.1. Hızlı korelasyon tabanlı filtre-Fast correlation based filter (FCBF)

FCBF, L. Yu and H. Liu tarafından 2003 yılında ortaya konulan bir öznitelik seçme algoritmasıdır [24]. FCBF, öznitelikler arası ilişki (relevance) ve artıklık (redundancy) kriterlerini temel alarak geliştirilmiştir. FCBF algoritması iki aşamada seçim işlemini gerçekleştirir [23].



Şekil 3.1. FCBF öznitelik seçme algoritması

İlişki Analizi (Relevance Analysis):

Öznitelikler ile sınıf etiketleri, H , arasındaki ilişki, diğer bir ifadeyle, her bir özneliğin sınıf etiketleri ile olan korelasyonu hesaplanır. Korelasyon değeri belli bir eşik değerine eşit ya da bu değerin altında olan öznitelikler elenmektedir. Hesaplanan doğrusal olmayan korelasyon değerinin her zaman pozitif olması sebebiyle, 0 eşik değeri durumunda, tüm öznitelikler, sınıf etiketleri ile en ilişkiliden en ilişkisize doğru sıralanmış olur. 0'dan büyük bir eşik değerinin kullanılması bazı özniteliklerin ilk etapta elenmesine yol açacağından algoritmanın hızlanmasında etkili olmaktadır [25]. A ve B rastlantı değişkenleri için doğrusal korelasyon katsayısı:

$$r = \frac{\sum_i (a_i - \bar{a}_i)(b_i - \bar{b}_i)}{\sqrt{\sum_i (a_i - \bar{a}_i)^2} \sqrt{\sum_i (b_i - \bar{b}_i)^2}} \quad (3.12)$$

FCBF'de her bir öznitelik ile sınıf etiketleri arasındaki korelasyonu ölçmek için Simetrik Belirsizlik (SU) değerleri hesaplanır. SU entropi tabanlı doğrusal olmayan bir korelasyon ölçütüdür. A ve B değişkenleri için SU değeri Denklem 3.13'deki gibi hesaplanır. Burada A ve B herhangi iki özneliği ya da bir öznitelik ve bir sınıf etiketi çiftini ifade etmektedir [26].

$$SU(A, B) = 2[IG(A|B)/H(A) + H(B)] \quad (3.13)$$

$$IG(A|B) = H(A) - H(A/B) \quad (3.14)$$

$$H(A) = -\sum_i P(a_i) \log_2(P(a_i)) \quad (3.15)$$

Burada $IG(A | B)$ A' nın B değerlerine bağlı bilgi kazancını ifade eder. $H(A)$ ve $H(B)$, A ve B değişkenlerinin entropileridir. $P(a_i)$ A değişkeninin olasılık değeridir ve A' nın B değerlerine bağlı entropisi olan $(H(A | B))$ denklemde görüldüğü gibi hesaplanır [27].

$$H(A|B) = -\sum_j P(b_j) \sum_i P(a_i | b_j) \log_2(P(a_i b_j)) \quad (3.16)$$

Artıklık Analizi (Redundancy analysis):

Bu aşamada, kalan öznitelikler arasından sınıf etiketleri ile en ilişkili olanlar kullanılarak artık öznitelikler saptanmaya çalışılır. Tespit edilen artık öznitelikler elenir. İki aşama sonucunda kalan öznitelikler, sınıf etiketleri ile en ilişkili, kendi aralarında ise en az artıklığa sahip öznitelik kümesi olarak değerlendirilmektedir.

Artıklık iki özneliğin birbiriyle gereğinden fazla ilişkili olması ya da iki öznitelik arasındaki korelasyon değerinin yüksek olması olarak değerlendirilebilir. Algoritmanın artık öznitelikleri eleme aşaması olan ikinci etabın çalışması şu şekildedir: Örneğin öznitelik a'nın sınıf etiketleri ile korelasyonu öznitelik b'den yüksektir. Eğer öznitelik a ile öznitelik b arasındaki korelasyon değeri, öznitelik b ile sınıf etiketleri vektörü arasındaki korelasyon değerinden yüksek ya da bu değere eşit ise öznitelik b artık bir öznitelik olarak tanımlanır ve öznitelik a'nın varlığında öznitelik b'ye gerek duyulmadığı için öznitelik kümesinden atılır [25].

$$SU_{a,b} \geq SU_{a,h} \rightarrow Elenen F_b \quad (3.17)$$

FCBF'nin verilen bilgiler doğrultusunda algoritması aşağıdaki gibidir. [23].

giriş: $S(F_1, F_2, \dots, F_3, C)$ // eğitim veri seti
 δ // eşik değerinin belirlenmesi
çıkış: S_{best} // seçilen bir alt kümesi

```

1: begin
// 1. Adım: İlişki Analizi
2:  $S' \leftarrow S$ 
3: for  $i = 1$  to  $N$  do
4:   Calculate  $SU_{i,c}$  for  $F_i$ ;
5:   if ( $SU_{i,c} > \delta$ )
6:     append  $F_i$  to  $S'$ ;
7:   end;
8: end;
9: Order  $S'$  in descending  $SU_{i,c}$  value;
// 2. Adım: Artıklık Analizi
10:  $F_j = \text{getFirstElement}(S')$ ;
11: do begin
12:    $F_i = \text{getNextElement}(S', F_j)$ ;
13:   if ( $F_i \neq \text{NULL}$ )
14:     do begin
15:       if ( $SU_{i,j} \geq SU_{i,c}$ )
16:         remove  $F_i$  from  $S'$ ;
17:          $F_i = \text{getNextElement}(S', F_i)$ ;
18:       end until ( $F_i == \text{NULL}$ );
19:    $F_j = \text{getNextElement}(S', F_j)$ ;
20: end until ( $F_j == \text{NULL}$ );
21:  $S_{best} = S'$ ;
22: end;

```

FCBF algoritmasının diğer öznitelik seçme algoritmalarına göre en avantajlı yönlerinden biri hızlı çalışmasıdır. Bir filtreleme yöntemi olan FCBF algoritması, öznitelikler ile sınıf etiketleri arasındaki ya da iki öznitelik arasındaki korelasyon değerlerini hesaplayarak bir özniteliliğin, sınıflandırmadaki değerine sınıflandırıcı algoritmalarının sonucuna gerek duymadan ortaya koyabilmektedir. Bu yönüyle diğer öznitelik seçme algoritmalarına göre hız açısından büyük avantaj sağlamaktadır [25].

3.2.2. Minimum fazlalık maksimum ilişki-minimum Redundancy Maximum Relevance (mRMR)

mRMR (Minimum Redundancy Maximum Relevance) algoritması, sınıf etiketleriyle en ilişkili (relevance) öznitelikleri seçmeye çalışırken aynı zamanda seçilen öznitelikler arasındaki artıklığı (redundancy), fazlalığı minimize etmeye çalışan bir filtreleme yöntemidir [28]. Algoritma her bir özniteliği ve sınıf etiketleri vektörünü ayrık bir rastlantı değişkeni olarak görür ve iki öznitelik arasındaki ya da bir öznitelik ile sınıf etiketleri vektörü arasındaki benzerlik seviyesini ölçmek için aralarındaki ortak bilgiyi (mutual information), $I(x,y)$, kullanır.

$$I(X,Y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (3.18)$$

x ve y öznitelikleri için, $p(x_i)$ ve $p(y_j)$ marjinal olasılık fonksiyonu ve $p(x_i, y_j)$ bağlantı olasılık dağılımı olmak üzere $I(x,y)$, x ve y 'nin karşılıklı bilgi miktarıdır (mutual information). Ortak bilgi iki rastlantı değişkeninin tamamıyla bağımsız olduğu durumlarda "0" değerini almaktadır. Bu değer negatif olamaz ve simetriktir ($I(X,Y) \geq 0, I(X,Y) = I(Y,X)$). Gösterimde kolaylık sağlamak amacıyla, her bir öznitelik $f(i)$, n tane gözlemin bir araya gelerek oluşturduğu bir vektör olarak tanımlanacaktır ($f_i = [f_i^1, f_i^2, f_i^3, \dots, f_i^n]$). Bir öznitelik vektörünün reel değerlerini barındıran f_i , vektörü ise F_i ayrık rastlantı değişkeninin bir örneği olarak ele alınacak ve öznitelik i ve j arasındaki ortak bilgi, $I(F_i, F_j)$ bu ayrık rastlantı değişkenleri üzerinden hesaplanacaktır. Burada $i, j = 1, 2, \dots, d$, d ise veri kümesinin boyut sayısını diğer bir deyişle veri kümesindeki öznitelik sayısını göstermektedir. Ortak bilgi kavramı sadece iki öznitelik arasındaki benzerlik seviyesini değil herhangi bir öznitelik i ile sınıf etiketleri vektörü, ($h(h = [h^1, h^2, h^3, \dots, h^n])$) ve ona denk düşen ayrık rastlantı değişkeni, H , arasındaki benzerliğin, $I(H, F_i)$, ölçülmesinde kullanılmaktadır.

S seçilmek istenen öznitelik kümesi olsun ve $|S|$ bu kümenin eleman sayısını göstere. Seçilmek istenen kümenin seçilebilecek en iyi küme olmasını garanti etmek için yukarıda bahsedilen iki koşulun karşılanması gerekir. Bunlardan ilki maksimum ilişki (maximum relevance):

$$\max W, \quad W = \frac{1}{|S|} \sum_{F_i \in S} I(F_i, H) \quad (3.19)$$

İkincisi ise minimum artıklık (minimum redundancy):

$$\min V, \quad V = \frac{1}{|S|^2} \sum_{F_i, F_j \in S} I(F_i, F_j) \quad (3.20)$$

koşuludur. MRMR algoritması bu iki koşulu iki basit kombinasyon ile birleştirmektedir:

$$\max(V - M) \text{ ve } \max(V/M) \quad (3.21)$$

İdeal olarak yukarıda tanımlanan iki koşulun karşılanması sonucunda en iyi öznitelik kümesinin elde edilmesi $O(N^{|S|})$ karmaşıklığa sahip bir algoritma gerektirmektedir. Arama uzayının çok büyük olması nedeniyle, MRMR algoritması bu optimizasyon sorununu şu şekilde çözer: Öncelikle ilk öznitelik eşitlik 3.19'a göre seçilir. Bu şunu gösterir seçilen ilk öznitelik sınıf etiketleri ile maksimum seviyede ortak bilgi içeren öznitelik olacaktır. Bundan sonraki her basamakta eşitlik 3.22 veya 3.23'ü sağlayan öznitelik F_i seçilir ve seçilen öznitelikler S kümesinde tutulur. Burada S , algoritmanın her basamağında seçilen özniteliklerin tutulduğu kümeyi, Ω tüm öznitelik kümesini, Ω_s ise seçilmiş öznitelikler dışında kalan tüm öznitelikleri ifade etmektedir ($\Omega_s = \Omega - S$).

$$MID(\text{Ortak Bilgi Farkı}) = \max_{F_i \in \Omega_s} \left[I(F_i, H) - \frac{1}{|S|} \sum_{F_j \in S} I(F_i, F_j) \right] \quad (3.22)$$

$$MIQ(\text{Ortak Bilgi Oranı}) = \max_{F_i \in \Omega_s} \left\{ I(F_i, H) / \left[\frac{1}{|S|} \sum_{F_j \in S} I(F_i, F_j) \right] \right\} \quad (3.23)$$

İlk özniteliğin seçilmesinin ardından eşitlik 3.19 ya da 3.20'i sağlayan öznitelik S kümesine dahil edilir ve istenilen öznitelik sayısına ulaşıldığında algoritma sonlanır. Maksimum ilişki ve minimum artıklık koşullarının bu iki basit kombinasyonu mRMR algoritmasının ayrık veriler üstünde kullanılan iki seçim kriterini ortaya çıkarır. Yukarıda geçen algoritmanın $O(Big O)$ notasyonuna göre karmaşıklığı $O(|S|N)$ 'dir

3.2.3. Fisher Puanlama-Fisher Score (FS)

Fisher puanlama, iki sınıf arasındaki ayırt ediciliği ölçen etkili ve basit bir metottur. Bu metotta veri setindeki her bir özelliğin Fisher puanlama değerleri eşitlik 3.24'e göre hesaplanır. Daha sonra veri setinden etkili öznitelikleri seçmek için tüm özelliklerin ortalama Fisher puan değerleri hesaplanarak eşik değeri bulunur. Eğer özneliğin Fisher puan değeri eşik değerinden fazla ise bu öznitelik veri setinin özellik uzayına eklenir.

İki sınıf durumunda her bir t özneliği için;

$$F_t = \frac{(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2} \quad (3.24)$$

değeri hesaplanır. Burada μ_1 , Sınıf 1'e ait t değerlerinin ortalaması; μ_2 , Sınıf 2'ye ait t değerlerinin ortalaması ve σ_1 ve σ_2 'de sırasıyla Sınıf 1 ve Sınıf 2'nin standart sapmalarıdır. Bunun arkasında yatan fikir; bir öznitelik, Sınıf 1 ve Sınıf 2 dağılımlarını ne kadar çok ayırabiliyorsa o kadar iyidir.

İki sınıftan daha fazla sınıf varsa Fisher ölçüt skoru,

$$F_t = \frac{(\mu_1 - \sum_{i=2}^M \mu_i)^2 + \dots + (\mu_j - \sum_{i=1}^M \mu_i)^2 + \dots + (\mu_M - \sum_{i=1}^{M-1} \mu_i)^2}{\sum_{i=1}^M \sigma_i^2} \quad (3.25)$$

biçiminde kolayca genelleştirilebilir. Burada M sınıf sayısıdır. Fisher ölçüt skoru aracılığıyla tüm öznitelikler için bir sıralama hesaplanır. Metodu karakterize eden şey, öznitelikler arasındaki tam bağımsızlık varsayımıdır. Her F_t katsayısı tek bir öznitelik hakkındaki bilgiyle hesaplanır ve öznitelikler arasındaki karşılıklı bilgiyi hesaba katmaz [29].

3.2.4. Relief algoritması

Kira ve Renden [30] tarafından önerilen iterasyona dayalı Relief algoritmasında öznitelikler, hedef fonksiyona uygunluklarına göre seçilir. Relief algoritması veri setinde yer alan niteliklerin ağırlıklandırmasında kullanılmaktadır. RELIEF yöntemi, özniteliklerin değerini aralarında bulunan ya da bulunmayan bağımlılıkları ortaya çıkarmaya çalışarak bulmayı hedefler. Bunu gerçekleştirmek için, özelliğin bulunduğu örneğin ait olduğu ve

olmadığı sınıflarda yer alan en yakın örnekleri ağırlıklandırarak karşılaştırır. Relief algoritmasının ana fikri, k en yakın komşu algoritmasının temel kuralına benzemektedir. Yani verilen bir mesafeye daha yakın uzaklıkların aynı sınıfa ait olmaları, çok daha olasıdır. Eğer bir öznitelik yararlı ise, aynı sınıfın en yakın mesafeleri o öznitelik boyunca verilen aralığa, diğer tüm sınıfların en yakın mesafelerinden daha yakın olması beklenecektir. Dolayısıyla, verilen i özniteliğinin ağırlığı (önem derecesi) Denklemde görüldüğü gibi hesaplanacaktır.

$$(W = W - \text{diff}(x_{ij}, \text{near_hit}_{ij})^2 + \text{diff}(x_{ij}, \text{near_miss}_{ij})^2) / m \quad (3.26)$$

biçiminde verilir. Burada m örnek boyutu (eğitim setinden rastgele seçilmiş bir alt küme) ve $\text{diff}(x_{ij}, \text{near_hit}_{ij})$, rastgele seçilmiş j mesafesi içindeki i özniteliğinin değeri ile aynı sınıfa sahip en yakın eğitim örneğindeki i özniteliğinin (near_hit_{ij}) değeri arasındaki farklılıktır. Paralel olarak, $\text{diff}(x_{ij}, \text{near_miss}_{ij})$ farklı bir sınıfa sahip en yakın eğitim örneğindeki i 'nin değeri olarak tanımlanır. Yararlı öznitelikler için x_{ij} ve near_hit_{ij} değerlerinin çok yakın olması beklenir. Eğer bir öznitelik yararlı değilse, her iki farklılığın da hemen hemen aynı dağılımı almaları beklenir [31].

