



**VOKAL KORD ÖLÇÜMLERİNDEN MAKİNE  
ÖĞRENMESİ YÖNTEMLERİ KULLANARAK  
PARKİNSON HASTALIĞININ TEŞHİSİ**

**BÜŞRA ZEYNEP GÜREL**

**YÜKSEK LİSANS TEZİ**

**ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA  
BİLİM DALI**

**DOÇ. DR. MAHMUT HEKİM**

**Temmuz-2019  
Her hakkı saklıdır**

T.C.  
TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ  
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

YÜKSEK LİSANS TEZİ

VOKAL KORD ÖLÇÜMLERİNDEN MAKİNE ÖĞRENMESİ YÖNTEMLERİ  
KULLANARAK PARKİNSON HASTALIĞININ TEŞHİSİ

BÜŞRA ZEYNEP GÜREL

TOKAT  
Temmuz - 2019

Her hakkı saklıdır

BÜŞRA ZEYNEP GÜREL tarafından hazırlanan "Vokal kord ölçümlerinden makine öğrenmesi yöntemleri kullanarak Parkinson hastalığının teşhisi" adlı tez çalışmasının savunma sınavı 11 TEMMUZ 2019 tarihinde yapılmış olup aşağıda verilen Jüri tarafından Oy Birliği ile Tokat Gaziosmanpaşa Üniversitesi Fen Bilimleri Enstitüsü ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI 'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

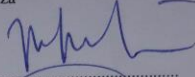
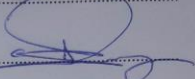
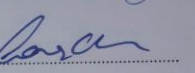
Jüri Üyeleri

İmza

Danışman  
Doç. Dr. Mahmut HEKİM

Üye  
Dr. Öğr. Üyesi Cem EMEKSİZ  
Tokat Gaziosmanpaşa Üniversitesi

Üye  
Dr. Öğr. Üyesi Canan ORAL  
Amasya Üniversitesi

  
.....  
  
.....  
  
.....

ONAY  
  
Prof. Dr. Cetin ÇEKİC  
Fen Bilimleri Enstitüsü Müdürü  
11/07/2019

## **TEZ BEYANI**

Tez yazım kurallarına uygun olarak hazırlanan bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezin içerdiği yenilik ve sonuçların başka bir yerden alınmadığını, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bu üniversite veya başka bir üniversitedeki başka bir tez çalışması olarak sunulmadığını beyan ederim.

**BÜŞRA ZEYNEP GÜREL**

**11 Temmuz 2019**

## ÖZET

### YÜKSEK LİSANS TEZİ

## VOKAL KORD ÖLÇÜMLERİNDEN MAKİNE ÖĞRENMESİ YÖNTEMLERİ KULLANARAK PARKİNSON HASTALIĞININ TEŞHİSİ

BÜŞRA ZEYNEP GÜREL

TOKAT GAZİOSMANPAŞA ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

(TEZ DANIŞMANI:DOÇ. DR. MAHMUT HEKİM)

Parkinson hastalığının (PD) en önemli semptomlarından biri, ağız, dil, boğaz, ses tellerinin fiziksel bir bozukluğu nedeniyle konuşmada güçlük çeken konuşma bozukluğudur. Yaygın olarak kullanılan Birleşik Parkinson Derecelendirme Ölçeği (UPDRS) gibi standart ölçeklere rağmen hastalığın erken evrelerde teşhisi zordur. Bu yüzden, bu çalışmada, hastalardan alınan ve sağlıklı gönüllülerden kaydettiğimiz vokal kord işaretlerinin zaman, frekans ve zaman-frekans domenlerinde istatistiksel parametreler yardımıyla özellik vektörleri çıkarılmış ve bu vektörler PD tanısı için makine öğrenmesi (ML) temelli sınıflayıcı modellere girdi olarak uygulanmıştır. Bu amaç için, vokal kord işaretlerine ayırık Fourier dönüşümü (DFT) kullanılarak elde edilen frekans spektrumu üzerinde ve ayırık dalgacık dönüşümü (DWT) uygulanarak ayrıştırılan frekans alt-bantları üzerinde istatistiksel parametreler yardımıyla özellik vektörleri elde edilmiştir. Bu özellik vektörleri PD'nin teşhisi için yaygın olarak kullanılan ML yöntemlerine dayalı sınıflayıcı modellere giriş olarak uygulanmıştır. PD teşhisi için zaman, frekans ve zaman-frekans domenlerinde gerçekleştirilen sınıflandırma deneylerinde, kullanılan ortalama, geometrik ortalama, harmonik ortalama, varyans gibi temel istatistiksel parametreleri ile hesaplanan özellik vektörleri sınıflayıcı modellere giriş olarak uygulandığında toplam doğru sınıflandırma (TCC) oranları açısından arzu edilen seviyelere ulaşmıştır. Buna ek olarak, PD teşhisine ilişkin ilk kez bu tezde kullanılan noktalar-arası eğimi hem zaman hem de frekans domeninde uygulanıp elde edilen yeni işaretin Ki-Karesi ve entropisi hesaplanarak gerçekleştirilen sınıflandırma deneyleri tekrarlandığında özellikle işaretlerin frekans domeninde sınıflayıcı modeller çok daha yüksek TCC oranlarını sağlamıştır. Sonuç olarak, hastaların ses işaretlerine ait frekans alt-bantlarında istatistiksel parametrelerle ve ses işaretlerinin eğimi alınıp Ki-Kare ve entropi uygulandığında elde edilen özellik vektörlerinden ML yöntemleri yardımıyla PD'nin teşhisinin yüksek başarı oranlarında gerçekleştirilmesi mümkün olmuştur.

2019, 29 SAYFA

**ANAHTAR KELİMELELER:** Vokal kord işaretleri, Parkinson hastalığı, Özellik çıkarma, Noktalar-arası eğim, Makine öğrenmesi, Ayırık Fourier dönüşümü, Ayırık dalgacık dönüşümü.

## ABSTRACT

### MASTER THESIS

#### THE DETECTION OF PARKINSON DISEASE USING MACHINE LEARNING MODELS BÜŞRA ZEYNEP GÜREL

TOKAT GAZIOSMANPASA UNIVERSITY  
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF ELECTRICAL ENGINEERING

(SUPERVISOR:ASSOC. PROF. DR. MAHMUT HEKİM)

Since one of the most important symptoms of Parkinson's disease (PD) is dysphonia that is difficulty in speaking due to PD and a physical disorder of the mouth, tongue, throat, vocal cords, the diagnosis of disease is difficult to diagnose early with the widely used standard scales such as Unified Parkinson's Disease Rating Scale (UPDRS). Therefore, in this study, feature vectors were extracted from patients and healthy volunteers by using statistical parameters in the time, frequency and time-frequency domains of vocal cords, and these vectors were applied as the inputs into machine learning (ML) based classifier models for PD diagnosis. For this aim, the frequency domain coefficients and frequency sub-bands coefficients were obtained by using discrete Fourier transform (DFT) and discrete wavelet transform (DWT) from vocal cord voice signals, respectively. The feature vectors were extracted from these obtained coefficients by using statistical parameters such as mean, geometric mean, harmonic mean and standard deviation. These feature vectors were applied as the inputs into classifier models based on commonly used ML methods for PD diagnosis. In the classification experiments implemented in time, frequency and time-frequency domains for PD diagnosis, the classifier models reached to the satisfactory levels in terms of total correct classification (TCC) ratios. In addition, when the classification experiments performed by using the parameters of entropy, Chi-inter-points slope that were evaluated in this thesis for the first time were repeated for PD diagnosis, the classifier models provided much higher TCC ratios especially when the feature vectors obtained from the frequency domain of the signals were used. As a result, this approach allowed to diagnose PD from feature vectors obtained by using statistical parameters in the frequency sub-bands of patients' voice signals in high success ratios with the help of ML methods.

2019, 29 PAGE

**KEYWORDS:** Vocal cord signals, Parkinson's disease, Feature extraction, Inter-points slope, Machine learning, Discrete Fourier transform, Discrete wavelet transform.