Orijinalinde iki sınıflı problemler için geliştirilen algoritma, ReliefF şeklinde çok sınıflı problemlere de uyarlanmıştır [32].

3.2.5. Gini index algoritması

ID3, C4.5 gibi karar ağacı metotlarında en ayırt edici niteliği belirlemek için her nitelik için bilgi kazancı ölçülür. Bilgi Kazancı ölçümünde Entropy kullanılır. Entropy rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir.

Eğer veri seti D , n tane sınıftan örnekler içeriyorsa, Gini Index:

$$\text{GINI}(t) = 1 - \sum_{j=1}^n p_j^2 \quad (3.27)$$

p_j , D veri setindeki j sınıfının relatif frekansıdır.

Eğer D veri setindeki örnek büyüklüğü sırayla $N1$ ve $N2$ olan, $D1$ ve $D2$ alt veri setlerine bölünmüşse; bölünmüş verinin gini indeksi n tane sınıftan örnekler içermektedir. Bu durumda gini indeksi eşitlik 3.28'deki gibi bulunur.

$$GINI_{split}(D) = \sum_{i=1}^k \frac{n_i}{n} GINI(D_2) \quad (3.28)$$

Bu algoritmada, değerler üzerinde öznitelikler sıralanır. Bu değerler doğrusal olarak taranır, her seferinde sayı matrisi güncellenir ve gini indeksi hesap edilir En küçük gini indeksine sahip olan nitelik seçilir.

3.3. Sınıflandırma Algoritmaları

Seçilen öznitelikler çok boyutlu bir öznitelik vektöründe toplanarak farklı sınıflandırma algoritmalarıyla tasarlanmış modellere verilmiştir. Bu modellerin kurulmasında bazı parametrik olmayan sınıflandırma yöntemleri kullanılmıştır. Tezin amacına uygun olarak, bu yöntemlerle kişinin hasta olup olmadığının tespiti yapılmış, sonucunda yöntemlerin çalışmadaki doğruluk oranları belirlenmiştir.

Çalışmada kullanılan ses verilerinin istatistiksel analiz ve örüntü tanıma işlemlerinin yapılabilmesi ve karşılaştırılmalarının yapılması için FFNN (ileri beslemeli sinir ağı), RBFN (radyal tabanlı fonksiyon ağı), DT (karar ağaçları), SVM (destek vektör makineleri) ve RF (rastgele orman) algoritmalarından faydalanılmıştır.

3.3.1. Yapay Sinir Ağları (YSA-ANN)

Bilinen en gizemli ve karmaşık hesaplayıcı insan beynidir [33]. Beynin öğrenme kapasitesi, nöronlar ve bunların birbiri ile olan bağlantısına bağlıdır [34]. Yapay sinir ağları (ANN), insan beyninin işleyişini taklit ederek yeni sistem oluşturmaya çalışan yaklaşımlardır [33].

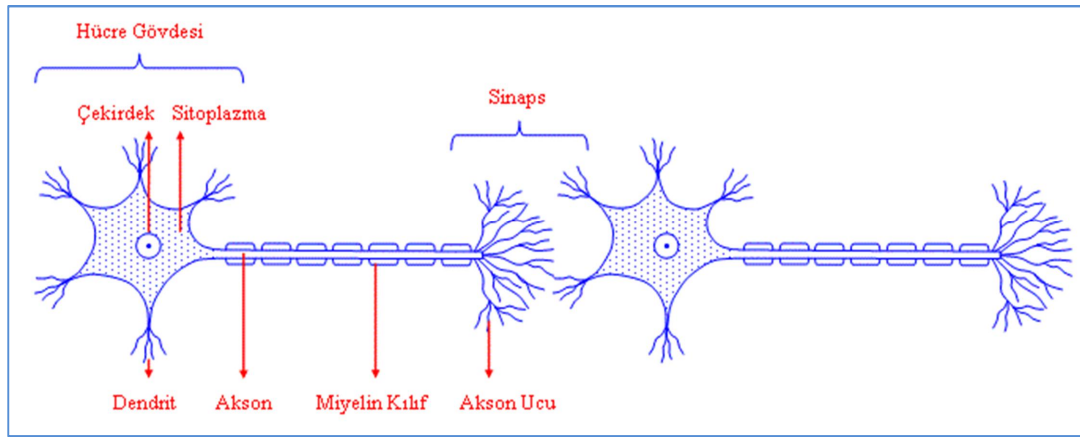
Yapay sinir ağları, birçok basit işlemci elemandan oluşan yapılardır. Bu elemanlar farklı formda ifade edilebilen nümerik verileri taşıyan “bağlantılar” veya “ağırlıklar” ile birbirlerine bağlıdır [35]. Sinir ağları, örüntü tanıma, belirleme, sınıflandırma uygulamaları ile ses, görüntü ve kontrol sistemlerini içeren çok çeşitli alanlarda karmaşık fonksiyonları çözümlmek için eğitilirler [36].

Yapay sinir ağları bilgisayar bilimine de bazı yenilikler getirmiştir. Özellikle olaylar hakkında bilgilerin olmadığı fakat örneklerin bulunduğu durumlarda çok etkin olarak kullanılabilecek bir karar verme aracı ve hesaplama yöntemi olarak görülebilir [37].

Geleneksel bilgisayar yazılım teknolojisi ile çözülemeyen birçok problemin yapay sinir ağları ile çözülebileceği aşıkardır. Mesela yapay sinir ağları, eksik, normal olmayan belirsiz bilgileri işleyebilen en güçlü problem çözme tekniğidir [37].

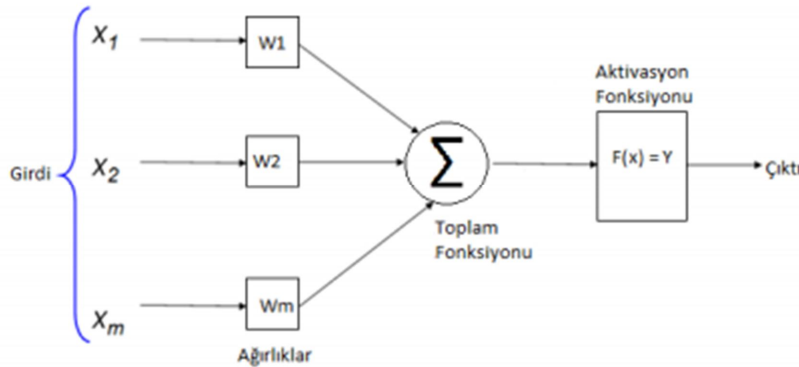
Yapay Sinir Ağlarının Yapısı

Sinir sistemi birbirleriyle iletişim halinde olan sinir hücrelerinden oluşmaktadır. Yapay sinir ağları biyolojik sinir sisteminden esinlenerek oluşturulmuştur.



Resim 3.2. Sinir hücresinin yapısı

Biyolojik sinir ağlarının sinir hücreleri olduğu gibi yapay sinir ağlarının da yapay sinir hücreleri vardır. Yapay sinir hücreleri mühendislik biliminde proses olarak da adlandırılmaktadır. Her proses elemanının beş ana bölümü vardır [38].



Şekil 3.2. Proses elemanının temel elemanları

Şekil 3.2 'de gösterilen yapay sinir hücresi temel olarak beş kısımdan oluşmaktadır. Bunlar; girişler, ağırlıklar, toplama işlemi, aktivasyon fonksiyonu ve çıkış olarak sıralanır.

Yapay sinir ağlarında öğrenme işlemi girişler ile gerçekleştirilir. Dış dünyadan veya bir önceki katmandan alınan bilgiler giriş olarak yapay sinir hücrelerine gönderilir. Ağırlıklar, girişlerin yapay sinir hücreleri üzerindeki etkisini belirleyen ve öğrenme işleminin gerçekleştirilmesini sağlayan katsayılardır. Yapay sinir hücresine gelen net girişin hesaplanması toplama işlemi olarak adlandırılır. En yaygın kullanılan yöntem ağırlıklı toplamı bulmaktır. Ağırlıklı toplamı bulabilmek için yapay sinir hücresinin girişindeki her bir ağırlığın ait olduğu girişlerle çarpımının toplamına eşik değeri ilave edilir. Eşik değeri biyolojik sinir hücrelerinin yüzeylerindeki potansiyel farkının bir benzerini yapay sinir hücrelerinde elde etmek için kullanılan kat sayıdır. Çıkış işlevi sonucunun dış dünyaya veya diğer sinirlere gönderildiği yerdir. Bir sinirin bir tek çıkışı vardır. Sinirin bu çıkışı kendinden sonra gelen diğer sinirlere giriş olabilir. Yapay sinir ağları sadece sayısal giriş bilgileri ile çalışırlar. Bu nedenle giriş bilgileri genellikle ölçeklendirilir [39].

$$net = \sum_i^n x_i w_i + \emptyset \quad (3.29)$$

Aktivasyon Fonksiyonları

Transfer veya aktivasyon fonksiyonları, öğrenme eğrisi olarak da adlandırılmaktadır. Bu fonksiyon hücreye gelen net girdiyi işleyerek hücrenin bu girdiye karşılık üreteceği çıktıyı belirler. Bu değerler çoğunlukla $[0,1]$ veya $[-1,1]$ arasındadır. Aktivasyon fonksiyonu genellikle doğrusal olmayan bir fonksiyon seçilir. Yapay sinir ağlarının bir özelliği olan “doğrusal olmama” aktivasyon fonksiyonlarının doğrusal olmama özelliğinden gelmektedir. Aktivasyon fonksiyonu seçilirken dikkat edilmesi gereken bir diğer nokta ise fonksiyonun türevinin kolay hesaplanabilir olmasıdır. Günümüzde en yaygın olarak kullanılan “Çok katmanlı algılayıcı” modelinde genel olarak aktivasyon fonksiyonu olarak “Sigmoid fonksiyonu” kullanılır.

Çizelge 3.2. Bazı aktivasyon fonksiyonları

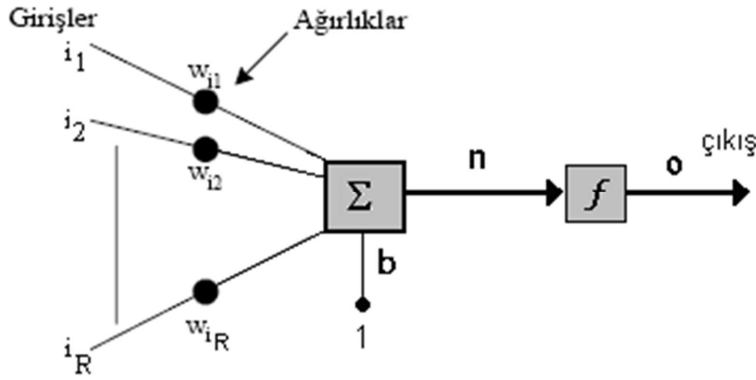
Aktivasyon Fonksiyonu	Matematiksel Tanımı
Lineer	$f(x) = x$
Lojistik Sigmoid	$f(x) = 1 / 1 + \exp(-x)$
Hiperolik Tanjant	$f(x) = \tanh(x)$
Gaussian	$f(x) = \exp(-x^2/2\sigma^2)$

Yapılarına Göre Yapay Sinir Ağları

Yapay sinir ağları, genel olarak ağı yapısına göre; ileri beslemeli ve geri beslemeli olarak ikiye ayrılabilir. İleri beslemeli ağlar genellikle örüntü tanıma problemlerinde, geri beslemeli ağlar da optimizasyon problemlerinde yaygın olarak kullanılmaktadır [38].

İleri Beslemeli Tek Katmanlı Ağlar

Girdi işaretlerinin ve ağırlıkların bir işlevini hesaplayıp, sonucu sonraki tüm yapay sinir hücrelerine iletir. Bu ağlar çok katmanlı ileri beslemeli ağlar olarak isimlendirilir. Şekil 3.3'te ileri beslemeli tek katmanlı ağ gösterilmiştir.

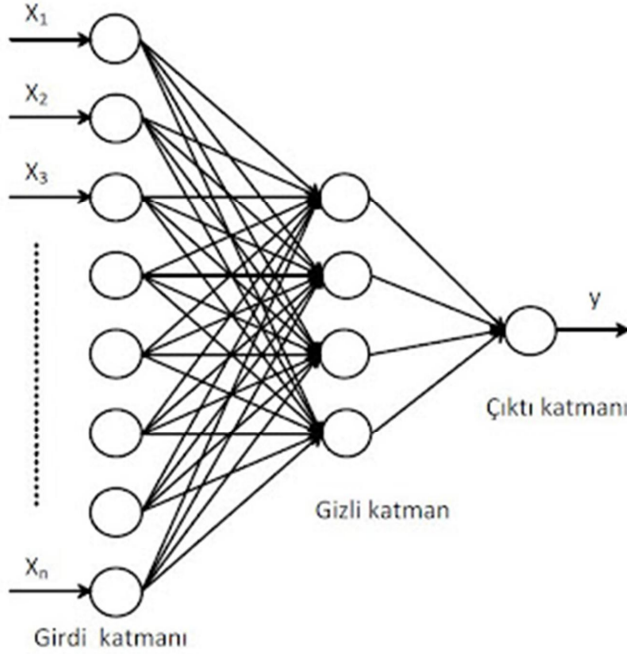


Şekil 3.3. İleri beslemeli tek katmanlı yapay sinir ağı

Perceptron ilk defa Rosenblatt tarafından örüntü (şekil) sınıflandırma amacı ile geliştirilmiştir. Perceptron bir sinir hücresinin birden fazla girdiyi alarak bir çıktı üretmesi prensibine dayanmaktadır. Ağıın çıkışı bir veya sıfırdan oluşan mantıksal değerdir. Çıktının değerinin hesaplanmasında eşik değer fonksiyonu kullanılır. Bu ağlarda proses elemanlarının değerlerinin ve dolayısıyla ağıın çıkışının sıfır olmasını önleyen birde eşik değeri vardır. Tek katmanlı ağlarda çıktı fonksiyonu doğrusal fonksiyondur. Yani ağı gösterilen örnekler iki sınıf arasında paylaştırılarak iki sınıfı birbirinden ayıran doğru bulunmaya çalışılır [40].

İleri Beslemeli Çok Katmanlı Ağlar

Bir yapay sinir ağının öğrenmesi istenen olayların girdi ve çıktıları arasındaki ilişkiler doğrusal olmayan ilişkiler olursa bu tür olayların öğrenilmesi için daha gelişmiş modellere ihtiyaç vardır. Çok katmanlı ağlar bunlardan birisidir [36].