## ÖNSÖZ

Tez çalışmalarım süresince yapmış olduğum arařtırmalarda beni yönlendiren ve çalışmalarımnda desteklerini esirgemeyen değerli hocam Doç. Dr. Mahmut HEKİM 'e ve öğrenim hayatım boyunca maddi manevi destekleriyle beni hiçbir zaman yalnız bırakmayan, bu günlere gelmemde büyük pay sahibi olan annem Ayşe GÜREL, babam İbrahim GÜREL, kardeşim Z. Ziya GÜREL ve sevgili anneannem Hanife KANIK'a, Parkinson hastalığı hakkında bildiklerini benimle paylaşan dayım Opt. Dr. İhsan KANIK'a, tez yazımında desteklerini esirgemeyen Dr. Öğr. Üyesi Mehmet AYAS'a ve manevi teyzem Öznur AYAS'a teşekkürlerimi sunarım.

**BÜŞRA ZEYNEP GÜREL**

**11 Temmuz 2019**

# İÇİNDEKİLER

## Sayfa

<b>ÖZET</b> .....	<b>i</b>
<b>ABSTRACT</b> .....	<b>ii</b>
<b>ÖNSÖZ</b> .....	<b>iii</b>
<b>İÇİNDEKİLER</b> .....	<b>iv</b>
<b>KISALTMALAR</b> .....	<b>v</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>vii</b>
<b>ÇİZELGE LİSTESİ</b> .....	<b>viii</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. KAYNAK ÖZETLERİ</b> .....	<b>3</b>
<b>3. MATERYAL ve YÖNTEM</b> .....	<b>6</b>
<b>3.1. PD Veri Seti</b> .....	<b>6</b>
<b>3.2. Özellik Vektörleri ve ML Yöntemleri Kullanılarak PD Teşhisi</b> .....	<b>6</b>
3.2.1. Normalizasyon.....	8
3.2.2. Ayırık Fourier dönüşümü (DFT) .....	8
3.2.3. Ayırık dalgacık dönüşümü (DWT).....	9
3.2.4. Noktalar-arası eğim .....	10
3.2.5. Ki-Kare (Chi-Square) .....	11
3.2.6. Özellik çıkarma.....	11
3.2.7. Parkinson hastalığının teşhisinde kullanılan sınıflayıcı modeller .....	13
3.2.8. Değerlendirme kriterleri .....	17
<b>4. BULGULAR ve TARTIŞMA</b> .....	<b>19</b>
<b>5. SONUÇ</b> .....	<b>26</b>
<b>6. KAYNAKLAR</b> .....	<b>27</b>
<b>7. ÖZGEÇMİŞ</b> .....	<b>29</b>



## KISALTMALAR

<b>Kısaltmalar</b>	<b>Açıklama</b>
A3M	Reel A3 katsayıların Ortalamaları
ANFIS	Adaptif Sinirsel-Bulanık Çıkarım Sistemi
BFS	Geriye Doğru Eleme (Backward Feature Selection)
BPNN	Bağımsız Geri Yayılımlı Sinir Ağı
CGP	Kartezyen Genetik Programlama ()
CM	Karışıklık Matrisi (Confusion Matrix)
CVANN	Karmaşık Değerli Yapay Sinir Ağı
D1S	D1 Katsayılarının Standart Sapmaları
D2M	Reel D2 Katsayıların Ortalamaları
D3H	Reel D3 Katsayıların Harmonik Ortalamaları
DFT	Ayrık Fourier Dönüşümü (Discrete Fourier Transform)
DT	Karar Ağacı (Decision Tree)
DWT	Ayrık Dalgacık Dönüşümü (Discrete Wavelet Transform)
ET	Temel Tremor (Essential Tremor)
FCM	Bulanık C-Ortalama
FCMFWF	FCM Kümeleme Tabanlı Özellik Ağırlıklandırma
FD	Frekans Domeni
FN	Yanlış Negatif (False Negative)
FP	Yanlış Pozitif (False Positive)
FPR	FP oranı (False Positive Rate)
FTIR	Fourier Dönüşümlü Infrared Mikro-Spektroskopisi
GMI	Sanal Katsayıların Geometrik Ortalamaları
HMI	Sanal Katsayıların Harmonik Ortalamaları
HP	Yüksek Geçiren (High Pass)
Ivar	Sanal Katsayıların Varyansları
KKNN	Bulanık $k$ -En Yakın komşu Sinir Ağı
KMCFW	$k$ -Ortalama Kümeleme Temelli Özellik Ağırlıklandırma
KNN	$k$ -en Yakın Komşu ( $k$ -nearest Neighbor)
LP	Alçak Geçiren (Low Pass)

LR	Lojistik Regresyon (Logistic Regression)
ML	Makine Öğrenmesi (Machine Learning)
MLPNN	Çok Katmanlı Algılayıcı Sinir Ağı (Multi-layer Perceptron Neural Network)
PCA	Temel Bileşen Analizi
PD	Parkinson Hastalığı (Parkinson Disease)
PDQ-39	Parkinson Hastalığı Anketi 39
PNN	Olasılıksal Sinir Ağı
PT	Parkinson Hastalığı Tremoru
Rvar	Reel Katsayıların Varyansları
SVM	Destek Vektör Makinesi (Support Vector Machine)
TCC	Toplam Doğru sınıflandırma (Total Correct Classification)
TN	Doğru Negatif (True Negative)
TNR	Doğru Negatif Sınıflandırma Oranı,Özgüllük (True Negative Rate)
TP	Doğru Pozitif (True Positive)
TPR	TP oranı, Hassasiyet (True Positive Rate)
UPDRS	Birleşik Parkinson Hastalığı Değerlendirme Ölçeği (Unified Parkinson's Disease Rating Scale)
vGRF	Zemin Reaksiyon Kuvvetinin Dikey Bileşeni
YSA	Yapay Sinir Ağı
ZD	Zaman Domeni

## ŞEKİL LİSTESİ

<b><u>Şekil</u></b>	<b><u>Sayfa</u></b>
Şekil 3.1 PD teşhisinde temel istatistiksel parametreleri değerlendirmek için tasarlanan sistemin akış diyagramı. ....	7
Şekil 3.2 PD teşhisinde noktalar arası eğimi alınan işaret için tasarlanan sistemin akış diyagramı. ....	7
Şekil 3.3 Üçüncü dereceden DWT akış diyagramı. ....	9
Şekil 3.4 RF akış şeması. ....	15
Şekil 3.5 Bir MLPNN modelinin yapısı. ....	16
Şekil 4.1 Normalizeli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri.....	20
Şekil 4.2 Normalizesiz işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri.....	21
Şekil 4.3 Normalizeli Eğimli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri .....	22
Şekil 4.4 ZD’de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların ROC eğrileri.....	23
Şekil 4.5 FD’de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların ROC eğrileri.....	24

## ÇİZELGE LİSTESİ

<b><u>Çizelge</u></b>	<b><u>Sayfa</u></b>
Çizelge 3.1 Karışıklık Matrisi.....	18
Çizelge 4.1 Normalizeli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri.....	19
Çizelge 4.2 Normalizesiz işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri.....	20
Çizelge 4.3 Normalizeli eğimli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri .....	22
Çizelge 4.4 ZD'de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların başarı ölçütleri.....	23
Çizelge 4.5 FD'de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların başarı ölçütleri.....	24
Çizelge 4.6 BFS'den sonra veri kümesinden çıkarılan öznitelikler .....	25

## 1. GİRİŞ

Parkinson hastalığı (PD), konuşma bozuklukları (disfoni), yürüyüş bozuklukları, bunama, titreme gibi ana belirtiler gösteren, beyinde dopamin salgılamaktan görevli nöronların zarar görmesinden kaynaklanan nörodejenaritif bir bozukluktur (Jankovic, 2007). 1817'de Dr. James Parkinson tarafından ilk defa tanımlanıp ve kendisi tarafından adlandırılmıştır (Langston, 2002). PD'nin dünya çapında 5 milyon insanı etkilediği ve bu durumun önmümüzdeki 10 yıl içerisinde önemli derecede artacağı tahmin edilmektedir. Türkiye'de ise belli bir yaş grubunda Alzheimer hastalığından sonra en sık görülen Parkinson hastalığı, 65 yaş ve üzerindeki her 100 kişiden birinde görülmekte ve her yıl yaklaşık 10 bin kişiye hastalık tanısı konmaktadır. PD'nin belirtilerinin bazıları günümüzde ilaç tedavisiyle veya cerrahi müdahale ile azaltılabilmekle birlikte, PD için kesin teşhis koyan bir yöntem henüz yoktur. Ancak, yine de uzmanlara Parkinson Hastalığı Anketi 39 (PDQ-39), Birleşik Parkinson Derecelendirme Ölçeği (MDS-UPDRS), Hoehn ve Yahr Ölçeği, Kısa Psikiyatrik Değerlendirme Ölçeği gibi ölçekler yardımcı olmakta, ama yeterli görülmemektedir. Bu yüzden, son zamanlarda destek vektör makinaları (SVM), yapay sinir ağı (YSA) gibi makine öğrenmesi (ML) tabanlı teşhis için ileri sürülen çözümlerin hastalığın erken teşhisinde ve yapılan ölçek sonuçlarının desteklenmesinde güvenilir katkı sağlayacağı düşünülmektedir (Guruler, 2017).

Mevcut kullanılan ölçeklerle erken teşhisi zor olan PD'nin en önemli belirtilerinden biri konuşma bozukluğu olduğu için, bu çalışmada hastalardan ve sağlıklı gönüllülerden alınan ses kayıtlarından oluşturulan işaretlerin zaman ve frekans domenlerinde istatistiksel parametreler kullanılarak çıkartılan özellik vektörlerinden ML yöntemleri yardımıyla PD teşhisi için bilgisayar destekli bir karar sistemine odaklanılmıştır. Bu amaç için, ses işaretlerine ayrık Fourier dönüşümü (DFT) kullanılarak elde edilen frekans spektrumu üzerinde ve ayrık dalgacık dönüşümü (DWT) uygulanarak ayrıştırılan frekans alt-bantları üzerinde istatistiksel parametreler yardımıyla özellik vektörleri elde edilmiştir. Bu özellik vektörleri PD'nin teşhisi için yaygın olarak kullanılan ML yöntemlerine dayalı sınıflayıcı modellere giriş olarak uygulanmıştır.

Gerçekleştirilen sınıflandırma deneylerinin sonucunda toplam doğru sınıflandırma (TCC) oranları işaretin normalizasyonu alınıp temel istatistiksel parametreler uygulandığında LR için 0.972, NB ve MLPNN için 1.0 doğruluk ölçülmüştür. İşaretin normalizasyonu alınmadan tekrarlandığında RF için 0.972 ve LR için 0.916 doğruluk ölçülmüştür.

Bu tezde, vokal kord işaretlerinden PD teşhisine ilişkin ilk defa kullanılan işaretin noktalar-arası eğimine ve kendisine Ki-Kare, entropi uygulanıp sınıflandırma deneyleri tekrarlandığında zaman domeninde LR için 0.972, MLPNN için 0.93 doğruluk ölçülmüştür. Vokal kord işaretlerinden DFT ile elde edilen katsayılar üzerinde yine aynı parametrelerle hesaplanan özellik vektörleri kullanıldığında KNN, LR, SVM kernel *RBF*, NB, DT, RF ve MLPNN için 1.0 TCC oranlarına ulaşılmıştır.

## 2. KAYNAK ÖZETLERİ

Parkinson hastalığının ML yardımıyla teşhis edilmesinde, veri setlerinin oluşturulması, sınıflayıcı modelin eğitimi, kullanılan özellikler ve işlenmesi çeşitlilik göstermektedir. Hamilton tarafından 2006 yılında DaTSCAN görüntülerinin yarı sayısal değerlendirmesinde, putamendeki izleyici birikiminin kaudat çekirdeğindeki oranına oranlanmasıyla yardımcı olabileceği ve PD'nin ilerlemesinin değerlendirilmesine izin verilebileceği öne sürülmüştür ve çalışmasında genel olarak yüksel bir başarı elde etmiştir. (Hamilton, 2016). Ön-işlemeli yürüyüş veri seti de PD'nin sınıflandırılması için kullanılmıştır ve %90'nın üzerinde başarı elde edilmiştir (Baby, 2017). 2009 yılında Muniz ve arkadaşları çalışmalarında, zemin reaksiyon kuvvetinin dikey bileşeninden (vGRF) türetilen temel bileşenleri girdi olarak alan, sağlıklı ve PD deneklerin arasındaki farkı belirlemek için bir olasılıksal sinir ağı (PNN) kullanmışlardır ve %90'nın üzerinde başarı elde etmişlerdir (Muniz ve ark., 2009). 2008 yılında Lingmei Ai ve arkadaşları yaptıkları çalışmada sık karşılaşılan klinik yanlış tanı konulan temel tremor (ET), PD tremoru (PT) ve fizyolojik tremor gibi üç çeşit bilindik titremeyi ayırt etmek için yeni bir yöntem önermişlerdir. Bunun için bağımsız geri yayımlı sinir ağı (BPNN) çıktı sonuçlarını D-S kanıt teorisi ile birleştirerek el titreme ivme işaretleri için yeni sentetik çoklu özellikli tanı yöntemi ortaya koymuşlardır ve titremeleri başarı bir şekilde ayırt edebilmişlerdir (Ai ve ark., 2008). 2012 yılında Ozcift, SVM'ye dayalı bir sınıflandırma yöntemi geliştirmiştir ve PD ilerlemesinin tahmini için % 97 doğruluk elde etmiştir (Ozcift, 2012). 2013 yılında Chen ve ark., Bulanık k-En Yakın Komşu (FKNN) yöntemi kullanarak bir PD tanı sistemi önermiştir (Chen ve ark., 2013). Gelişmiş FKNN tabanlı sistemin sonuçlarını SVM tabanlı yaklaşımların sonuçlarıyla karşılaştırmıştır. Ayrıca PD teşhis doğruluğunu artırmak amacıyla analiz aşamasındaki veri setinin boyutlarını azaltmak için Temel Bileşen Analizi (PCA) yaklaşımını kullanmıştır. Deneysel sonuçlar, FKNN tabanlı sistemin, sınıflandırma doğruluğunu (%96.07) önemli ölçüde geliştirdiğini ve literatürde SVM tabanlı yaklaşımları ve diğer yöntemleri geride bıraktığını göstermiştir. 2014 yılında Chandrashekar yaptığı çalışmada YSA ve konuşma ölçümleri kullanarak PD'nin doğru bir şekilde teşhis edilmesine yönelik yeni bir yöntem geliştirilmiştir (Chandrashekar, 2014). Bu yöntem için hastaların bir mikrofonla konuşması gerekmektedir. Konuşma işaretleri daha sonra Praat (fonetikte konuşmanın bilimsel analizi için ücretsiz bir bilgisayar yazılımı paketi) kullanarak analiz edilmiş ve