Şekil 3.4. İleri beslemeli çok katmanlı yapay sinir ağları

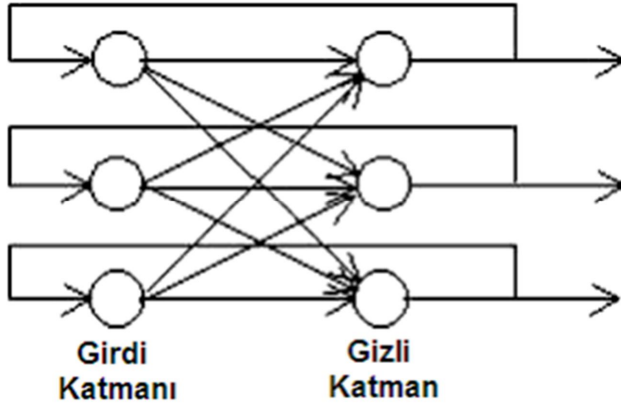
Bu ağlar üç katmandan oluşur. Girdi katmanı, dış dünyadan bilgileri alır ve bu katmanda herhangi bir bilgi işleme olmaz. Buradaki çıktı bir sonraki katmanda bulunan bütün proses elemanlarına gönderilir. Ara (gizli) katmanlar, girdi katmanından gelen bilgileri işlerler. Bir adet ara katman ile birçok problemi çözmek mümkündür. Eğer ağın öğrenmesi istenilen problemin girdi çıktı arasındaki ilişki doğrusal olmaz ve karmaşıklık artarsa birden fazla sayıda ara katman kullanılabilir. Çıktı katmanı, ara katmandan gelen bilgileri işleyerek ağı girdi katmanından sunulan girdi için ağın ürettiği çıktıyı bulur. Bu çıktı dış dünyaya iletilir [41].

Çok katmanlı ağların eğitilmesi sırasında ağın ürettiği çıktılar ile beklenen çıktılar arasındaki farkın veya hatanın ağın ağırlıklarına dağıtılarak zaman içinde bu farkın an aza indirgenmesidir. Öğrenme sırasında önce girdiler ağı sunularak bu girdilere karşılık gelen

çıktılar üretilir. Bu işleme ileri doğru hesaplama denir. Daha sonra üretilen çıktı ile beklenen çıktı karşılaştırılarak aradaki hata geriye doğru dağıtılarak ağırlıklar değiştirilir. Buna da geriye doğru hesaplama denir [41].

Geri Beslemeli Ağlar

Geri beslemeli ağlarda isminden de anlaşılacağı gibi bir tür geri besleme işlemi vardır. Hopfield ağı bu tür mimariye sahip bir yapay sinir ağıdır. Geri beslemeli yapay sinir ağlarında bir hücrenin çıktı değeri kendisine, başka bir hücreye veya giriş katmanına giriş olabilmektedir. Bağlantı çift yönlüdür. İleri beslemeli yapay sinir ağlarından farklı olarak bağlam (context) tabakası yer almaktadır. Bu tabaka gizli tabaka çıktılarını ağırlıklandırılarak tekrar gizli tabakaya girdi olarak vermektedir [42]. Şekil 3.5'te geri beslemeli yapay sinir ağı mimarisi gösterilmiştir.



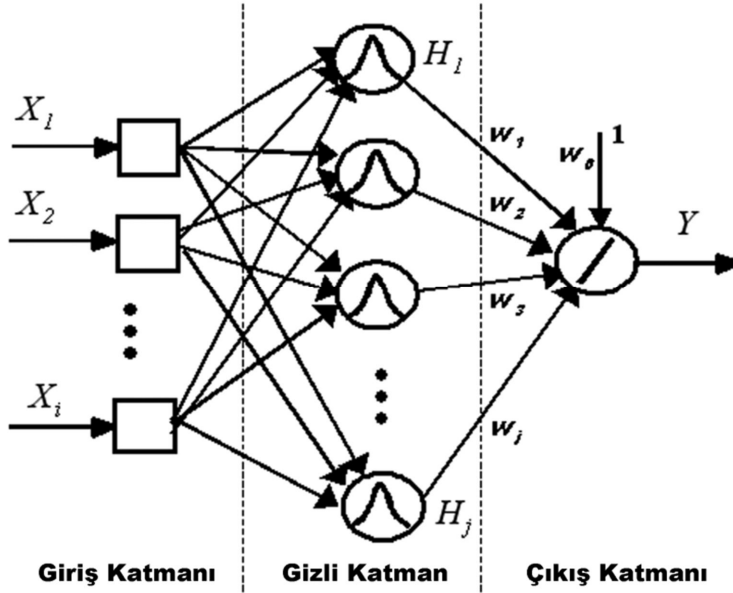
Şekil 3.5. Geri beslemeli yapay sinir ağı

3.3.2. Radyal Tabanlı Fonksiyon Ağları (RBFN)

Moody ve Darken (1989) tarafından popüler hale getirilen RBFN, danışmalı öğrenme kapsamında değerlendirilen ileri beslemeli bir YSA modelidir. RTYSA modellerinin eğitimini çok boyutlu uzayda eğri uydurma yaklaşımı olarak görmek mümkündür [45].

RBF ağı; giriş katmanı, gizli katman ve çıkış katmanı olmak üzere üç katmandan oluşan ileri beslemeli bir ağ yapısıdır. Fakat klasik YSA yapılarından farklı olarak bu ağlarda, girdi katmanından gizli katmanına geçişte radyal tabanlı aktivasyon fonksiyonları ve

doğrusal olmayan bir kümeleme analizi kullanılmaktadır. Gizli katman ile çıkış katmanı arasındaki yapı ise ağırlıkların atandığı ve hesaplanması için ilgili öğrenme algoritmalarının kullanıldığı kısımdır [48].



Şekil 3.6. Radyal tabanlı fonksiyon ağıнын yapısı [44]

Radyal olarak simetrik olan gizli katman işlem elemanı için üç bileşen vardır. İlki giriş ve gizli katman arasındaki ağırlık vektörü olarak saklanan giriş uzayındaki bir merkez vektörüdür. İkincisi, bir giriş vektörünün merkezden ne kadar uzak olduğunu belirlemek için kullanılan uzaklık ölçütüdür. Üçüncüsü ise, tek değişkenli olan ve uzaklık fonksiyon çıkışını giriş olarak alan işlem elemanının çıkış değerini belirleyen bir aktivasyon fonksiyonu yapısıdır. İlk katmandaki işlem elemanları girişlerin ağırlıklandırılmış şeklini kullanmaz. İlk katmandaki işlem elemanlarının çıkışları, yapay sinir ağı girişleri ile temel fonksiyonun merkezi arasındaki uzaklığa göre belirlenir. Radyal tabanlı yapay sinir ağı yapılarının son katmanı lineerdir ve ilk katmanın çıkışlarından ağırlıklandırılmış toplam çıkışı üretir [44].

RBF'in genel matematiksel ifadesi aşağıda verilmiştir:

$$s_j(x) = \sum_{i=1}^N w_{ij} \varphi_i(x) + b_j \quad (3.30)$$

Burada w_{ij} i. gizli katman ile j. çıkış nöronu arası dönüşümdeki ağırlığı göstermekte, b_j muhtemel bir yanlılık, $\varphi(x)$ ise aktivasyon fonksiyonudur. RBFN’de gizli katman aktivasyon fonksiyonu genellikle standart Öklid uzaklıklarını üstel fonksiyondan geçiren Gauss fonksiyonudur ve aşağıdaki gibi tanımlanır:

$$\varphi_i(x) = \exp\left\{-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right\} \quad (3.31)$$

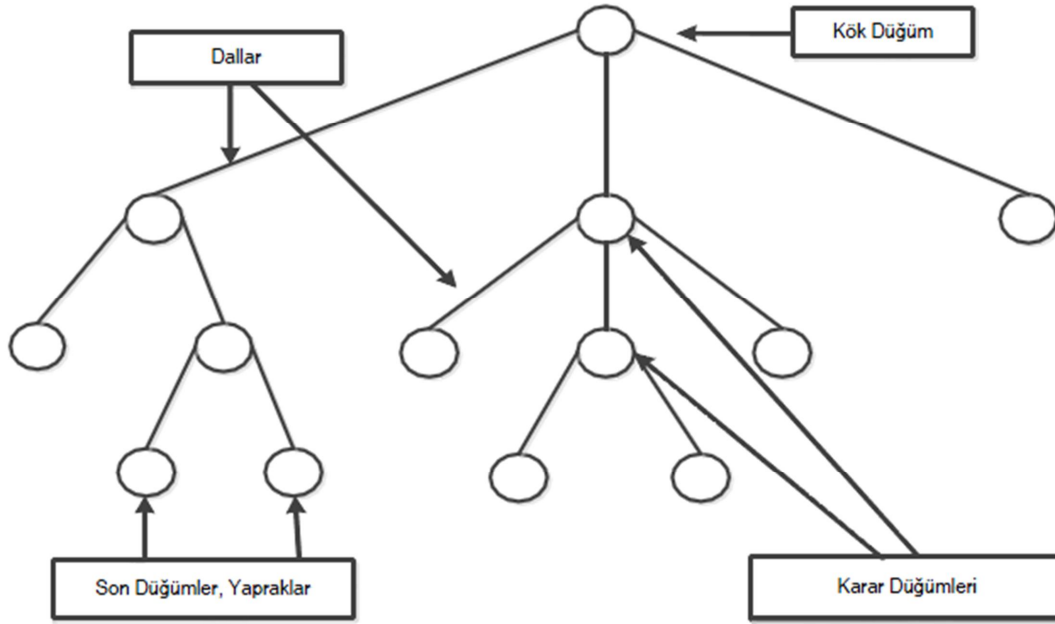
Bu ifadede x giriş vektörü c_i merkez vektörüdür.

3.3.3. Karar Ağaçları (DT)

Karar ağacı, çok sayıda kayıt içeren bir veri kümesini, bazı kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir ağaç grafiktir. Karar ağacı, düğümler ve çizgilerden oluşur.

Karar vericinin kontrolü dışındaki olaylar ağaç üzerinde daireler gösterilmesine olay düğüm noktası denir. Daire biçimindeki düğüm noktalarından çıkan her dal bir olayı simgeler. Olay düğüm noktalarından çıkan dallar, karar vericiyi bir başka düğüm noktasına veya bir karar noktasına götürebilir.

Karar ağaçları tüm kullanıcılar tarafından açık ve anlaşılır bir şekildedir. Problemin çözümünde sondan başa doğru hesaplama yöntemi uygulanır. Her bir karar noktası kendisinden sonraki yolu belirleyecek soru vardır. Bu sorunun cevabına göre ağaç dallanır ve sonuç kümesine doğru ilerler. Burada önemli olan en az soruyla en doğru sınıfa doğru gitmektir. Bu ilerleyiş sırasında karşılaşılan karar noktalarının her birinde, o düğümün beklenen değeri hesaplanır. Beklenen değerlerden en iyi olanı o düğümün üzerine yazılır. O düğümdeki tüm öğeler aynı sınıfa dahil olması durumunda ve sınıflandırılacak değer kalmadıysa döngü sonlandırılır ve karar ağacı oluşturulmuş olur [43,49]. Şekil 3.7’de karar ağacı örneği gösterilmiştir.



Şekil 3.7. Karar ağacı yapısı [46]

Karar Ağaçlarında Dallanma Kriterleri

Karar ağaçları oluştururken dallanmanın nasıl yapılacağını dikkat edilmesi gereken en önemli hususlardan biridir. Her farklı kriter bir karar ağacı algoritmasını oluşturur. Bu algoritmaları Entropiye dayalı algoritmalar ve Sınıflandırma ve regresyon ağaçları (CART) şeklinde ayrılabilir [46,50].

Entropiye Dayalı Algoritmalar

Bir sistemdeki belirsizliğin ölçüsüne “entropi” adı verilir [45]. Entropi verilerin birbirinden farklılığı olarak ele alınırsa aynı değere sahip iki ifadenin entropileri 0 çıkmaktadır. Diğer bir ifadeyle bir verinin entropi ölçüsü ne kadar fazla ise, o alan kullanılarak oluşturulan sonuçlar da bir o kadar belirsiz ve kararsızdır [45]. Bu sebeple karar ağacının kökünde entropi ölçüsü en az belirsizlik ve kararsızlığı olan alanlar tecih edilir. Eğer $P = (p_1, p_2, \dots, p_n)$ şeklinde bir olasılık dağılımı verilirse P ’nin entropisini hesaplayan formül aşağıdaki gibidir :

$$I(P) = - (p_1 * \log(p_1) + p_2 * \log(p_2) + \dots + p_n * \log(p_n)) \quad (3.32)$$

Karar ağaçlarının oluşturulmasında dallanmaya hangi nitelikten başlanacağı önemlidir. Çünkü sınırlı sayıda veriden oluşan bir eğitim kümesinin tüm ağaç yapılarını ortaya çıkarmak ve en uygununu seçerek ondan başlamak kolay değildir. Bu nedenle karar ağacı algoritmalarının dallanması entropinin alacağı değere göre gerçekleştirilebilir [46,50].

Sınıflandırma ve Regresyon Ağaçları (CART)

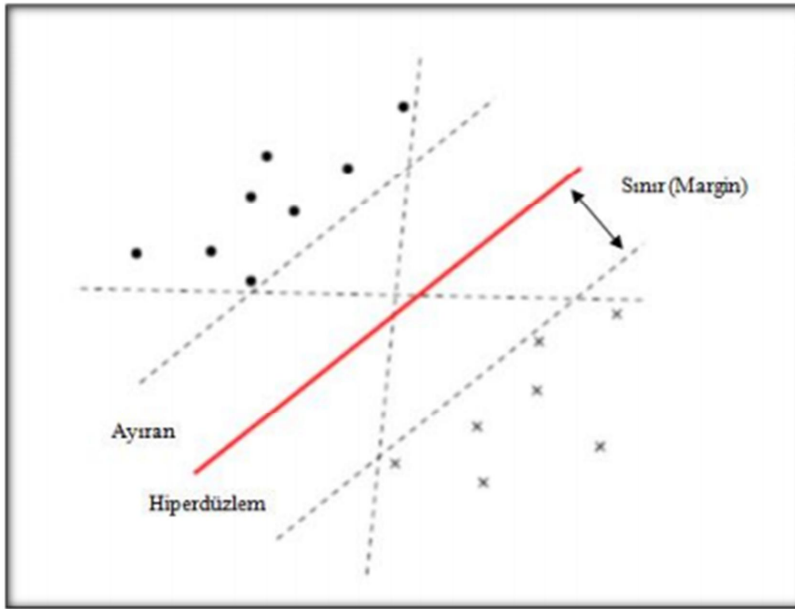
1984'te Breiman tarafından sınıflandırma ve regresyon ağaçları (CART - Classification and Regression Trees) olarak ortaya atılmıştır. CART karar ağacında, her bir düğümünün iki dala ayrılması gerekir. CART algoritmasında, bir düğümde belli bir kriter uygulanarak bölünme işlemi gerçekleştirilir. Dallara ayırma kriteri hesaplarken kayıp verileri kullanmaz. Bu bölünmeler üzerinde seçme işlemi uygulanır [46,50].

3.3.4. Destek Vektör Makineleri (SVM)

Destek Vektör Makineleri (DVM) ikili sınıflandırma için kullanılan popüler bir yöntemdir. Bunun en önemli sebebi; doğrusal ve doğrusal olamayan şekli ile iyi sınıflandırma sağlamasıdır.

DVM'lerin temelindeki en iyi yöntem; destek vektörlerini maksimum uzaklıkta birbirinden ayıran hiper düzlemi bulmadır. Bu işlemi hiper düzlemin verileri en iyi şekilde ayırmasını sağlayacak şekilde yapar.

Hiper düzlemin sınıfları mümkün olduğunca birbirinden uzakta ayıracak şekilde olması tercih edilir. Böylece hiper düzlemin görünmeyen verileri de bu ayırma ile en iyi şekilde genelleştirmesini sağlanır. Hiper düzlemin uzaklığı, veri setindeki verileri en iyi şekilde sınıflandıracak en geniş sınır anlamına gelir. Bir veri setini iki sınıfa ayıran iki boyutlu bir uzay olan hiper düzlemi Şekil 3.8'de gösterilmektedir [46,51].



Şekil 3.8. Destek vektör makineleri [46]

Doğrusal olarak ayrılabilen iki sınıflı bir sınıflandırma probleminde DVM'nin eğitimi için n tane örnekten oluşan eğitim verisinin $\{x_i, y_i\}$, $i = 1 \dots n$ kadar olduğu kabul edilirse, en uygun hiper düzleme ait denklemler aşağıdaki şekilde hesaplanır:

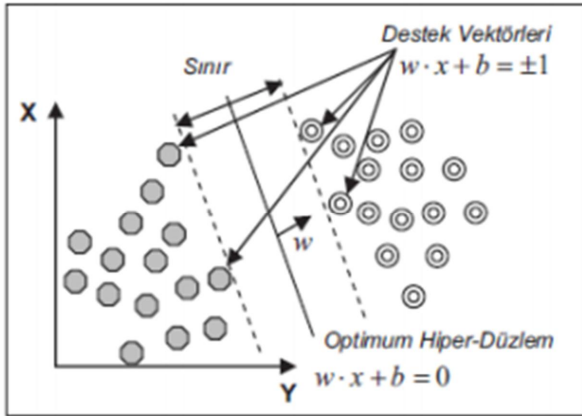
$$y_i = +1 \text{ için } w^T \cdot x_i + b \geq +1 \quad (3.33)$$

$$y_i = -1 \text{ için } w^T \cdot x_i + b \geq -1 \quad (3.34)$$

x : N boyutlu bir uzayı, $y: \{-1, +1\}$ sınıf etiketlerini, w : ağırlık vektörünü, b : eğilim değerini ifade etmektedir.

DVM'de en uygun hiper düzlemin belirlenmesi sınıflandırmanın başarısını etkiler. Bunu için veri setindeki verilerin, hiper düzlemin doğru tarafında olmalarının yanı sıra iyi bir genelleştirme için hiper düzlemden belirli bir uzaklıkta bulunmaları da gerekmektedir.

En uygun hiper düzlemin belirlenmesi için hiper düzleme paralel ve sınırları belirleyecek destek vektörleri adı verilen ve Şekil 3.9'da gösterilen iki hiper düzlem belirlenir [46].



Şekil 3.9. Destek vektörleri [46]

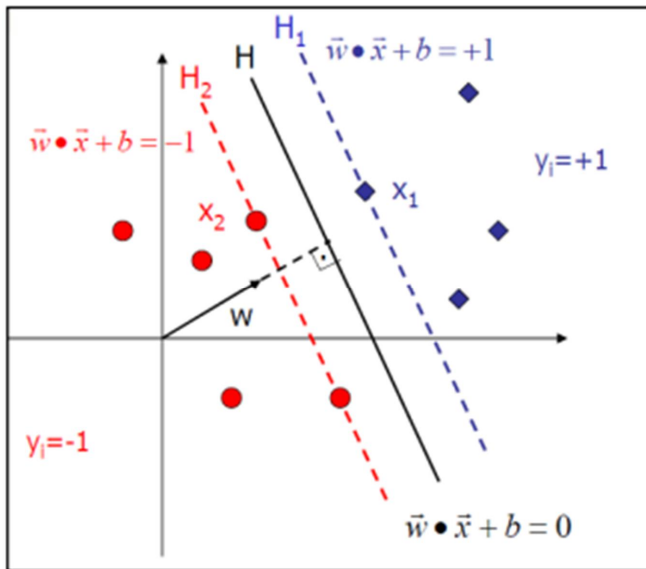
Destek vektör makinelerinde uzaklık hesaplama işlemi Şekil 3.10'daki için model alınırsa aşağıda gösterildiği gibi yapılmaktadır.

$$(w \cdot x_1) + b = y_i = +1 \quad (3.35)$$

$$(w \cdot x_2) + b = y_i = -1 \quad (3.36)$$

$$\Rightarrow w \cdot (x_1 - x_2) = 2$$

$$\Rightarrow \frac{w}{\|w\|} \cdot (x_1 - x_2) = \frac{2}{\|w\|}$$



Şekil 3.10. Destek vektör makinelerinde uzaklık hesaplama [46]

En uygun hiper düzlemin sınırının yani uzaklığın ($\frac{2}{\|w\|^2}$ ifadesinin) maksimuma çıkarılması için $y_i(w \cdot x + b) \geq +1$ olacak şekilde $\frac{\|w\|^2}{2}$ nin minimum olması gerekir. Bu problemin çözümünde Lagrange denklemleri kullanılır.

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^k a_i \left[y_i(w \cdot x_i + b) + \sum_{i=1}^k a_i \right] \quad (3.37)$$

Bu işlem sonrasında iki sınıflı doğrusal ayrılabilen bir problem için karar fonksiyonu denklem 3.38'daki gibi yazılabilir [46].

$$f(x) = \text{sign} \left(\sum_{i=1}^k \lambda_i y_i(x \cdot x_i) + b \right) \quad (3.38)$$

3.3.5. Rastgele Orman (RF)

Rastgele orman birçok karar ağacından oluşan bir topluluk öğrenme algoritmasıdır. Bireysel olarak oluşturulan ağaçlar arasında oylama sonucuna göre kazanan sınıfı çıktı olarak veren bir makine öğrenmesi yöntemidir. Rastgele Orman, Tin Kam Ho'nun Rastgele Karar Ağaçları fikrinden esinlenerek Leo Breiman ve Adele Cutler tarafından geliştirilmiştir [47,52].

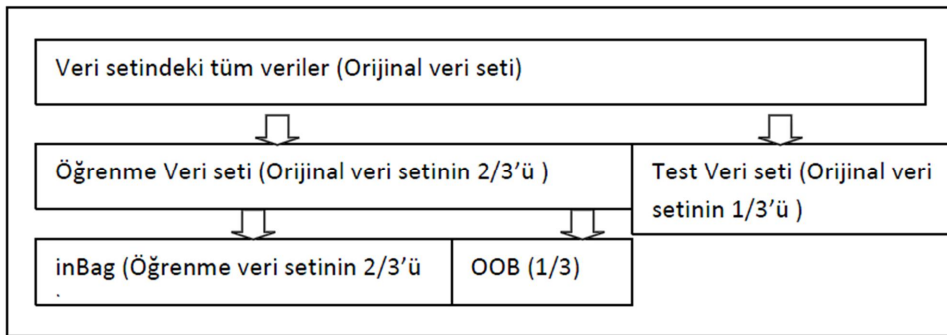
Rastgele Orman algoritmasında karar ağaçları şu şekilde oluşturulur:

- R adet özelliği olan, K elemanlı bir eğitim kümesi için,
- Bir ağacın düğümündeki kararı verebilmek için gerekli olan giriş değişkenlerinin sayısı olan r verilir. Bu r değeri R değerinden çok daha küçüktür.
- K adet eğitim kümesindeki örneklerden K kez yenisiyle değiştirme yaparak rastgele eğitim örnekleri seçilir. Kalan örnekler ise ağacın hata oranını hesaplamak için kullanılır.
- Ağacın her düğümünde seçilmiş olan r adet değişken özellikten yararlanılarak bir karar verilir. Oylamayla verilen karara göre o düğümdeki en iyi ayırım hesaplanır [47].

RF yönteminde sonradan eklenen veriye ait tahmin yapıldığı gibi değişkenlerin önem derecesi de hesaplanmaktadır. Çok sayıda değişkenin yer aldığı veri setinde, RF yöntemiyle önem derecesine göre elde edilen değişkenlerle kurulacak olan yeni modelde, daha sağlıklı ve etkin tahminlerin yapılması gerçekleştirilebilir.

RF yönteminde, oluşturulan modeli test etmek için ayrı bir test verisi yoksa veya orijinal veri, test verisinden ayrılmamışsa, sınıf dağılımına göz önünde bulundurup orijinal veri setinin $2/3$ 'ü öğrenme verisi (inBag), $1/3$ 'ü ise test verisi (Out-Of-Bag (OOB)) olarak ayrılır. Eğer ayrı bir test verisi varsa veya orijinal veri, test verisinden ayrılmışsa, modelin oluşturulması için ayrılan öğrenme verisi, kendi içinde $2/3$ oranında öğrenme verisi (inBag), $1/3$ 'ü ise test veri seti(OOB) olarak ayrılır. RF yönteminde karar ağacı oluşturmak için orijinal veri setinden “inBag” verisi ve “OOB” verisinin hangi oranlarda belirlendiği Şekil 3.11’de gösterilmiştir.

Karar ormanında, oluşturulacak karar ağacı sayısı kadar sayıda bootstrap tekniği ile örneklem oluşturulur. Her örneklem sayısı kadar inBag ve OOB verisi ayrılır. Kurulan her karar ağacı için ayrılan OOB verisi ile testi yapılarak hata oranı tahmini yapılır ve bütün karar ağaçları için yapılan OOB hata oranlarının ortalaması alınarak, karar ormanının OOB hata oranı hesaplanır. RF algoritmasında yeteri kadar karar ağacı oluşturulduysa OOB hata oranının sapmasız ve doğru hesaplandığı yapılan çalışmalar ortaya koymaktadır [53].



Şekil 3.11. Rastgele Orman yönteminde veri seçimi

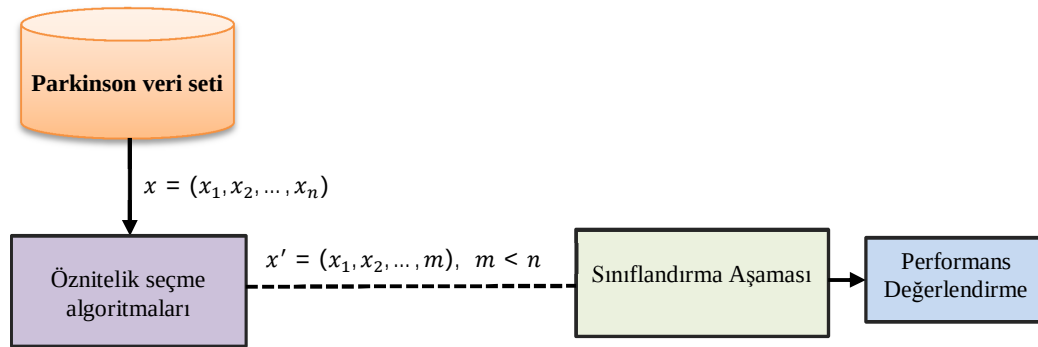
Karar ağaçları kurumasından sonra, test verisindeki her biri farklı bir deneğe ait olan veriler önceden kurulmuş ağaçlar üzerinde yukarıdan aşağıya satır satır yerleştirilir. Bu işlem veri setindeki tüm veriler için tekrar edilir. Hesaplanan OOB hata oranına göre oluşturulan her ağaca bir ağırlık verilir. Hata oranı en düşük ağaç en yüksek ağırlığı, hata

oranı en yüksek ağaç ise en düşük ağırlığı alır. Belirlenen ağırlığa göre her ağaç, sınıflama tahmini için bir oy verme işlemine tabi tutulur. RF algoritmasında ağaçların oyu oluştuktan sonra bu ağırlıklı oylar toplanır. Tüm ağaçlardan ağırlıklı olarak en çok oyu alan sınıf, nihai sınıf tahmini olarak belirlenir. Her bir farklı bir deneğe ait olan veriler satır satır yerleştirildiğinde nihai olarak orman karar vermiş olur. [53].

4. WEB ARAYÜZ YAZILIMI VE DENEYSEL SONUÇLAR

Bu bölümde tez çalışması süresince geliştirilen ve kullanılan yazılım ve web arayüzü tanıtılmaktadır. Ayrıca arayüz ile uygulanan deneyler ve elde edilen sonuçlar yer almaktadır.

Uygulanan yöntemin blok diyagramı Şekil 4.1’de görülmektedir. Sistemde ilk olarak Parkinson hastası ve sağlıklı kişilere ait ses ölçümlerini içeren öznitelik kümesi sisteme sunulmaktadır. Sonraki aşamada etkili özniteliklerin tespiti için öznitelik seçme algoritmalarından yararlanılmıştır. Bu algoritmalar sırasıyla; Fisher score, ReliefF, FCBF, mRMR ve Gini index algoritmalarıdır. Elde edilen etkili öznitelik kümesi sınıflandırma algoritmasına giriş olarak sunulmaktadır. Bu tez çalışması kapsamında sınıflandırma algoritması olarak rastgele orman, karar ağacı, radyal sinir ağı, ileri beslemeli sinir ağı ve destek vektör makinaları tercih edilmiştir.



Şekil 4.1. Parkinson hastalığı için uygulanan metot

4.1. Web Arayüz Yazılımı

Çalışmada uygulamalar MATLAB yazılım platformunda geliştirilmiştir. Web arayüz uygulamasında ise yine MATLAB’da web tabanlı uygulamalar için MATLAB Builder NE ve MATLAB Web Figure araçları tercih edilmiştir.

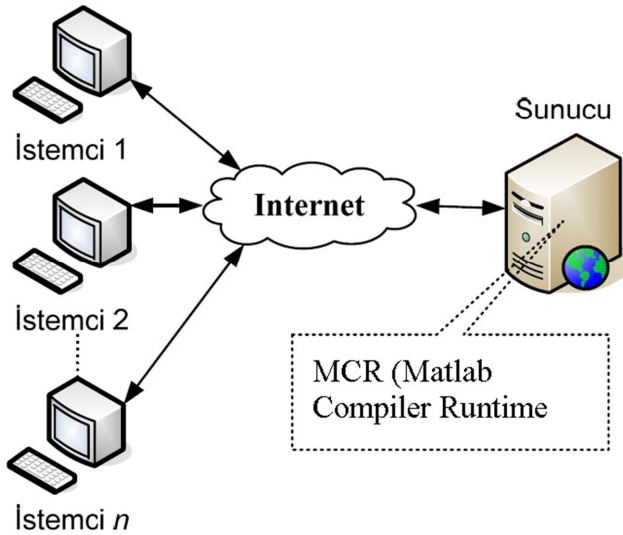
Arayüz tasarımı için MATLAB Web Figure ile ASP.NET platformu kullanılmıştır. Böylece yöntemin yaygınlaştırılmasını sağlayan, aynı anda birden fazla kullanıcının kullanabileceği, akademik amaçlara uygun bir yapı ortaya konmuştur.

4.2. Matlab Builder NE ve Matlab Web Figure

Matlab Builder NE, Matlab Compiler ürününün bir aracıdır ve Matlab fonksiyonlarına .Net programcılarının CLS-compliant dilleri olan C#, VB.Net ve C++ programlama dilleriyle ulaşmalarına imkan verir. Matlab Builder, Matlab kodlarını derleyerek Matlab fonksiyonlarını .Net metodlarına uygun hale dönüştürür. [54,55].

Matlab Builder NE veri dönüşümü, veri sıralama ve dizi biçimlendirme imkanı sağlayarak, yazılan .NET kod çağrıldığında Matlab'ta yazılan ve derlenen fonksiyonları ve özelliklerinin esnek biçimde kullanılmasını sağlar. Matlab Builder NE'de Matlab data tiplerinin desteklenmesi için MWArray.dll dosyasında tanımlı olan MWArray veri dönüşüm sınıfları bulunmaktadır. Kullanılan veri ve dizileri Matlab'a uygun biçime dönüştürmek ve Matlab'tan gelen verileri yazılan programda kullanabilmek için MWArray.dll dosyası uygulamaya referans olarak eklenmesi gerekmektedir. [54,55].

Matlab Builder NE'deki WebFigure özelliği ise Matlab figürlerini bir web sitesinde gösterebilme ve bu figürler üzerinde görsel olarak işlem yapabilme olanağı sağlamaktadır. Bu özellik son kullanıcılara, Matlab programı ve herhangi bir araçlar olmadan grafiksel uygulamaları istedikleri bir yerden yalnızca web tarayıcı aracılığı ile internet üzerinden kullanabilme imkanı sağlamaktadır [54,55]. Burada yapılan uygulamalar bir sunucu üzerinden çalıştırılarak kullanıcılara hizmet verir. Uygulamanın çalışma yapısı Şekil 4.2'deki gibidir.