teşhisi koyabilmek için YSA temelli sınıflayıcı kullanılmıştır. MATLAB kullanarak oluşturulan ve gerçek hasta verisi ile eğitilen ve test edilen YSA % 96.55'ini doğru bir şekilde sınıflandırabilmiştir. 2016 yılında Hlavica ve arkadaşlarının bir çalışmasında, konuşma işareti ölçümlerini klinik çıktılara eşleyen iki hesaplama modeli önerilmiş ve sınıflayıcı olarak YSA ve ANFIS modelleri kullanılmıştır (Hlavica, 2016). Burada, Levenberg-Marquardt optimizasyon algoritması ile eğitilen sinir ağı modeli kullanıldığında esnek geri yayılım ve eşlenik gradyan yöntemleri kullanılarak eğitilen ağlara göre daha uygun korelasyon katsayısına ulaşılmıştır. 2017 yılında M. Mahsal Khan ve ark., Kartezyen Genetik Programlama (CGP) kavramını kullanarak hem ayırt edilebilir hem de ayırt edilemez parametreleri eğitmek için evrimsel bir yöntem önermiştir (Khan ve ark., 2017). Bu yaklaşımda, meme kanseri ve Parkinson hastalığının saptanması için gerçek dünyadaki veri setleri üzerinde değerlendirilmiştir. Önerilen yöntemin performansı, çeşitli standart sınıflandırma yöntemleriyle karşılaştırılabilir olmuştur. 2016 yılında M. Nilashi ve ark., gürültü giderme, kümeleme ve tahmin yöntemlerini kullanarak PD ilerlemesinin tahmin edilmesi için yeni bir hibrit akıllı sistem önermiştir (Nilashi ve ark., 2016). Çalışmalarında, PCA ve Beklenti Maksimizasyonu (EM), deneysel veri kümelerindeki çoklu ortak problemleri ele almak ve verileri kümelemek için kullanılmıştır. Ardından, PD ilerlemesinin tahmini için Adaptif Sinirsel-Bulanık Çıkarım Sistemi (ANFIS) ve Destek Vektör Regresyonunu (SVR) uygulamıştır. Halka açık Parkinson veri setlerindeki deneysel sonuçlar, önerilen yöntemin, PD ilerlemesinin tahmin doğruluğunu önemli ölçüde geliştirdiğini göstermiştir. Ortaya koydukları hibrit akıllı sistemin, PD'nin erken tespiti için sağlık pratisyenlerine yardımcı olabileceğini öne sürmüşlerdir. 2017 yılında H. Gürüler tarafından yapılan bir çalışmada k-ortalama kümeleme temelli özellik ağırlıklandırma (KMCFW) yönteminin ve karmaşık değerli yapay sinir ağı (CVANN) birleşimini kullanılarak yeni bir teşhis sistemi amaçlanmıştır (Guruler, 2017). Modelin ilk aşamasında PD veri setindeki özellikler KMCFW yöntemi kullanılarak ağırlıklandırılmış ve elde edilen yeni özellikler karmaşık bir sayı formatına dönüştürülmüştür. Bu özellik değerleri, CVANN'ye bir girdi olarak sunulmuştur. Önerilen sistemin verimliliği ve etkinliği beş farklı değerlendirme yöntemi açısından PD veri grubu ile değerlendirilmiştir. Deneysel sonuçlar, KMCFW-CVANN olarak adlandırılan hibrit sistemin, gerçek değerli YSA'dan belirgin şekilde üstün olduğunu ve şimdiye kadar bildirilen en yüksek sınıflandırma sonuçlarına ulaştığını ve %99.52'lik bir



sınıflandırma doğruluğu ile gösterildiğini ortaya koymuştur. 2017 yılında K. Yılcıoğlu'nun çalışmasında, YSA modeli ile regresyon analizi yöntemi kullanılarak motor, toplam Birleşik Parkinson Hastalığı Değerlendirme Ölçeği (BPHDÖ) klinik sonuçları ve vokal kord işaretleri kullanılarak PD teşhisi için lineer regresyon sinir ağı modeli sunulmuştur (Yılcıoğlu, 2017). Buna ek olarak farklı özelliklerin önemini göstermek için regresyon modelinde kullanılan eğitim veri setlerinden bazı özellikler çıkarılmış ve sonuçlar ile öngörülen UPDRS skorları arasındaki korelasyon skorları incelenmiştir. Toplam UPDRS özelliği eğitim veri setinden çıkarılmış ve analiz tekrarlanmıştır. Tahmini motor UPDRS ve klinik motor UPDRS'nin korelasyonu %72'ye düşmüştür. Motor UPDRS özelliklerini eğitim setinden çıkarıldığında, toplam UPDRS tahmini etkinliği %76'ya kadar azalmıştır. Sonra, tüm *Jitter* değerleri eğitim veri setinden çıkarılmış, motor UPDRS skorları tahmin edilmiştir. Tahmini motor UPDRS skoru ve klinik motor UPDRS skoru arasındaki korelasyon %97 olarak bulunmuştur. Daha sonra *Jitter* değerlerini ve toplam UPDRS skorları kaldırılmış ve motor UPDRS skorları tahmin edilmiştir. Tahmini ve klinik motor UPDRS skoru arasındaki korelasyon %79 olarak bulunmuştur.

Bu tezde, zaman domeninde doğrudan vokal kord işaretleri üzerinde, DFT kullanılarak elde edilen frekans domeninde ve DWT kullanılarak elde edilen frekans alt-bantları üzerinde temel istatistiksel parametreler kullanılarak elde edilen özellik vektörleri PD tespiti için LR, DT, RF, SVM, DT, KNN ve MLPNN temelli sınıflayıcı modellerine giriş olarak uygulanmıştır. Gerçekleştirilen sınıflandırma deneylerinde kullanılan temel istatistiksel parametrelerin yanında PD teşhisine ilişkin ilk kez bu tezde önerilen noktalar-arası eğim parametresi kullanılarak elde edilen özellik vektörleri sınıflayıcılara giriş olarak uygulandığında memnun edici toplam doğru sınıflandırma (TCC) oranlarına ulaşılmıştır.

### **3. MATERYAL ve YÖNTEM**

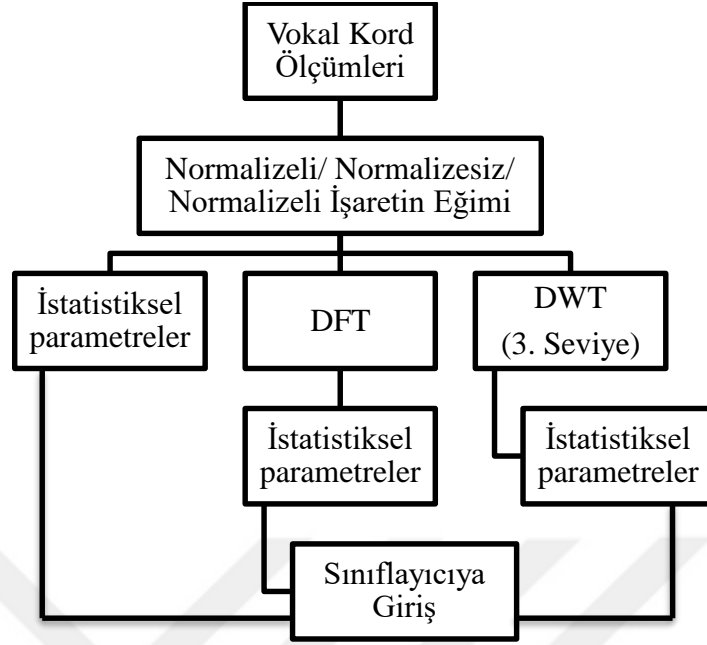
#### **3.1. PD Veri Seti**

Bu çalışmada 54 sağlıklı, 54 PD teşhisi konulmuş kişilerden oluşan veri kümesi kullanılmıştır. PD tanısı konulmuş kişilerin vokal kord ses işaretleri UCI veri deposundaki 20 kişinin ses örneklerinin 5'er saniyelik bölünmesi ve 48000 Hz ile örnekleme ile elde edilmiştir. (Erdogdu ve ark., 2013).

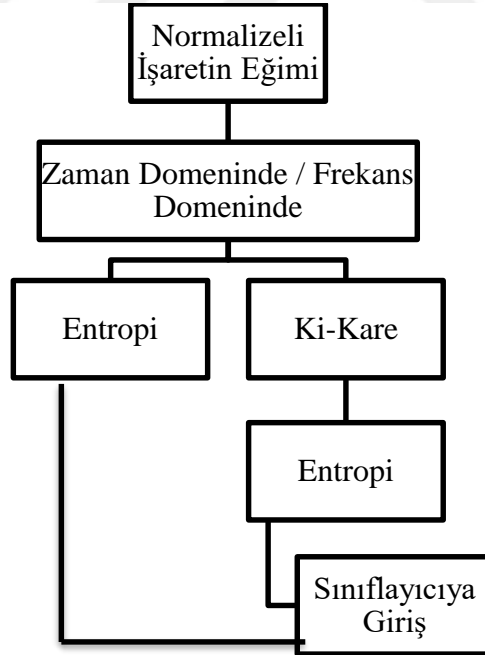
Sağlıklı sesler ise yaşları 16-50 arasında değişen gönüllülerden elde edilmiştir. Gönüllülerden 'a', 'o', 'u' harflerini 5'er saniye seslendirilerek elde edilen bu sesler standart olarak kullanılan 48000 Hz'lik bir örnekleme hızında örneklendikten sonra *m4a* biçiminde kaydedilmiştir.

#### **3.2. Özellik Vektörleri ve ML Yöntemleri Kullanılarak PD Teşhisi**

Bu çalışmada, vokal kord işaretlerinden ML temelli sınıflayıcı modellerle PD'nin yüksek başarı oranlarında teşhis edilebilmesi için sınıflayıcının girişine uygulanacak en uygun özellik çıkartma parametrelerinin saptanması için tasarlanan sistemin blok akış diyagramı Şekil 3.1 ve 3.2'de gösterilmektedir.



Şekil 3.1 PD teşhisinde temel istatistiksel parametreleri değerlendirmek için tasarlanan sistemin akış diyagramı.



Şekil 3.2 PD teşhisinde noktalar arası eğimi alınan işaret için tasarlanan sistemin akış diyagramı.

Akış diyagramlarında gösterilen parametreler aşağıdaki alt-başlıklarda detaylıca açıklanmaktadır.

### 3.2.1. Normalizasyon

Veri kümelerini oluşturan işaretlerin arasındaki uygunluğu sağlamak ve sınıflandırma başarısını artırmak için tüm ölçülen değerleri  $[0,1]$  aralığına çekmek için gerçekleştirilen işlemdir. Amaç ölçülen değerler arasındaki farkı en aza indirgeyerek hesaplamayı kolaylaştırmak ve doğruluğu arttırmaktır. Bu tezde Eşitlik 3.1 ile gösterilen *Min-Max* normalizasyon yöntemi kullanılmıştır.

$$X_N = \frac{X - X_{MIN}}{X_{MAX} - X_{MIN}} \quad (3.1)$$

Burada,  $N$  işaretin uzunluğu,  $X_{MIN}$  işaretteki en küçük değer,  $X_{MAX}$  ise işaretteki en büyük değerdir.

### 3.2.2. Ayrık Fourier dönüşümü (DFT)

Ayrık Fourier Dönüşümü (DFT) uygulanarak bir işaret, farklı genlik, frekans ve fazlarda kosinüs ve sinüs temel bileşenlerinin toplamı olarak ifade edilebilmektedir. Elde edilen temel bileşenler işaretlerin işlenmesinde ve analizinde önemli kolaylık sağlamaktadır.  $X(k)$  işaretinin DFT'si Eşitlik 3.2'deki tanımlanmaktadır.

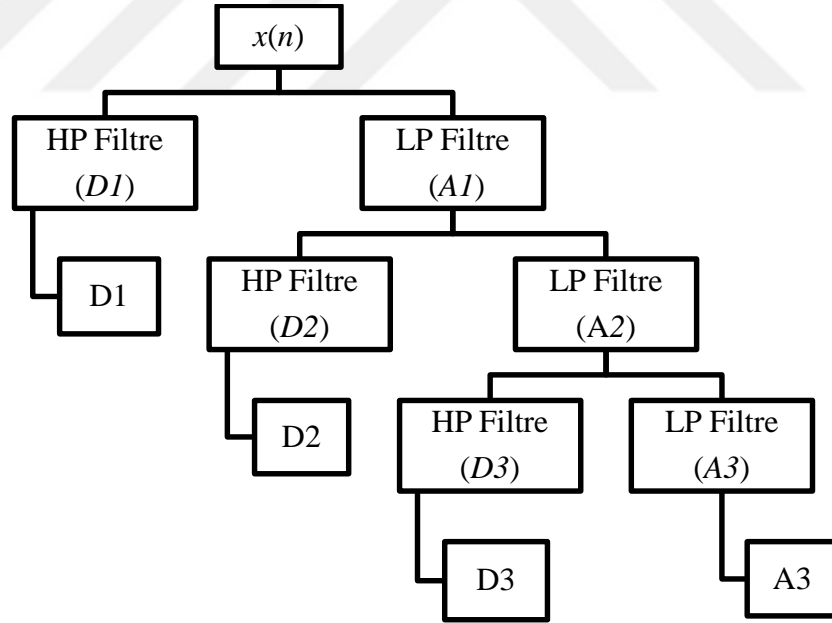
$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\left(\frac{2\pi n}{N}k\right)} \quad (3.2)$$

Bu çalışmada hem normalizasyon yapılmış hem de normalizasyon yapılmamış işaretin DFT'si alınarak elde edilen frekans spektrumu katsayıları PD hastalığı teşhisi için ML temelli sınıflayıcı modellere uygulanmıştır.

### 3.2.3. Ayrık dalgacık dönüşümü (DWT)

Ayrık Dalgacık dönüşümü (DWT), DFT'den farklı olarak düşük frekanslar için geniş, yüksek frekanslar için dar olacak şekilde değişen pencere boyutlarına sahiptir. Böylece, bütün frekans aralıklarında en iyi zaman-frekans çözünürlüğü sağlanabilmektedir. DWT farklı frekanslarda durağan olmayan güce sahip zaman serisi işaretlerinin analizinde kullanılabilir. Bunun yanı sıra, DFT'nin aksine sadece frekans bilgisi içeriğini değil, aynı zamanda zaman-frekans bilgisi içeriğini birlikte kullanır.

Ayrık zamanlı  $x(n)$  işareti yüksek geçiren (HP) filtreden geçirilir ve  $D1$  katsayısı elde edilir. Daha sonra  $x(n)$  işareti alçak geçiren (LP) filtreden geçirilir ve  $A1$  katsayısı elde edilir. LP'den geçirilmiş  $x(n)$  işareti tekrar HP ve LP filtrelerden geçirilir,  $A2$  ve  $D2$  katsayıları elde edilir. Bu işlem istenen seviyeye kadar devam ettirilebilir. Şekil 3.3'te herhangi bir işarete uygulanabilecek olan 3. seviyeden dalgacık dönüşümü gösterilmektedir.



Şekil 3.3 Üçüncü dereceden DWT akış diyagramı.

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}h(2^jx - k) \quad (3.3)$$

$$\psi_{j,k}(x) = 2^{\frac{j}{2}}g(2^jx - k) \quad (3.4)$$

Eşitlik 3.3'te  $\phi_{j,k}(x)$  alçak geçiren filtre ve dalgacık fonksiyonu, Eşitlik 3.4'te ise  $\psi_{j,k}(x)$  yüksek geçiren filtre verilmiştir. Burada  $x=0,1,2,\dots,N-1$ ,  $j=0,1,2,\dots,J-1$  ve  $k=0,1,2,\dots, 2^j - 1$  'dir.  $J$  burada  $\log_2(N)$  'dir. İşaretin  $x$  eksenindeki fonksiyon konumunu gösteren  $k$  örnekleme oranı,  $j$  ise fonksiyonun  $x$  eksenindeki genişliğini gösteren iyileştirmedir.

Eşitlik 3.5'te ise  $A$  yaklaşıklık, Eşitlik 3.6'da  $D$  detay katsayıları verilmiştir.

$$A_i = \frac{1}{N} \sum_x X(x) \phi_{j,k}(x) \quad (3.5)$$

$$D_i = \frac{1}{N} \sum_x X(x) \psi_{j,k}(x) \quad (3.6)$$

Bu çalışmada hem normalizeli hem de normalizesiz vokal kord işarete uygulanan olan DWT'nin ayrıştırma seviyesi doğru sınıflandırma başarı oranlarına belirlenmiştir. Deneysel çalışmalarda en yüksek TCC oranlarını 3. seviye sağladığı için bu seviyede dönüşüm kullanılmıştır. Bu yüzden çalışmada 3. seviyeden dalgacık dönüşümü yapılarak elde edilen A3, D1, D2, D3 katsayılarından hesaplanan özellik vektörleri PD teşhisi için ML temelli sınıflayıcı modellere giriş olarak uygulanmıştır.

#### 3.2.4. Noktalar-arası eğim ( $s_i$ )

İki nokta arasındaki yükselti farkının yatay uzaklığa oranı olan eğim sayısal ses işaretlerinde örnekleme aralığı eşit olduğu için iki örnek arasındaki fark olarak da tanımlanabilir. Bu fark (eğim) işaretin yükselen örnek değerlerinde pozitif, azalan örnek değerlerinde negatif işaretli bir büyüklük olacaktır. Komşu örnek değerleri arasındaki korelasyon büyük olacağı için bu fark değerleri çok büyük değerli olması beklenmez. Dolayısıyla ses işaretlerinden elde edilen eğim işaretlerinin maksimum ve minimum değerleri arasındaki fark küçülecektir ve analiz edilen işaretin aralığı daha da küçülecektir. Bu da işaretlerin frekans alt-bantlarına ayrıştırıldıklarında PD hastalığının teşhisinde kullanılan sağlıklı ve hasta işaretlerinin farklı bir dağılım göstereceği izlenimi vermektedir. Bu yöntem verinin boyutunun küçüleceğini garanti etmez. Ancak minimum seviyede yani neredeyse kayıpsız bir sıkıştırma sağlar.