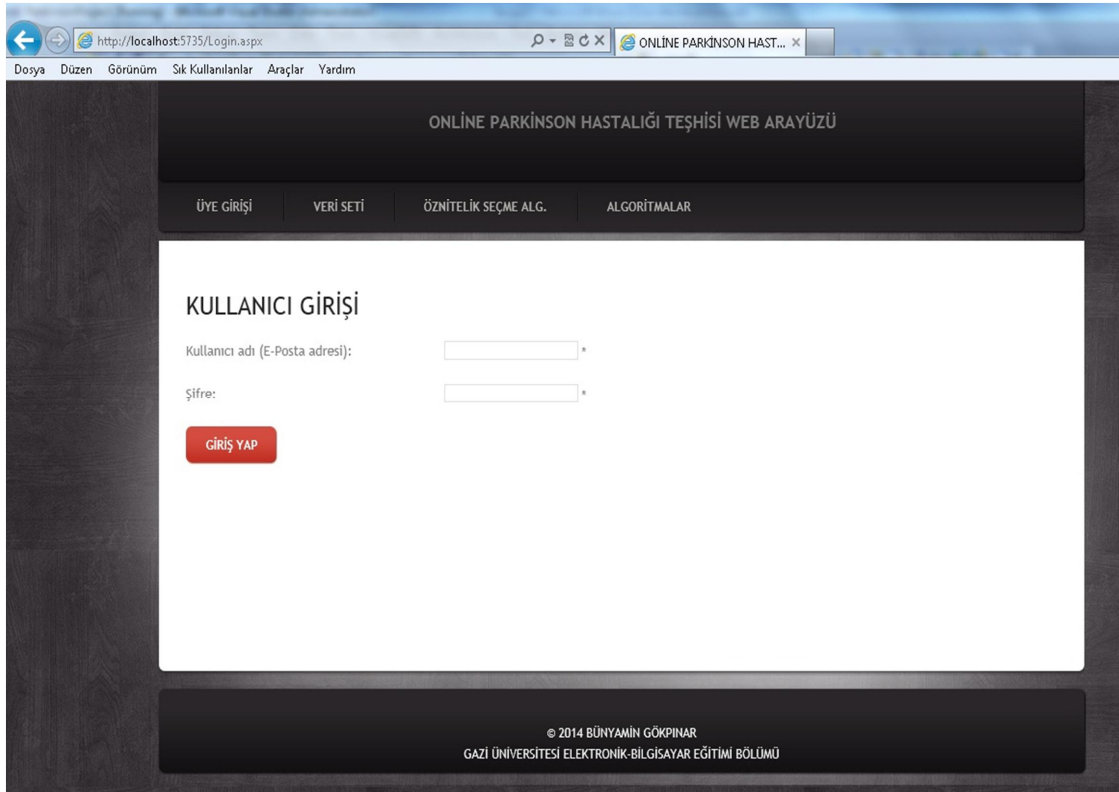


Şekil 4.2. MATLAB Web Figure içeren bir uygulamanın çalışma yapısı

4.3. Tasarlanan Web Arayüz Uygulaması

Gerçekleştirilen web arayüzü tasarımında; Matlab programı ve Matlab Builder NE ile hazırlanmış .Net bileşeni, bu bileşen ile uyumlu biçimde çalışabilen ve arayüz tasarımının gerçekleştirildiği Microsoft Visual Studio 2013 programı ve sunucu hizmeti olarak da IIS 7.5 (Internet Information Service) kullanılmıştır. Web tasarımının alt yapısı ASP.NET ve kullanılan programlama dili ise C# olarak seçilmiştir. Uygulamaya gelen bilgiler SQL Server veri tabanına kaydedilmekte olup Matlab ve Asp.Net için veriler buradan sağlanmıştır.

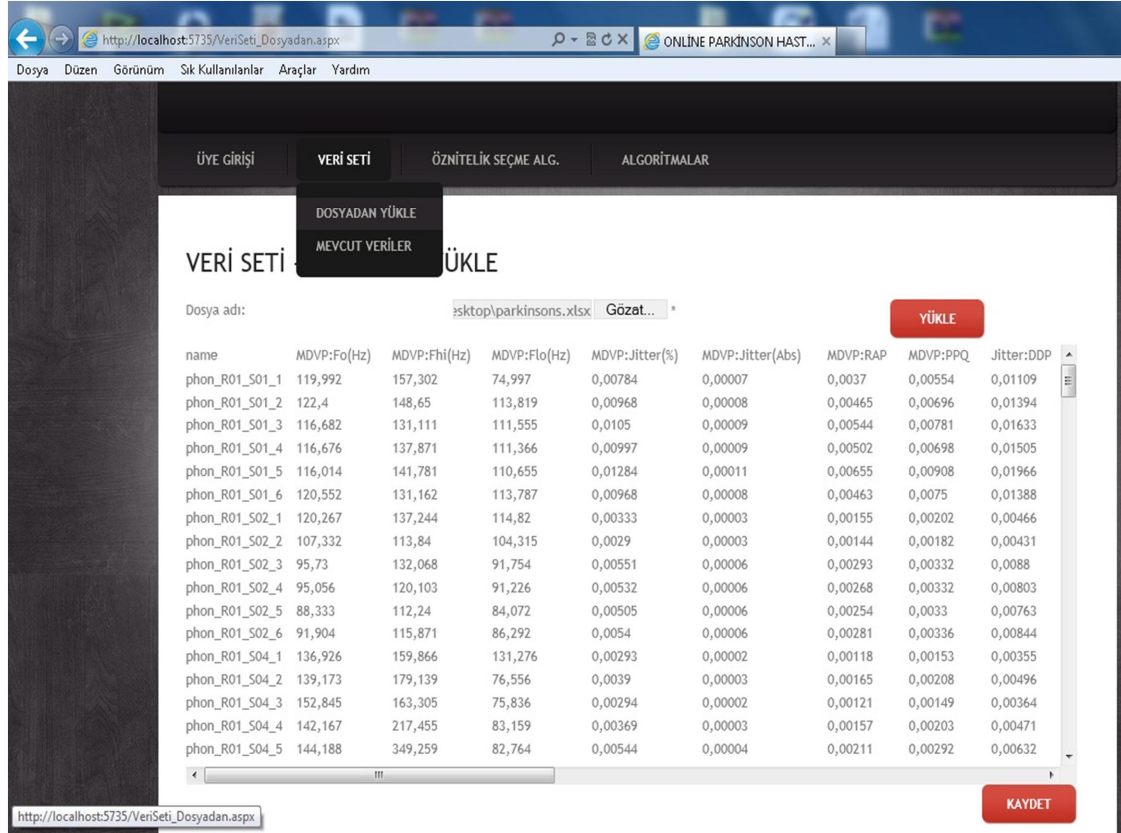
Hazırlanan kullanıcı web arayüzü Resim 4.1'deki gibi tasarlanmıştır. Bu web arayüz sayesinde öznitelik seçme ve sınıflandırma aşamaları gerçekleşmektedir. Tasarlanan web arayüzü dört farklı kısımdan oluşmaktadır. Web arayüzü çalıştırıldığında ekrana ilk olarak Resim 4.1'deki giriş sayfası gelmektedir.



Resim 4.1. Uygulama web arayüzü ve login ekranı

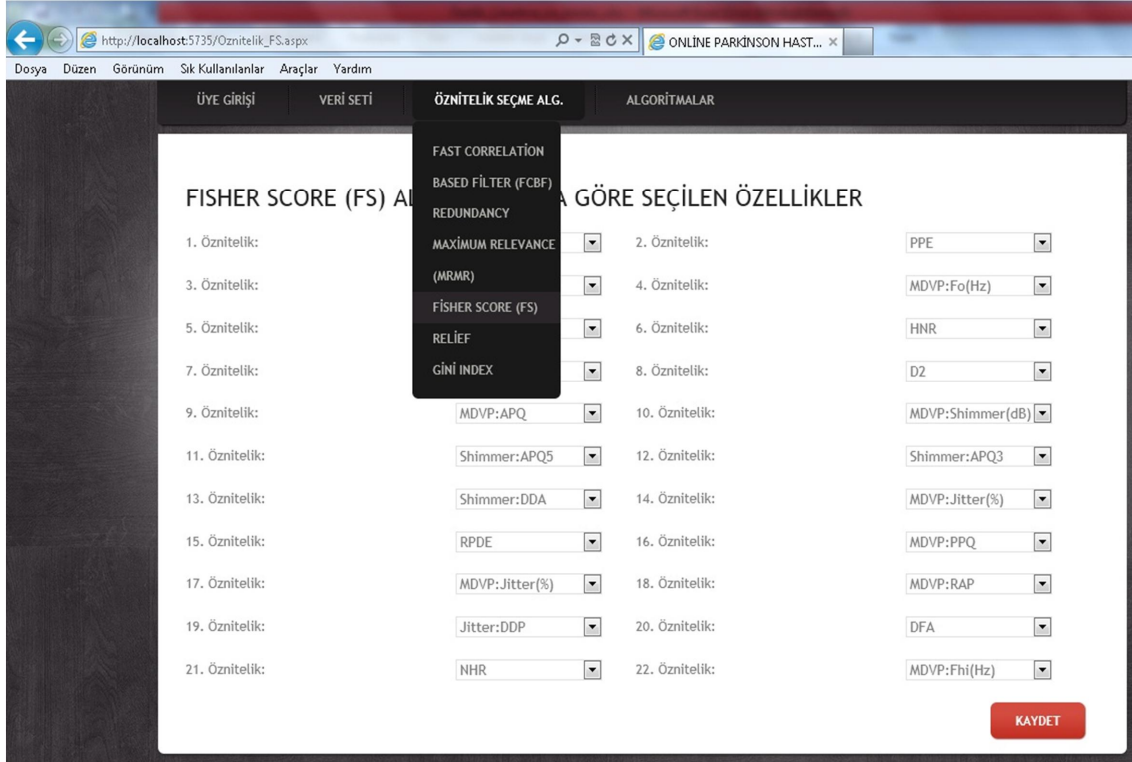
Kullanıcı adı ve şifre girildikten sonra “Giriş Yap” butonuna basıldığında Veri Seti menüsünden ister hazır verilerin kullanılması istenirse dosyadan veri seti yüklenmesi

şeklinde olmak üzere sayfa aktif hale gelmektedir. “Dosyadan Yükle” linkine tıklandığında veri setinin Resim 4.2’deki sayfa açılmakta ve “Gözet” butonuna tıklanmasıyla seçilen dosyadaki veriler ekranda gösterilmektedir. “Kaydet” butonuna tıklandığında veritabanına kaydedilmektedir.



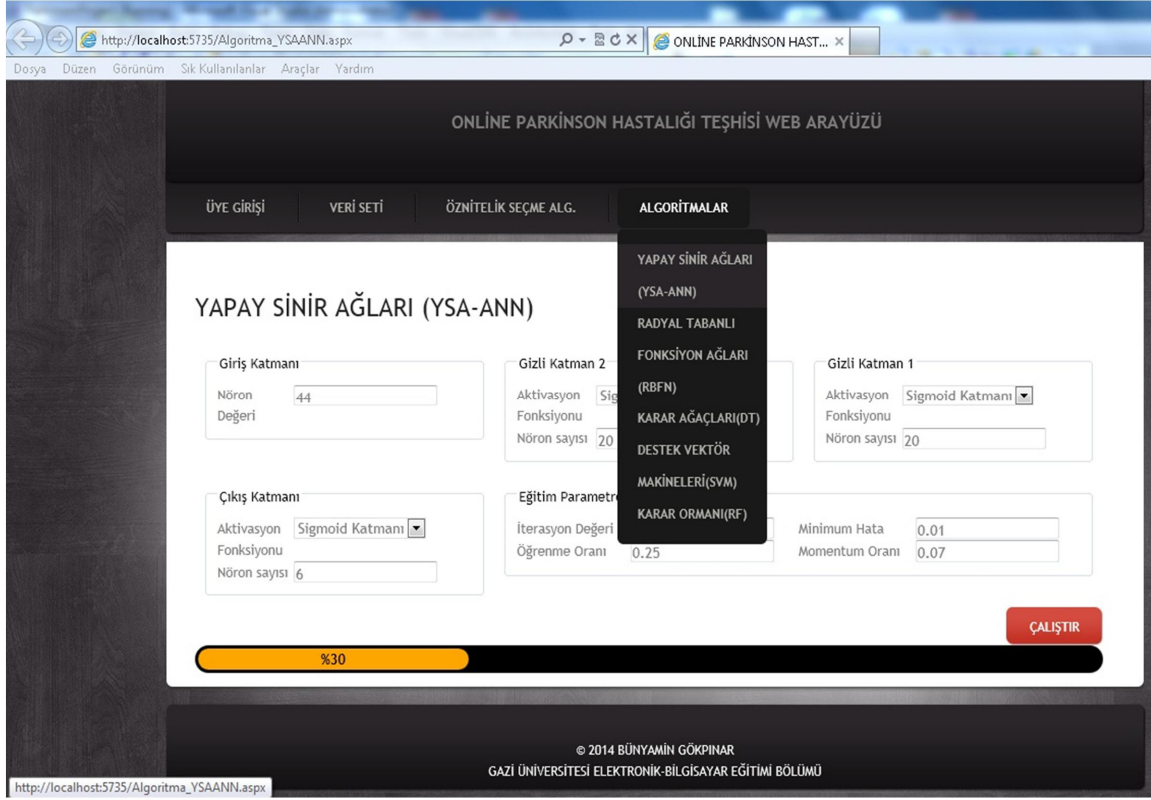
Resim 4.2. Veri setinin yüklenmesi

Yüklenen veri setinden sonra uygulanmak istenen Öznitelik seçme algoritması seçilmesi gerekmektedir. Öznitelik seçme için geliştirilen web arayüz sayfası Resim 4.3’te görüldüğü gibi “Öznitelik Seçme Alg.” Menüsünün üzerine gelince 5 farklı öznitelik seçme algoritması (Fast correlation based filter (FCBF)- Redundancy Maximum Relevance (mRMR) - Fisher Score (FS) – Relief - Gini Index) kullanılmaktadır. Seçilen öznitelik algoritmasına göre ekranda etkili öznitelikleri ve sırasını görmek mümkün olmaktadır. “Kaydet” butonuna tıklandığında belirlenen öznitelikler veritabanına kaydedilmektedir.