Bu yöntem genellikle haberleşmede ses işaretlerinin sayısal iletiminde kullanılmakla birlikte, hastalık teşhisinde işaret dönüşümü olarak ilk kez bu tez çalışmasında kullanılmıştır. Noktalar-arası eğim dönüşümü aşağıdaki Eşitlik 3.7 ile tanımlanabilir:

$$s_i = (x_{i+1} - x_i) \quad (3.7)$$

Burada  $x_i$  ve  $x_{i+1}$  işaretteki  $i$ . ve sonraki elemanlardır.

### 3.2.5. Ki-Kare (Chi-Square)

Ki-Kare dağılımı genel olarak, veri kümesindeki verilerin birbiri ile uyumluluğunu ölçer. Deneysel çalışmalarda Ki-Kare olasılıksal dağılım yöntemi kullanılmıştır ve yöntemin eşitliği Eşitlik 3.8 ile verilmiştir.

$$y = f(x|v) = \frac{x^{(v-2)/2} e^{-x/2}}{2^{v/2} \Gamma(\frac{v}{2})} \quad (3.8)$$

Burada  $\Gamma(\cdot)$  gamma fonksiyonudur.  $v$  ise serbestlik derecesidir ve en yüksek başarıyı sağlama için deneysel olarak  $v=4$  olarak seçilmiştir.

### 3.2.6. Özellik çıkarma

Özellik çıkarma, kullanılacak sınıflandırıcının eğitimi ve testi için veri setlerini temsil edecek daha öz ve daha ayırt edici özellikleri oluşturma işlemidir. Bu tezde, PD teşhisi için, zaman ve frekans düzlemini oluşturan frekans alt-bant katsayılarının ortalama, harmonik ortalama, geometrik ortalama, varyans ve ses işaretlerinin eğimine de Ki-Kare, entropi ve değerleri sınıflandırıcının girişine uygulanmıştır. Vokal kord işaretlerinden özellik vektörlerini çıkarmak için kullanılan parametreler aşağıda gösterilmektedir:

Ortalama ( $x_{ort}$ ): Bir serideki elemanların toplamının toplam eleman sayısına oranıdır ve bu oran Eşitlik 3.9 ile hesaplanmaktadır.

$$x_{ort} = \frac{\sum_{i=1}^n x_i}{n} \quad (3.9)$$

Burada,  $x_i$  işaretteki  $i$ . elemanı,  $n$  işaretin uzunluğunu göstermektedir.

Varyans ( $\sigma^2$ ): Bir serideki elemanların dağılımı hakkında bize bilgi verir. varyans Eşitlik 3.10 ile verildiği gibi bir serideki elemanların ortalamaya uzaklıklarının kareleri toplamının, toplam eleman sayısına oranıdır.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - x_{ort})^2}{n} \quad (3.10)$$

Geometrik ortalama ( $G$ ): Bir serideki elemanların çarpımının, o serideki eleman sayısı derecesinde kökünün alınmasıdır. Geometrik ortalama Eşitlik 3.11 ile hesaplanmaktadır.

$$G = \sqrt[n]{\prod_{i=1}^n x_i} \quad (3.11)$$

Harmonik ortalama ( $H$ ): Bir serideki elemanların terslerinin toplamının yine tersinin eleman sayısı ile çarpımına eşittir ve Eşitlik 3.12 ile tanımlanmaktadır.

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (3.12)$$

Entropi ( $E_n$ ): Bir rastgele örneğin doğal belirsizlik veya rastlantısallık miktarını ölçen önemli bir bilgi teorisidir (Zhang ve ark., 2011). Eşitlik 3.13'te Shannon entropi eşitliği verilmiştir.

$$E_n = -\sum_{i=1}^n p_i \log_2(p_i) \quad (3.13)$$

Burada,  $n$  ses işaretinin örnek uzunluğudur  $p_i$  her bir ses işareti örneğinin bulunma olasılığıdır.

Geriye doğru eleme (BFS): BFS yaklaşımı eğitim için en uygun özneliği bulmaya yarayan yöntemdir. Ancak bulunan özelliklerin en uygun özellikler olduğunu garanti edemez. Bu yöntemde H1 ve H0 hipotezi belirlenir. H0 hedef, H1 ise buna karşıt hipotezdir. H0 doğruluğu, hassasiyeti arttırmak gibi herhangi bir performans unsuru olabilir. Burada kullanılan performans ölçütü ise  $P$  değeridir. Geriye doğru özellik seçiminde özellik setinden bir özellik çıkarılarak H0 ya da H1 hipotezinden birini sağlayıp sağlamadığına bakılır. Eğer H0 sağlanıyorsa (hedeflenen performans unsuru) o özellik eğitim setinden çıkarılarak, H1 sağlanıyorsa o özellik çıkarılmadan diğerleri için de aynı adım uygulanarak tekrar edilir ve model sonlandırılır.



### 3.2.7. Parkinson hastalığının teşhisinde kullanılan sınıflayıcı modeller

#### Lojistik Regresyon (LR):

Lojistik regresyonda bağımlı değişkenin tahmini değerleri olasılıksal olarak hesaplanır ve buna göre sınıflama yapar (Ayhan 2006). Deneysel çalışmalarda ikili lojistik regresyon yöntemi kullanılmıştır. İkili lojistik regresyon modeli Eşitlik 3.14 ile ifade edilir (Özdamar, 2009).

$$P(Y) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (3.14)$$

#### k-en yakın komşu (KNN):

Bu algoritma ML algoritmaları içerisinde en çok bilinen ve kullanılan algoritmalarından biridir. Seçilen bir özelliğin kendine en yakın olan özellikler arasındaki yakınlığı kullanarak sınıflandırma yapılır (Kılınç ve ark., 2016). Burada bulunan  $k$  değeri en yakın kaç komşuya bakılacağını ifade eder. Örneğin  $k=3$  ise en yakın 3 komşuya bakar. Nesnelere arasındaki mesafelerin belirlenmesinde genel olarak Eşitlik 3.15'teki formül kullanılmaktadır.

$$d(i, j) = \sqrt{\sum_{k=1}^N (x_{ik} - x_{jk})^2} \quad (3.15)$$

Burada,  $k=1,2,3,\dots,N$ 'dir.  $d(i, j)$  uzaklığı  $i$  ve  $j$  noktaları arasındaki uzaklığı,  $x_{ik}$   $i$ . gözlem için  $k$  değişkeninin almış olduğu değeri,  $x_{jk}$  ise  $j$ . gözlem için  $k$  değişkeninin almış olduğu değeri göstermektedir (Soner, 211).

#### Destek vektör makinası (SVM):

Bu yaklaşım sınıflandırma ve regresyon problemleri için kullanılabilir. Veri kümesini destek vektörleri ile böler ve buna göre sınıflandırma yapar. Veri kümesini destek vektörleri ile böler ve buna göre sınıflandırma yapar.  $i$ , 1 ile işaretin uzunluğu arasında bir değer alabilen tam sayı olmak üzere her  $i$  için  $x_i$ 'ye karşılık gelen ikili etiketlerle  $y_i = \epsilon\{-1,1\}$  giriş alanındaki giriş vektörleri olsun. Bu durumda  $x_i$ 'nin sınıfı 1 ise 1, değilse ise -1 alır.

$$y_i(\langle w, x_i \rangle + b) - 1 \geq 0 \quad (3.16)$$

Burada,  $w$  düzlemin normali aynı zamanda ağırlık vektörü ve  $b$  de sabit olarak tanımlanmıştır (Ayhan ve Erdoğan, 2014). Eşitlik 3.17 ile ise bu veri kümesini en iyi şekilde ayıran hiper düzlem verilmiştir.

$$\phi(w) = \frac{1}{2} \|w\|^2 \quad (3.17)$$

Burada,  $\phi(w)$  veri kümesini en iyi şekilde ayıran düzlemi,  $w$  düzlemin normalini göstermektedir.

#### Naive Bayes (NB):

Bu yaklaşım önceden belirli bir oranda sınıfı belli olan veri sunulur ve öğretilir. Öğretilmiş verilerle yeni veriler karşılaştırılarak yeni verinin hangi kategoriye ait olduğu tahmin edilir. Önceden sınıfı belli olan veri ne kadar çoksa tahmin o kadar gerçeğe yakın olur. NB’de önemli olan bir olayın olma olasılığıdır.

$$P\left(\frac{Y_i}{X}\right) = \frac{P\left(\frac{X}{Y_i}\right)P(Y_i)}{P(X)} \quad (3.18)$$

Burada,  $Y_i$  önceden etiket olarak 1 ya da -1 olarak sınıflandırılmış sınıflardan biridir. NB istekli öğrenmedir, yani tahmin edilecek veri gelmeden önce öğrenir, tüm olasılıkları hesaplar. Öğrendikten sonra eğitim için kullandığı veriyi silip sadece olasılıklarını sakladığı için, bellek açısından kullanışlıdır.

#### Karar ağacı (DT):

Bu yaklaşıma dayalı sınıflandırma veri kümesinin belli gruplara bölünerek öğrenilmesine dayanır. Karar ağacı bir düğüm ile başlar ki bu düğüm veri kümesindeki en ayırt edici kategoridir. En ayırt edici kategori ise kazancı en yüksek olandır. Eşitlik 3.19 ile tahmin edilecek sınıfın entropisi hesaplanır.

$$E(D) = - \sum_{i=1}^N P_i \log_2(P_i) \quad (3.19)$$

Burada,  $P_i$  tahmin edilecek sınıftır. Eşitlik 3.20 ile  $A$  niteliğinin  $D$  veri kümesindeki bilgi kazancı hesaplandıktan sonra, Eşitlik 3.21 ile kazanç hesaplanır ve kazancı en yüksek olan düğüm seçilir.

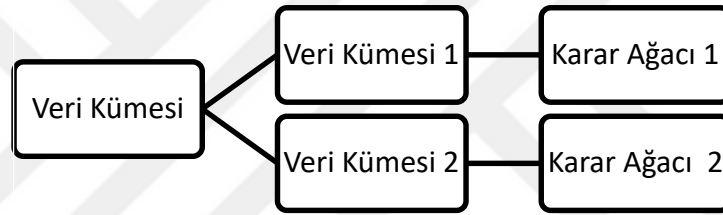
$$info_A(D) = \sum_{j=1}^N \left( \frac{|D_j|}{|D|} E(D_j) \right) \quad (3.20)$$

$$Gain(A) = E(D) - info_A(D) \quad (3.21)$$

Burada,  $Gain(A)$  A niteliğinin kazancını,  $D_j$  veri kümesindeki  $j$ . elemanı göstermektedir.  $N$  veri kümesinin uzunluğudur.

#### Rassal ağaçlar (RF):

Bu yaklaşım birlikte öğrenme yöntemidir. Yani, RF yaklaşımında birden fazla karar ağacı aynı anda kullanılarak daha yüksek başarı hedeflenmektedir. Şekil 4 RF modelin akış şemasını göstermektedir.



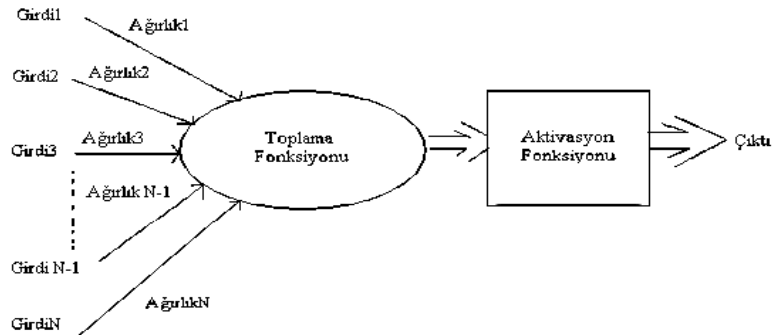
Şekil 3.4 RF akış şeması.

Bu şemanın sonucunda çoğunluğun oyu karar olarak alınabilir (tahmin yapılacaksa ortalamaları alınabilir). RF'te NB'nin aksine veri arttıkça sonuçlar yanlış olabilir. Bunun sebebi ise aşırı öğrenmedir. Ayrıca dallar attıkça hesaplama süresi de uzar. Bu yüzden küçük ve orta büyüklükteki veri setleri için kullanışlıdır.

### Çok katmanlı algılayıcı sinir ağı (MLPNN):

MLPNN girdileri uygun çıktılarla eşleştiren ileri beslemeli yapay sinir ağı modelidir. 3 kısımdan oluşur: Giriş katmanı, gizli katman, çıkış katmanı. Giriş katmanı girdilerden oluşur ve bu girişlerin her biri girdi düğümleri olarak adlandırılır. Bu kısımda veriler herhangi bir ön işleme tabii tutulmaz. Gizli katmanda giriş katmanından alınan bilgiler ağırlıklandırılıp toplama fonksiyonuna daha sonra aktivasyon fonksiyonuna verilerek elde edilen çıktılar bir sonraki katman olan çıkış katmanına iletilir. Aktivasyon fonksiyonu olarak sigmoid, hiperbolik tanjant, doğrusal adım fonksiyonu kullanılabilir. Bir MLPNN’de birden fazla gizli katman bulunabilir.

Çıkış katmanında gizli katmandan gelen bilgiler işlenerek girdilere karşılık gelen çıktılar belirlenir. Aktivasyon fonksiyonu olarak çıkış katmanında sigmoid fonksiyonu kullanılır. Bir yapay sinir ağında geri yayılım algoritması, esnek yayılım algoritması, delta bar algoritması, geliştirilmiş delta bar algoritması, hızlı yayılım algoritması, *Levenberg-Marquard* yöntemi, eşleştirmeli eğim algoritması, genetik algoritma gibi pek çok algoritma kullanılabilir. Bu çalışmada literatürde yaygın olarak kullanılan ileri ve geri yayılım algoritmaları kullanılmıştır. Şekil 3.5’te bir MLPNN modelinin yapısı gösterilmiştir.



Şekil 3.5 Bir MLPNN modelinin yapısı.

Bir sinir ağı oluştururken her bir girdi ağırlıklandırılarak ( $w$ ) ve bir *bias* katsayısı ile ( $b$ ) toplama fonksiyonuyla toplanır.

$$\hat{y} = (Girdi_n)w_n + b \quad (3.22)$$

Bu toplam bir aktivasyon fonksiyonunun girişine verilir. Aktivasyon fonksiyonunun amacı belirli bir eşik değerine göre sınıflandırmaktır. Aktivasyon fonksiyonu olarak *relu*, *sigmoid*, *adım* ve *lineer* kullanılabilir. Aktivasyon fonksiyonunun türevi alınabilir olması gerekmektedir. Burada eğer  $\hat{y}$  eşik değerini aşarsa sinir ağı hasta, aşmazsa sağlıklı olarak sınıflandırmaktadır. Aktivasyon fonksiyonu olarak 1. Katmanda *relu*, çıkış katmanında ise *sigmoid* kullanılmıştır. Çıkış katmanında genel olarak *sigmoid* kullanılır. Daha sonra hata ve toplam hata hesabı yapılır. Hata tahmin edilen ( $\hat{y}$ ) ve gerçekteki  $y$  değerlerinden ne kadar farklı ölçüldüğüdür. Toplam hata ise bu hataların toplamıdır. Eşitlik 3.23 ile bir MLPNN için hata fonksiyonu ve Eşitlik 3.24 ile toplam hata fonksiyonu verilmiştir.

$$Hata = (1 - y)\log(1 - \hat{y}) + (y\log(\hat{y})) \quad (3.23)$$

Burada,  $y$  gerçekteki çıktı değerini,  $\hat{y}$  ise tahmin edilen çıktı değeridir.

$$Toplam\ Hata = \sum(Hata) \quad (3.24)$$

Eğer toplam hata çok büyükse model kötü demektir ve parametreler yeniden güncellenir. Bu adımlara “*ileri yayılım*” adı verilir. Daha sonraki adımda elde edilen çıkışlar hata fonksiyonuna verilerek ne kadar hata olduğu hesaplanır. Hatayı en aza indirmek için gradyan iniş metodu kullanılmıştır. Gradyan iniş metodunda çıkışın fonksiyon üzerindeki noktasına teğeti eğimi verir. Bu eğim fonksiyon artış hızını ve yönünü verir. Eğer eğim pozitifse minimum noktanın negatif yönde, eğim negatifse minimum noktanın pozitif yönde olur. Artış hızı ve yönü kullanılarak hatanın en az olduğu nokta bulunur. Gradyan iniş metodunda bir diğer önemli değişken ise adım katsayısıdır ( $\eta$ :*eta*). Eğim azaldıkça adım katsayısı da küçülür. Ancak adım katsayısı başlangıçta çok küçük seçildiğinde minimum noktaya ulaşılamayabilir. Yani adım katsayısının küçük olarak seçilebilmesi için eğimin de az olduğundan emin olunmalıdır. Bu adımlar “*geri yayılım*” olarak adlandırılır.