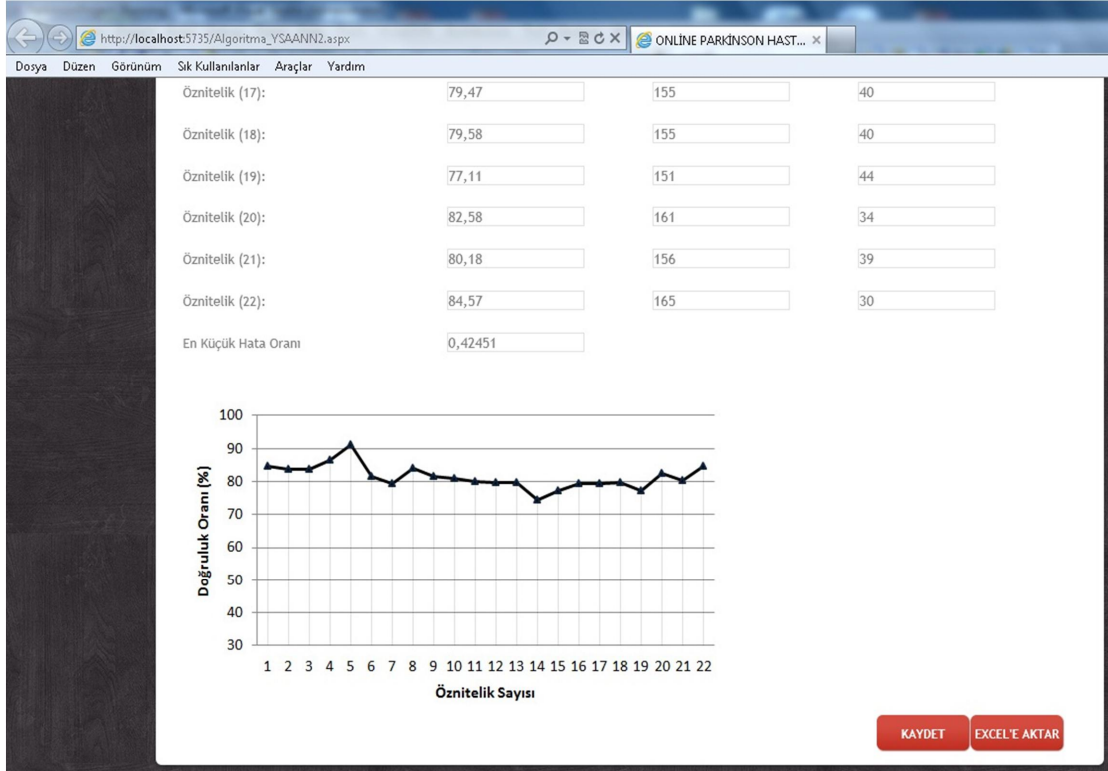
Resim 4.3. Öznitelik seçme algoritmasının seçilmesi

Son olarak seçilen özniteliklerin sınıflandırma aşamasında 5 farklı sınıflandırma algoritması (Yapay Sinir Ağları (YSA-ANN) - Radyal Tabanlı Fonksiyon Ağları (RBFN) - Karar Ağaçları (DT) - Destek Vektör Makineleri (SVM) - Rastgele Orman(RF)) bulunmaktadır. Kullanıcı hangi algoritma ile sınıflandırma işlemini gerçekleştirmek istiyorsa bunu “Algoritmalar” menüsünden ilgili linke tıklayarak gerçekleştirebilir. Resim 4.4’te gösterildiği gibi örneğin “Algoritmalar” menüsünden Yapay Sinir Ağları(YSA-ANN) linkine tıklandığında ekranda parametre ayarları yapılmaktadır. Gizli katman nöron sayıları, gizli katman aktivasyon fonksiyonları ve çıkış katmanı aktivasyon fonksiyonları seçilebilmektedir. Giriş katmanı nöron sayısı ve çıkış katmanı nöron sayısı seçilen veri türüne bağlı olarak otomatik seçilmektedir. Eğitim parametreleri alanında iterasyon sayısı, öğrenme oranı, minimum hata değeri ve momentum katsayısı değerleri girilebilmektedir.



Resim 4.4. Sınıflandırma algoritmasının seçilmesi

“Çalıştır” butonuna tıklandığında işlemler gerçekleştirilmeye başlamakta progressbar’ın %100 olmasıyla Resim 4.5’deki gibi sonuç sayfası ekrana gelmektedir. Burada seçilen öznelik algoritmasıyla önceliği belirlenen öznelikler sırasıyla işleme konulmakta başarı yüzdesi, doğru tahmin edilen kayıt sayısı, hatalı tahmin edilen kayıt sayısı ve en küçük hata değerleri görülebilmektedir. Bu başarı yüzdeleri de grafikte incelenebilmektedir. İstenirse sonuçlar “Kaydet” butonu ile veritabanına kaydedilmekte, “Excele Aktar” butonuyla da excele aktarılmaktadır.



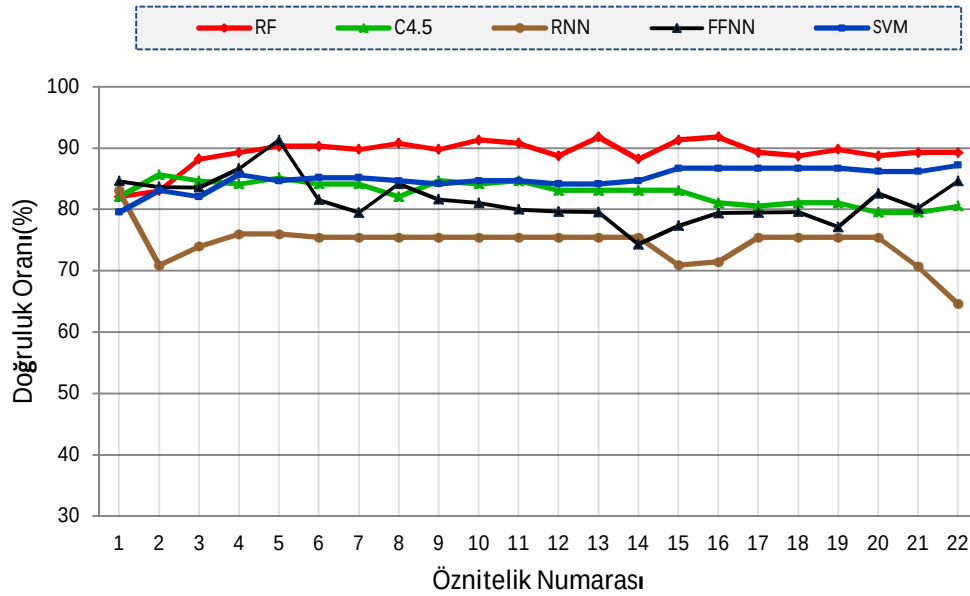
Resim 4.5. Seçilen sınıflandırma algoritmasına göre elde edilen sonuçların gösterilmesi

4.4. Yapılan Deneyler ve Sonuçların Değerlendirilmesi

Uygulanan sistemin değerlendirme aşamasında classification accuracy, sensitivity, specificity values and confusion matrix değerlerinden yararlanılmıştır. Bütün deneylerin eğitim ve test verilerinin seçimi 10-fold cross validation (CV) yöntemi ile gerçekleştirilmiştir. Deneylerde kullanılan öznitelik numaraları Bölüm-3'te verilen Çizelge 3.1'e göre gösterilmektedir. Aşağıdaki deneyler web arayüzü kullanılarak gerçekleştirilmiştir.

Deney-1: Fisher Skor Öznitelik Seçme Algoritması

Deney-1 aşamasında öznitelik kümesine, etkili öznitelik seçme algoritmalarından Fisher Score algoritması uygulanmıştır. Fisher Score algoritmasının uygulanması ile elde edilen etkili özniteliklerin sıralaması sırasıyla şu şekildedir: 19, 22, 3, 1, 20, 16, 9, 21, 13, 10, 12, 11, 14, 5, 17, 7, 4, 6, 8, 18, 15, 2. Şekil 4.8'de Fisher Score algoritması ile belirlenen öznitelik kümelerine ait sınıflandırma doğruluk oranları görülmektedir.



Şekil 4.3. Fisher Skor algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler

Şekil 4.3 incelendiğinde bazı algoritmalarda tek bir öznitelik değeri ile bile yüksek sınıflandırma doğruluklarının elde edildiği görülmektedir. Bu seçilen öznitelik kümelerinin etkili olduğu göstermektedir. Grafik incelendiğinde genel olarak en iyi sonucun Rastgele Orman algoritması ile elde edildiği görülmektedir. Rastgele orman algoritmasını destek vektör makinelerinin takip ettiği görülmektedir. Yer yer ileri beslemeli sinir ağı ve karar ağaçlarının da iyi sonuçlar verdiği görülmektedir.

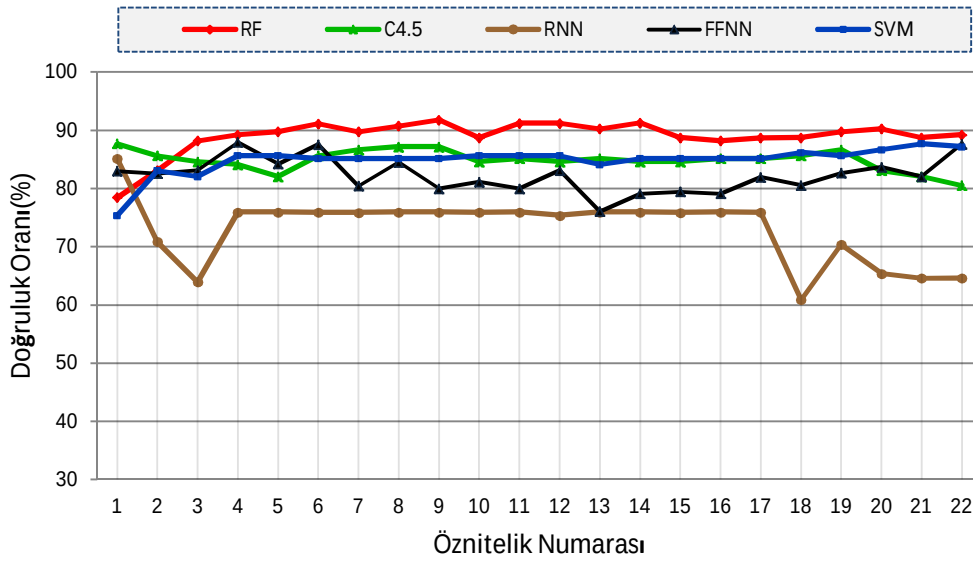
Deney-1 aşamasında en iyi sonuç; Rastgele Orman Algoritması ve ilk 13 özniteliğin birlikte kullanılması ile elde edilmiştir. İlk 13 özniteliğin Rastgele Orman algoritması ile %91.79 doğruluk oranı elde edilmiştir. Bu 13 öznitelik algoritması sırasıyla şu şekildedir: spread1, PPE, MDVP:Flo(Hz), MDVP:Fo(Hz), spread2, HNR, MDVP:Shimmer, D2, MDVP:APQ, MDVP:Shimmer(dB), Shimmer:APQ, Shimmer:APQ3, Shimmer:DDA.

Çizelge 4.1. Fisher Skor algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları

Methot	Öznitelik Sayısı	Sınıflandırma Yüzdesi (%)
RF	13	91.79
FFNN	5	91.28
SVM	22	87.17
C4.5	2	85.64
RNN	1	83

Deney-2: Gini İndeks Öznitelik Seçme Algoritması

Deney-2 aşamasında öznitelik kümesine, etkili öznitelik seçme algoritmalarından Gini indeks algoritması uygulanmıştır. Gini indeks algoritmasının uygulanması ile elde edilen etkili özniteliklerin sıralaması sırasıyla şu şekildedir: 22, 19, 3, 1, 5, 13, 20, 6, 8, 12, 15, 7, 9, 4, 10, 11, 14, 2, 16, 17, 21, 18. Şekil 4.4’de Gini İndeks algoritması ile belirlenen öznitelik kümelerine ait sınıflandırma doğruluk oranları görülmektedir.



Şekil 4.4. Gini indeks algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler

Şekil 4.4 incelendiğinde bazı algoritmalarda tek bir öznitelik değeri ile bile yüksek sınıflandırma doğruluklarının elde edildiği görülmektedir. Bu seçilen öznitelik kümelerinin etkili olduğu göstermektedir. Grafik incelendiğinde genel olarak en iyi sonucun Rastgele Orman algoritması ile elde edildiği görülmektedir. Rastgele orman algoritmasını destek vektör makineleri ve karar ağacının takip ettiği görülmektedir.

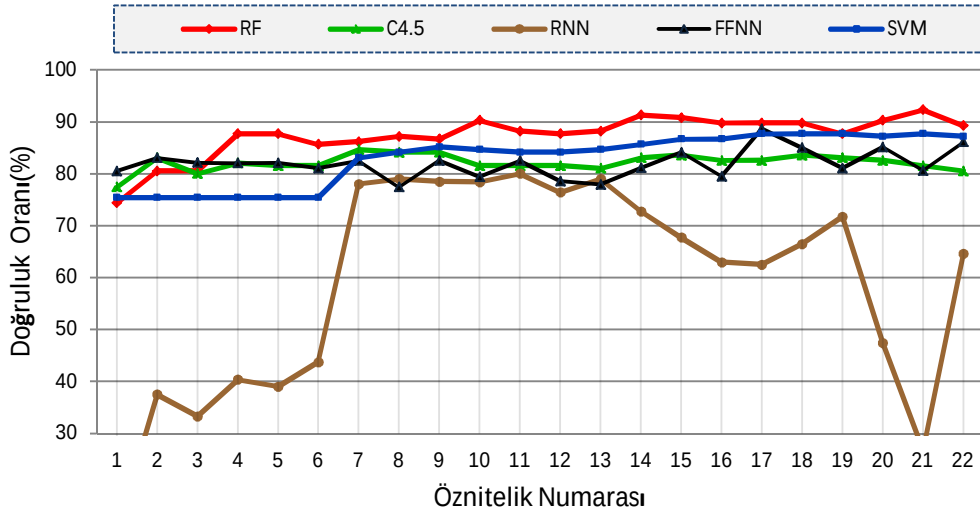
Deney-2 aşamasında en iyi sonuç; Rastgele Orman Algoritması ve ilk 9 özniteliğin birlikte kullanılması ile elde edilmiştir. İlk 9 özniteliğin Rastgele Orman algoritması ile %91.79 doğruluk oranı elde edilmiştir. Bu 9 öznitelik algoritması sırasıyla şu şekildedir: PPE, spread1, MDVP:Flo(Hz), MDVP:Fo(Hz), MDVP:Jitter(Abs), MDVP:APQ, spread2, MDVP:RAP, Jitter:DDP.

Çizelge 4.2. Gini indeks algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları

Methot	Öznitelik Sayısı	Sınıflandırma Yüzdesi (%)
RF	9	91.79
FFNN	4	87.95
C4.5	1	87.69
SVM	21	87.69
RBFN	1	85.13

Deney-3: mRMR Öznitelik Seçme Algoritması

Deney-3 aşamasında öznitelik kümesine, etkili öznitelik seçme algoritmalarından mRMR algoritması uygulanmıştır. mRMR algoritmasının uygulanması ile elde edilen etkili özniteliklerin sıralaması sırasıyla şu şekildedir: 2, 4, 5, 6, 7, 8, 19, 9, 11, 12, 13, 14, 15, 1, 17, 18, 20, 22, 10, 21, 3, 16. Şekil 4.10'da mRMR algoritması ile belirlenen öznitelik kümelerine ait sınıflandırma doğruluk oranları görülmektedir.



Şekil 4.5. mRMR algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler

Şekil 4.5 incelendiğinde bazı algoritmalarda tek bir öznitelik değeri ile bile yüksek sınıflandırma doğruluklarının elde edildiği görülmektedir. Bu seçilen öznitelik kümelerinin etkili olduğu göstermektedir. Grafik incelendiğinde ilk 3 öznitelik değerinden oluşan öznitelik kümesinde ileri beslemeli sinir ağının etkili olduğu görülmektedir. Genel olarak

en iyi sonucun Rastgele Orman algoritması ile elde edildiği görülmektedir. Rastgele orman algoritmasını Destek Vektör Makineleri ve karar ağacının takip ettiği görülmektedir.

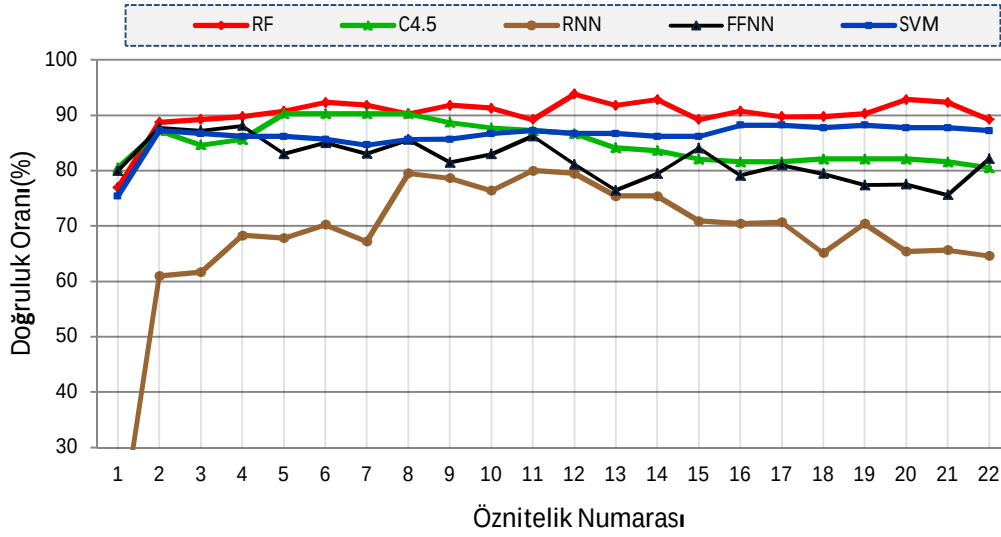
Deney-3 aşamasında en iyi sonuç; Rastgele Orman Algoritması ve ilk 21 özneliğin birlikte kullanılması ile elde edilmiştir. İlk 21 özneliğin Rastgele Orman algoritması ile %92.3 doğruluk oranı elde edilmiştir. Bu 21 öznelik algoritması sırasıyla şu şekildedir: MDVP:Fhi(Hz), MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP, spread1, MDVP:Shimmer, Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA, NHR, MDVP:F0(Hz), RPDE, DFA, spread2, PPE, MDVP:Shimmer(dB), D2, MDVP:Flo(Hz).

Çizelge 4.3. mRMR algoritmasıyla seçilen öznelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları

Methot	Öznelik Sayısı	Sınıflandırma Yüzdesi (%)
RF	21	92.3
FFNN	17	88.63
SVM	17	87.69
C4.5	7	84.61
RBFN	11	80.02

Deney-4: ReliefF Öznelik Seçme Algoritması

Deney-4 aşamasında öznelik kümesine, etkili öznelik seçme algoritmalarından ReliefF algoritması uygulanmıştır. ReliefF algoritmasının uygulanması ile elde edilen etkili özneliklerin sıralaması sırasıyla şu şekildedir: 1, 19, 22, 18, 11, 14, 9, 20, 12, 17, 10, 16, 3, 21, 13, 2, 7, 6, 8, 5, 4, 15. Şekil 4.11’de ReliefF algoritması ile belirlenen öznelik kümelerine ait sınıflandırma doğruluk oranları görülmektedir.



Şekil 4.6. ReliefF algoritmasıyla seçilen özniteliklerin sayısının artması ile sınıflandırma doğruluğundaki değişimler

Şekil 4.6 incelendiğinde bazı algoritmalarda tek bir öznitelik değeri ile bile yüksek sınıflandırma doğruluklarının elde edildiği görülmektedir. Bu seçilen öznitelik kümelerinin etkili olduğu göstermektedir. Grafik incelendiğinde genel olarak en iyi sonucun Rastgele Orman algoritması ile elde edildiği görülmektedir. Rastgele orman algoritmasını Destek Vektör Makineleri ve karar ağacının takip ettiği görülmektedir.

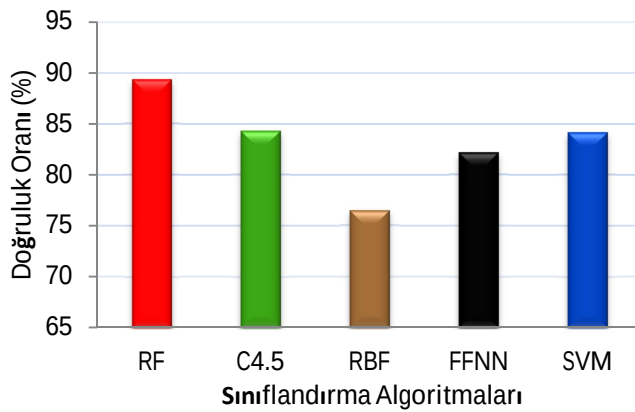
Deney-4 aşamasında en iyi sonuç; Rastgele Orman Algoritması ve ilk 12 özneliğin birlikte kullanılması ile elde edilmiştir. İlk 12 özneliğin Rastgele Orman algoritması ile % 93.84 doğruluk oranı elde edilmiştir. Bu 12 öznitelik algoritması sırasıyla şu şekildedir: MDVP:Fo(Hz), spread1, PPE, DFA, Shimmer:APQ3, Shimmer:DDA, MDVP:Shimmer, spread2, Shimmer:APQ5, RPDE, MDVP:Shimmer(dB), HNR.

Çizelge 4.4. ReliefF algoritmasıyla seçilen öznitelik sayısına göre tüm yöntemlerde elde edilen en iyi başarı oranları

Methot	Öznitelik Sayısı	Sınıflandırma Yüzdesi (%)
RF	12	93.84
C4.5	5	90.25
SVM	16	88.2
FFNN	4	88.1
RNN	11	80

Deney-5: FCBF Öznitelik Seçme Algoritması

Deney-5 aşamasında öznitelik kümesine, etkili öznitelik seçme algoritmalarından FCBF algoritması uygulanmıştır. FCBF algoritması eleme usulü çalışmaktadır. Bu nedenle bahsi geçen diğer öznitelik seçme algoritmalarından farklı olarak tek bir öznitelik kümesi elde edilmektedir. FCBF algoritmasının uygulanması ile elde edilen etkili özniteliklerin sıralaması sırasıyla şu şekildedir: 22, 3, 13, 2, 21. Şekil 4.12’de FCBF algoritması ile belirlenen öznitelik kümelerine ait sınıflandırma doğruluk oranları görülmektedir.



Şekil 4.7. FCBF algoritmasıyla seçilen özniteliklerin sınıflandırma doğruluğundaki değişimler

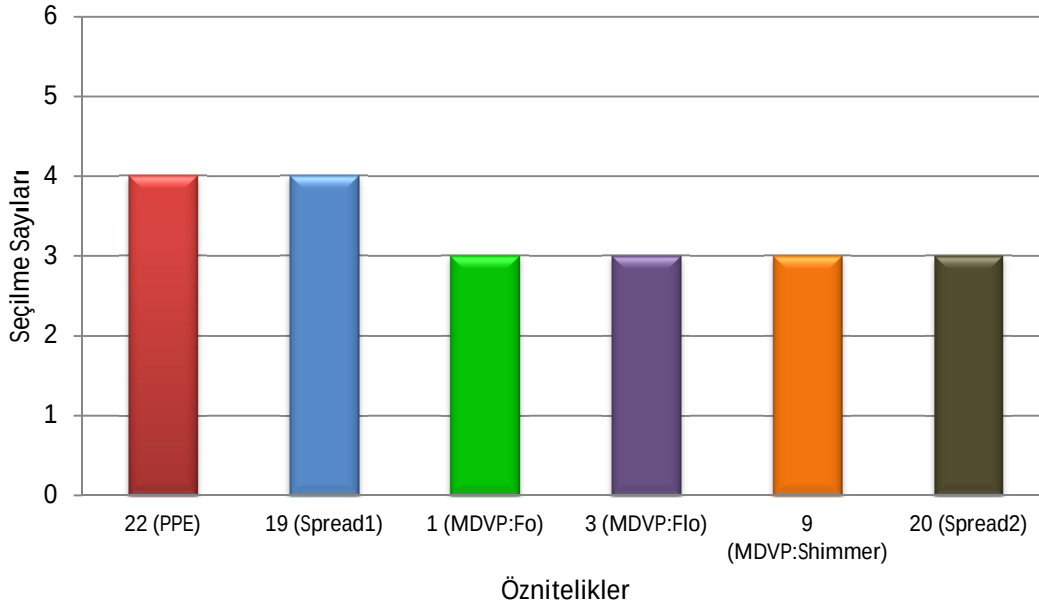
Deney-5 aşamasında en iyi sonuç; Rastgele Orman Algoritması ve 5 özniteliğin birlikte kullanılması ile elde edilmiştir. 5 özniteliğin Rastgele Orman algoritması ile % 89.28 doğruluk oranı elde edilmiştir. Bu 5 öznitelik kümesi sırasıyla şu şekildedir: PPE, MDVP:Flo(Hz), MDVP:APQ, MDVP:Fhi(Hz), D2.

Çizelge 4.5. FCBF algoritmasıyla seçilen öznitelik kümesine göre tüm yöntemlerde elde edilen en iyi başarı oranları

Methot	Öznitelik Sayısı	Sınıflandırma Yüzdesi (%)
RF	5	89.28
C4.5	5	84.26
SVM	5	84.1
FFNN	5	82.1
RNN	5	76.42

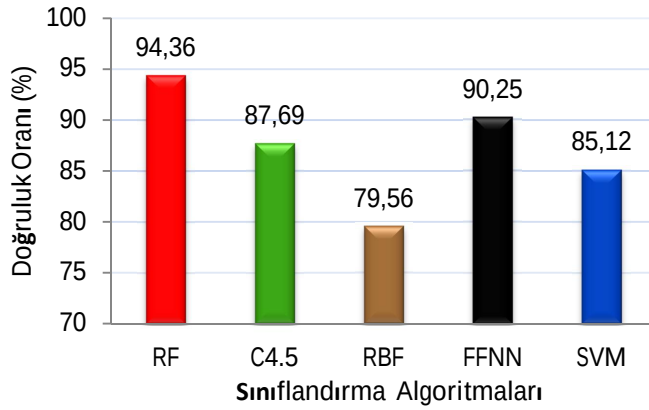
Deney-6: Hibrit Yaklaşım

Çalışmada gerçekleştirilen 5 deneyde, 5 farklı öznitelik algoritmasına göre en etkili öznitelik sıralamaları elde edildi. Bu önem sırasına göre belirlenen özniteliklerle deneyler gerçekleştirildi. Deney-6'nın yapılmasında hibrit bir yaklaşım amaçlanmaktadır. Bu yaklaşımda, 5 farklı deney sonucunda sıralanan etkili özniteliklerden, öznitelik sayısının $1/3$ 'ü yani ilk 8 arasında tekrarlanan özniteliklerin sayısı tespit edildi. Buna göre her algoritma sonucu elde edilen ve 3 veya daha fazla seçilen 6 öznitelikten yeni bir öznitelik kümesi oluşturuldu. Bu elde edilen yeni öznitelik kümesine göre deneyler gerçekleştirildi. Elde edilen öznitelikler ve seçilme sayılarını gösteren histogram grafiği Şekil 4.8'de görülmektedir.



Şekil 4.8. Öznitelik seçme algoritmaları uygulamasıyla en çok seçilen öznitelikler

Şekil 4.8'de görüldüğü gibi Deney-6 aşamasında 6 öznitelik algoritması seçilmiştir. Bunlar sırasıyla PPE, Spread1, MDVP:F0(Hz), MDVP:F0(Hz), MDVP:Shimmer ve Spread2 öznitelik değerleridir. Bu 6 öznitelik algoritması 5 farklı sınıflandırma algoritmasına giriş olarak sunulmuştur. Deney 6 aşamasında en iyi sonuç Rastgele Orman algoritması ile %89.28 doğruluk oranı elde edilmiştir. Elde edilen başarı oranları Şekil 4.9'da sunulmaktadır.



Şekil 4.9. Hibrit yaklaşımla seçilen özniteliklerin sınıflandırma doğruluğundaki değişimler

Deneyler incelendiğinde bazı algoritmalarda tek bir öznitelik değeri ile bile yüksek sınıflandırma doğruluklarının elde edildiği görülmektedir. Bu seçilen öznitelik kümelerinin etkili olduğu göstermektedir. Sonuçlara bakıldığında genel olarak en iyi sonucun Rastgele Orman algoritması ile elde edildiği görülmektedir. Rastgele Orman algoritmasını Karar Ağacı ve Destek Vektör Makinelerinin takip ettiği görülmektedir.

6 deney incelediğinde en iyi sonucun hibrit yaklaşımla elde edilen 6 öznitelik değerinden oluşan öznitelik kümesi ve RF algoritmasının beraber kullanılmasıyla elde edilmiştir. Deney-6 aşamasında % 94.36 doğruluk oranı elde edilmiştir.

Sonuç olarak deney sonuçlarına göre başarı sıralamasında en iyi algoritmanın RF algoritması olduğu görülmektedir. RF algoritmasını takiben FFNN, SVM, C4.5 ve RBFN algoritmalarının farklı öznitelik kümelerinde etkili sonuçlar verdiği görülmektedir.

Çizelge 4.6. Yapılan deneylerde tüm yöntemlerde elde edilen en iyi başarı oranları

Deney Adı	Öznitelik Seçme Algoritması	Sınıflandırma Algoritması	Öznitelik Sayısı	Doğruluk Oranı (%)
Deney 1	Fisher Score	RF	13	91.79
Deney 2	Gini İndex	RF	9	91.79
Deney 3	mRMR	RF	21	92.3
Deney 4	Relief	RF	12	93.84
Deney 5	FCBF	RF	5	89.28
Deney 6	Hibrit Yaklaşım	RF	6	94.36

Ayrıca bulunan en iyi sonuçla ilgili confusion matrix, sensitivity ve specificity değerleri [56] aşağıdaki tabloda sunulmaktadır.

Çizelge 4.7. En iyi sonucun confusion matrix, accuracy, sensitivity ve specificity değerleri

	Sağlıklı	Parkinson Hastası
Sağlıklı	40	8
Parkinson Hastası	3	144
Accuracy (Doğruluk)	94.36 %	
Sensitivity(Duyarlılık)	94.74 %	
Specificity(Özgüllük)	93.02 %	

Yapılan çalışmada %94.36'lık bir doğruluk değeri ile diğer çalışmalarla kıyaslanabilir olduğu Çizelge 4.8'de görülmektedir.

Çizelge 4.8. En iyi sonucun diğer yapılan çalışmalara kıyaslanması

Çalışma	Yöntem	Doğruluk (%)
Little et al. (2009)	Pre-selection filter + Exhaustive search + SVM	91.4 (bootstrap with 50 replicates)
Shahbaba and Neal (2009)	Dirichlet process mixtures	87.7 (5-fold CV)
Das (2010)	ANN	92 (hold-out)
Sakar and Kursun (2010)	Mutual information based feature	92.75 (bootstrap selection + SVM with 50 replicates)
Psorakis et al. (2010)	Improved mRVMs	89.47 (10-fold CV)
Guo et al. (2010)	GP-EM	93.1 (10-fold CV)
Ozcift and Gulten (2011)	CFS-RF	87.1 (10-foldCV)
Li et al. (2011)	Fuzzy-based non-linear Transformation+SVM	93.47 (hold-out)
AStröm and Koker (2011)	Parallel NN	91.20 (hold-out)
Önerilen Yaklaşım	22 (PPE), 19 (Spread1), 1 (MDVP:F0), 3 (MDVP:F0), 9 (MDVP:Shimmer), 20 (Spread2) + RF	94.36 (10-fold CV)

5. SONUÇ VE ÖNERİLER

Ülkemizde, Parkinson hastalığı teşhisi ile ilgili yapılan çalışmalar sınırlı, bilgisayar destekli karar sistemlerinin hastanelerde kullanımı da pek yaygın değildir. Parkinson teşhisi ile ilgili 2000’li yıllarda yurt dışında çalışmalar yapılmaya başlasa da, ses ve yazı ölçümlerini kullanarak hastaların Parkinson hastalığıyla ilgilerini tespit edecek olan çalışma oldukça az sayıdadır.

Bu tezde, hastaların biyomedikal ses parametreleri kullanarak, bu verilerin istatistiki analizini yapan ve böylece doktorlara bilgisayarlı bir karar destek sistemi olarak kullanılabilecekleri bir sistem geliştirilmiştir.

Tüm özniteliklerin kullanılması da artık veya sınıf hakkında hiçbir bilgi içermeyen değerleri barındırabileceğinden kullanılabilir özniteliklerin de ayrıştırılması gerekmektedir. Bu sebeple veri kümesine öznitelik seçme algoritmaları uygulanarak veriler önem sırasına göre oluşturulmuştur. Bu algoritmalar sırasıyla; Fisher Score, ReliefF, FCBF, mRMR ve Gini Index algoritmalarıdır. Oluşturulan farklı öznitelikler farklı beş yöntemle sınıflandırma yapılmıştır. Bu tez çalışması kapsamında sınıflandırma algoritması olarak rastgele orman, karar ağacı, radyal sinir ağı, ileri beslemeli sinir ağı ve destek vektör makinaları tercih edilmiştir. Sınıflandırma algoritmaları beş farklı öznitelik seçme yöntemi ile beslenmiş ve en ideal öznitelik seçme yöntemi bulunması amaçlanmıştır.

En iyi özniteliklerin sayısı birer birer artırılarak rastgele orman, karar ağacı, radyal sinir ağı, ileri beslemeli sinir ağı ve destek vektör makinaları başarı oranları gözlemlenmiştir. Deneylerin sonucuna göre en iyi sonucun hibrit yaklaşım ile elde edilen ilk 6 öznitelik değerinden oluşan öznitelik kümesinin RF algoritmasının beraber kullanımı ile %94.36 doğruluk oranı elde edilmiştir. Bu oranlara göre öznitelik seçiminin başarı performansına olumlu yönde katkısı olduğu söylenebilmektedir. Bu deneyler sonunda elde edilen sonuçlar, önerilen ses örneklerinden Parkinson hastalığı teşhis yöntemini literatürdeki diğer çalışmalar ile kıyaslanabilir olduğunu göstermiştir.

Sonuç olarak, farklı yöntemler kullanılarak ses verilerinden hastalık bilgilerinin tespitini gösterebilecek bir çalışma yapılmıştır. Farklı özniteliklerin bulunması ve farklı

sınıflandırma yöntemleri kullanılması sistemin başarı performansını olumlu yönde etkileyecektir.

Çalışma hem problemin temsili için kullanılan öznitelikler içerisinde en etkili özniteliklerin tespiti hem de seçilen beş popüler sınıflandırma algoritmalarından problemin sınıflandırılması için en etkili olan algoritmanın tespiti yönüyle de farklı bir analiz sunmaktadır. Başta ülkemizde olmak üzere, dünyada da birçok çalışmaya referans olacak niteliktedir. Elde edilen verilerin hastalığın teşhis ve tedavisinde yol gösterici olabilmesi için kullanılan farklı örüntü tanıma algoritmalarından daha iyi sonuçlar alabilmek için ses ve bir ileriki aşama olan el yazısı gibi verilerin bir arada kullanılması ve kullanılan algoritmaların geliştirilmesi ve Parkinson verilerine uyarlanması yeni yöntem ve algoritmaların çıkışına ışık tutacaktır.

Ayrıca çalışmamda her ne kadar sadece Parkinson hedef alınmış olsa da, geliştirilen sistem Alzheimer gibi diğer nörolojik hastalıklarda da kullanılabilecektir.

Ses verilerinden öznitelik çıkartılmasında öznitelik seçme ve sınıflandırma algoritmalarında MATLAB teknik programlama dili ve web arayüzü için .NET platformu kullanılmıştır. Matlab Builder NE ve Matlab Web Figure araçları kullanılarak .NET platformunda Asp.NET tabanlı bir web tasarımı yapılmıştır. Böylece hem Matlab programının yeni versiyonlarına hem de her geçen gün yaygınlaşan ve gelişen .NET teknolojisine uyumlu bir arayüz oluşturulmuştur. Yapılan çalışmanın sunduğu en büyük avantaj Matlab programının özellikleri kullanılırken sunucu bilgisayarında Matlab programının kurulu olması gereksinimini ortadan kaldırılmasıdır. Ayrıca çalışmanın bir diğer üstünlüğü ise Matlab Web Figure aracının kullanıcılara sağlamış olduğu dinamik ve esnek yapıdır.

Sunulan uygulama bir prototip ve başlangıç çalışması olup, gelecek çalışmalarda daha farklı analiz sonuçları, grafik ve karşılaştırma seçenekleri üzerine odaklanması düşünülmektedir.

KAYNAKLAR

1. Apaydın, H. (2001). Parkinson hastalığının klinik özellikleri. *Parkinson Hastalığı ve Hareket Bozuklukları Dergisi*, 4(2), 114-24.
2. Mauro, T., Saitoh, Y., Hosomi, K., Kishima, H., Shimokawa, T., Hirata, M., Goto, T., Morris, S., Harada, Y., Yanagisawa, T., Aly, M.M., Yoshimine, T. (2011). Deep brain stimulation of the subthalamic nucleus improves temperature sensation in patients with Parkinson's disease. *PAIN*, 860-865.
3. 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.
4. Little, M. A., MCSarry, 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, 1015-1022.
5. Hoehn, M.M., Yahr, M.D. (1967). Parkinsonism: onset, progression and mortality. *Neurology*, 17, 427-42.
6. Tugwell, C. (2008). *Parkinson's Disease on Focus*. Chicago: Pharmaceutical Press, 7-8.
7. Aydın, I. (2005). Ağrının Patofizyolojisi. *Türk Fiz. Tıp Rehab. Derg.* , 51 (özel ek B), 8-13.
8. Ford, B., Louis, E.D., Greene, P. and Fahn, S. (1996). Oral and genital pain syndromes in Parkinson's disease. *Mov Disord*, 11(4), 421-6.
9. Fahn, S. And Przedborski, S. (2000). *Parkinsonizm : In Rowland LP ed. Merritt's Textbook Of Neurology* (10nd edition). Philadelphia: Lippincott Williams & Wilkins, 679-693.
10. Apaydın, H., Özekmekci, S. (2008). *Parkinson Hastalığı: Hasta ve Yakınları için El Kitabı*. İstanbul: Parkinson Hastalığı Derneği, 6-15.
11. Ertan, S. (2005). Parkinson hastalığının klinik özellikleri. *Nörolog Olmayanlar için Nöroloji*, 42, 249-254.
12. Delil, Ş. (2001). *Parkinson Hastalığında Levodopa'ya Bağlı Motor Komplikasyonların Gelişmesinde Rol Oynayan Etkenler*, Uzmanlık Tezi, İstanbul Üniversitesi, İstanbul, 17-18.
13. Dewey, R.B. (2000). *Jr. Clinical Features of Parkinson's Disease. In: Parkinson's Disease and Movement Disorder, Ahlskog*. New Jersey: Humana Press, 71-84.
14. İsenkul, M. E. (2011). *Parkinson Hastalığı'nın Teşhisi İçin Veri Toplama ve Örüntü Tanıma Sistemi*, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

15. İnternet: Parker, J. Speech Dysfunctions In Parkinson's Disease. [http://www.webcitation.org/query?url=http%3A%2F%2Fwww.selfgrowth.com%2Farticles%2FSpeech Dysfunctions In Parkinson s Disease.html&date=2014-04-29](http://www.webcitation.org/query?url=http%3A%2F%2Fwww.selfgrowth.com%2Farticles%2FSpeech%20Dysfunctions%20In%20Parkinson%20s%20Disease.html&date=2014-04-29), Son Erişim Tarihi: 29.04.2014.
16. 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.
17. 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*.
18. Revett, 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.
19. 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.
20. İnternet: UCI Machine Learning Repository. Parkinson Dataset. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Farchive.ics.uci.edu%2Fml%2Fdatasets%2FParkinsons&date=2014-04-29>, Son Erişim Tarihi: 29.04.2014.
21. Caglar, M.F., Cetisli, B. And Toprak, I.B. (2010). Automatic recognition of Parkinson's disease from sustained phonation tests using ANN and adaptive neuro-fuzzy classifier. *SDU Journal of Engineering and Design*, 1(2), 59-64.
22. Kılıç, M. A. ve Okur, E. (2001). CSL ve Dr. Speech ile ölçülen temel frekans ve pertürbasyon değerlerinin karşılaştırılması. *KBB İhtisas Dergisi*, 8(2), 152-157.
23. 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.
24. 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)*, Washington DC.
25. Gülgezen, G. (2009). *Kararlı ve başarıımı yüksek öznitelik seçimi*, Yüksek Lisans Tezi, İstanbul Teknik Fen Bilimleri Enstitüsü, İstanbul, 15-17.
26. Press, W.H., Flannery, B.P., Teukolsky, S.A. And Vetterling, W.T. (1992). *Numerical Recipes in C* (2nd edition). Cambridge: Cambridge University Press.
27. Quinlan, J.R. (1993). *C4.5: Programs For Machine Learning*. San Mateo: Morgan Kaufmann.

28. Ding, C. and Peng, H.C. (2003). Minimum redundancy feature selection from microarray gene expression data. *Proc. Second IEEE Computational Systems Bioinformatics Conf.*, Stanford, 523-528.
29. Fröhlich, H., Chapelle, O. and Schölkopf, B. (2004). Feature selection for support vector machines using genetic algorithms. *International Journal on Artificial Intelligence Tools*, 13(4), 791-800.
30. Kira, K. And Rendell, L.A. (1992). The feature selection problem: traditional methods and a new algorithm. *In Proceedings of the 9th National Conference on Artificial Intelligence*, California, 129-134.
31. Last, M., Bunke, H. and Kandel, A. (2002). A feature-based serial approach to classifier combination. *Springer-Verlag London Limited*, 5(4), 385-395.
32. Kononenko, I. (1994). Estimating attributes: analysis and extensions of RELIEF. *European Conference on Machine Learning*, Italy, 171-182.
33. Civelekoğlu, G. (2006). *Arıtma proseslerinin yapay zekâ ve çoklu istatistiksel yöntemler ile modellenmesi*, Doktora Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Isparta, 33-44.
34. Tektaş, M., Akbaş, A. ve Topuz, V. (2002). Yapay zeka tekniklerinin trafik kontrolünde kullanılması üzerine bir inceleme. *I. Uluslararası Trafik ve Yol Güvenliği Kongresi*, Ankara, 4-7.
35. Perendeci, A. (2004). *Şeker fabrikası anaerobik atık su arıtma tesisinin yatışkın olmayan durumda modellenmesi çalışmaları*, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
36. Bilgin, S. (2008). *Kalp Hızı Değişkenliğinin Dalgacık Dönüşümü Ve Yapay Sinir Ağları Kullanarak Analizi*, Doktora Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
37. Öztemel, E. (2003). *Yapay Sinir Ağları*. İstanbul: Papatya Yayıncılık.
38. Çelik, E. (2011). *Görüntü İşlemeye Dayalı Avuç İçi İzinin Yapay Sinir Ağı İle Tanınması*, Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
39. Haykin, S.S. (2005). *Neural Networks: A Comprehensive Foundation*, USA: Pearson Prentice Hall.
40. Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the Brain. *Psychoanalytic Review*, 65, 386-408.
41. Rumelhart, D., Hinton, G. and Williams, R. (1986). Learning representations by backpropagation errors. *Nature*, 323, 533-536.

42. Hopfield, J.J. And Tank, D.W. (1985). Neural computation of fecisions in optimization problems. *Biological Cybernetics*, 52, 141-152.
43. Elmas, M. (2012). *Destek Vektör Makineleri İle Fiyat Tahminleri Ve Kuyumculuk Sektöründe Bir Uygulama*, Doktora Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
44. Şen, Z. (2004). *Yapay Sinir Ağları İlkeleri*. İstanbul: Su Vakfı Yayınları.
45. Mahanty, R.N. And Gupta, P.B.D. (2004). Application of RBF neural network to fault classification and location in transmission lines. *IEEE Proceedings of Gener. Transm. Distrib.*, 151, 201-212.
46. Pala, T. (2013). *Tıbbi Karar Destek Sisteminin Veri Madenciliği Yöntemleriyle Gerçekleştirilmesi*, Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
47. Kaban, Z. (2008). *Yapay Bağışıklık Sistemlerini Kullanarak Türkçe Metinlerde Tür, Yazar Ve Cinsiyet Tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
48. Bolat, B., Küçük, Ü., Yıldırım, T. (2004). Aktif Öğrenen PNN İle Konuşma/Müzik Sınıflandırma. *Akıllı Sistemlerde Yenilikler Ve Uygulamalar Sempozyumu*, 187-189.
49. Gürsoy, U. T. (2009). *Veri Madenciliği ve Bilgi Keşfi*. Ankara: Pegem Akademi.
50. Yalçın, Ö. (2008). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.
51. İnternet: Menon, A. K. Large-Scale Support Vector Machines: Algorithms and Theory. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fftp%3A%2F%2Ffcseweb.ucsd.edu%2F%7Eakmenon%2FResearchExam.pdf&date=2014-04-29>, Son Erişim Tarihi: 29.04.2014.
52. Breiman, L. (2001). Random forests. *Kluwer Academic Publishers*, 45, 1-30.
53. Akman, M. (2010). *Veri madenciliğine genel bakış ve random forests yönteminin incelenmesi: sağlık alanında bir uygulama*, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, 33-35.
54. İnternet: The Mathworks Inc.. Matlab Builder NE 3 User's Guide. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.mathworks.com%2Fproducts%2Fdatasheets%2Fpdf%2Fmatlab-builder-ne.pdf&date=2014-04-29>, Son Erişim Tarihi: 29.04.2014.
55. Kaçar, S., Bayılmış, C., Çankaya, İ., Çakıroğlu, M. (2009). Kablosuz Algılayıcı Ağlar İçin Matlab Builder Ne Ve Matlab Webfigure İle Asp.Net Tabanlı Web Arayüzü Tasarımı. *New World Sciences Academy*, 4, 360-370.

56. İnternet: Wikipedia. Sensitivity and specificity. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fen.wikipedia.org%2Fwiki%2FSensitivity_and_specificity&date=2014-04-29, Son Erişim Tarihi: 29.04.2014.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : GÖKPINAR, Bünyamin
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 10.03.1983, Beykoz/İSTANBUL
 Medeni hali : Evli
 Telefon : 0 (506) 930 74 30
 e-posta : bunyamin.gokpinar@gazi.edu.tr

Eğitim Derecesi	Okul/Program	Mezuniyet tarihi
Yüksek Lisans	Gazi Üniversitesi Elektronik-Bilgisayar Eğitimi Bölümü	Devam Ediyor
Lisans	Eskişehir Anadolu Üniversitesi Açıköğretim İşletme Fakültesi İşletme Bölümü	2007
Lisans	Karabük Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümü	2007
Ön Lisans	Sakarya Üniversitesi Karasu Meslek Yüksek Okulu Bilgisayar Programcılığı	2003
Lise	Üsküdar İmam Hatip Lisesi	2000

İş Deneyimi, Yıl	Çalıştığı Yer	Görev
2002-2003	Edam Eğitim Danış. ve Araş. Mer. (Staj)	Bilg. Programcısı
2003-2005	MindShare Medya IT Department	Yazılım Geliştirme
2008–devam ediyor	75. Yıl D.M.O. Tek. ve End. Mes. Lisesi	Öğretmen/Md.Yrd.

Yabancı Dil

İngilizce

Yayınlar

1. Atacak, İ., Gökpınar, B. (2013). A computer-aided diagnosis system for detection of Parkinson's disease using fisher score feature selection and neural network. *3rd World Conference on Innovation and Computer Science (INSODE-2013)*, Antalya, 4(2), 639-642.
2. Gökpınar, B., Çeşiçi, M., Bakış, N. (2013). Okul tek kartla kolay. *Aydın Üniversitesi Eğitimde İyi Örnekler Paylaşımı Konferansı*, İstanbul.
3. Gökpınar, B., Peker, M., İzdaş, H., ve Gökçimen, F. (2013). Mesleki yarışmalar: keşfet - eğlen - yarış. *Sabancı Üniversitesi Eğitimde İyi Örnekler Konferansı (İÖK)*, İstanbul.



GAZİ GELECEKTİR...