### 3.2.8. Değerlendirme kriterleri

Karışıklık matrisi (CM): Doğru negatif (TN), yanlış negatif (FN), doğru pozitif (TP), yanlış pozitif (FP) sınıflandırılan veri sayısı veren matristir. Çizelge 3’te bir karışıklık matrisi (CM) verilmiştir.

Çizelge 3.1 Karışıklık Matrisi

TP	FP
FN	TN

Toplam doğru sınıflandırma (TCC) oranı: Bir modelde verilerin yüzde kaçının doğru sınıflandırıldığıdır. Eşitlik 3.25 ile TCC oranı verilmiştir.

$$TCC = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.25)$$

TP oranı (TPR, Hassasiyet): Pozitif etiketli verileri doğru tespit etme oranıdır. Eşitlik 3.26 ile TPR verilmiştir.

$$TPR = \frac{TP}{TP+FN} \quad (3.26)$$

FPR oranı (FPR): Negatif etiketli verilere hangi olasılıkla yanlış pozitif olarak sıralandığının oranıdır. Eşitlik 3.27 ile FPR verilmiştir.

$$FPR = \frac{FP}{FP+TN} \quad (3.27)$$

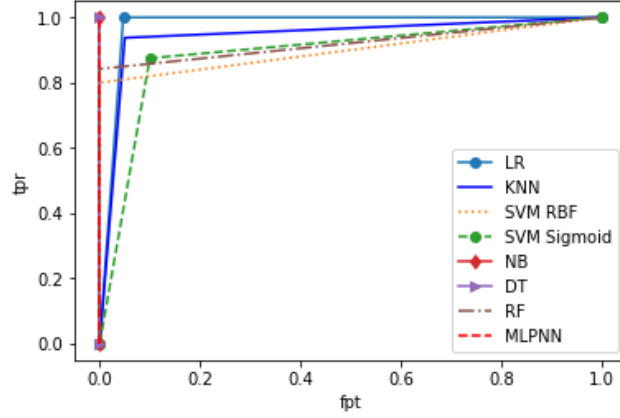
#### 4. BULGULAR ve TARTIŞMA

Veri ön işleme ve MLPNN modeli *MATLAB 16a* ara yüzünde gerçekleştirilmiştir. Veri kümesi *MATLAB*'da normalize edildikten sonra tekrar ML yöntemleri için *Anaconda Spyder*'da test edilmiştir. Vokal kord işaretlerinden oluşturulan veri setleri normalize edildikten sonra veri kümesinde ortalamalar, reel katsayıların ve 1., 2. ve 3. dalgacıkların hem *A* hem de *D* katsayılarının standart sapmalarının korelasyonları %95'ten büyüktür. Bunlar öğrenimi kötü etkilediği için veri kümesinden çıkarılarak, normalizeli işaretlere temel istatistiksel parametreler uygulanıp ML modellerinin girişine verilmiştir.

Vokal kord işaretlerinden oluşturulan veri setleri normalize edilmeden öğrenime katkıları araştırıldığında korelasyon katsayıları %95'ten büyük olanlar ezberden kaçınmak için veri kümesinden çıkarılmışlardır. İşaretlerin frekans düzlemine taşındıktan sonra elde edilen sanal katsayıların geometrik ortalamaları (*GMI*) ve harmonik ortalamaları (*HMI*), reel *D2* katsayılarının ve *A3* katsayılarının ortalamalarının (*D2M* ve *A3M*) istikrarları %100'e yakın oldukları için veri kümesinden çıkarılmışlardır. *DI* katsayılarının standart sapmaları (*DIS*) korelasyonları %0'a yakın olduğu hesaplanmıştır. Sonuca etkisi olmadığı için *DIS* veri kümesinden çıkarılabilir. Buna göre elde edilen sonuçlar Çizelge 4.1 ve 4.2'de verilmiştir.

Çizelge 4.1 Normalizeli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri.

	<b>LR</b>	<b>KNN</b> ( <i>Minkowski</i> )	<b>SVM</b> ( <i>RBF</i> )	<b>SVM</b> ( <i>Sigmoid</i> )	<b>NB</b>	<b>DT</b>	<b>RF</b>	<b>MLPNN</b>
<b>TCC</b>	<b>0.972</b>	0.944	0.889	0.889	<b>1.0</b>	<b>1.0</b>	0.916	<b>1.0</b>
<b>TPR</b>	1.0	0.9375	0.8	0.875	1.0	1.0	0.8421	1.0
	0.047	0.05	0	0.1	1.0	1.0	0	1.0
<b>(CM)</b>	$\begin{bmatrix} 15 & 1 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 15 & 1 \\ 1 & 19 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 4 & 16 \end{bmatrix}$	$\begin{bmatrix} 14 & 2 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 3 & 17 \end{bmatrix}$	$\begin{bmatrix} 10 & 0 \\ 0 & 6 \end{bmatrix}$



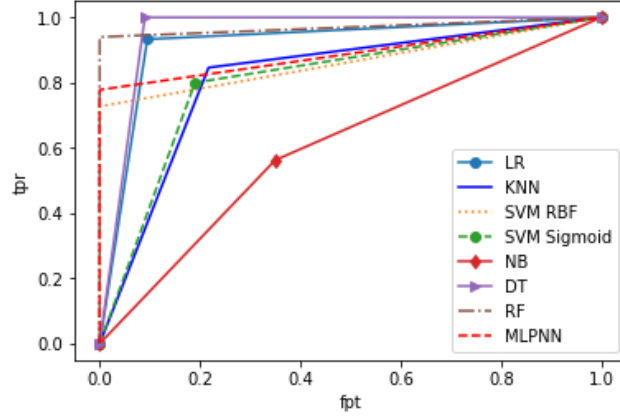
Şekil 4.1 Normalizeli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri

Normalizeli işaretler ile oluşturulan veri kümesi ile yapılan deneysel çalışmada en yüksek başarı TCC=1.0 ile NB, DT ve MLPNN, TCC=0.972 ile LR'dir. Ayrıca diğer ML sınıflayıcılar da yüksek doğruluk oranlarına sahiptir. Şekil 4.1'de ROC eğrilerinin altına bakılarak da bu görülmektedir.

Çizelge 4.2 Normalizesiz işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri.

	LR	KNN (Minkowski)	SVM (RBF)	SVM (Sigmoid)	NB	DT	RF	MLPNN
<b>TCC</b>	0.916	0.805	0.834	0.806	0.611	0.944	<b>0.972</b>	0.875
<b>TPR</b>	0.933	0.8461	0.727	0.8	0.5625	1.0	0.94	0.788
<b>FPR</b>	0.095	0.217	0.0	0.19	0.35	0.09	0.0	0.0
<b>CM</b>	$\begin{bmatrix} 14 & 2 \\ 1 & 19 \end{bmatrix}$	$\begin{bmatrix} 11 & 5 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 12 & 4 \\ 3 & 17 \end{bmatrix}$	$\begin{bmatrix} 9 & 7 \\ 7 & 13 \end{bmatrix}$	$\begin{bmatrix} 14 & 2 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 1 & 19 \end{bmatrix}$	$\begin{bmatrix} 7 & 2 \\ 0 & 7 \end{bmatrix}$





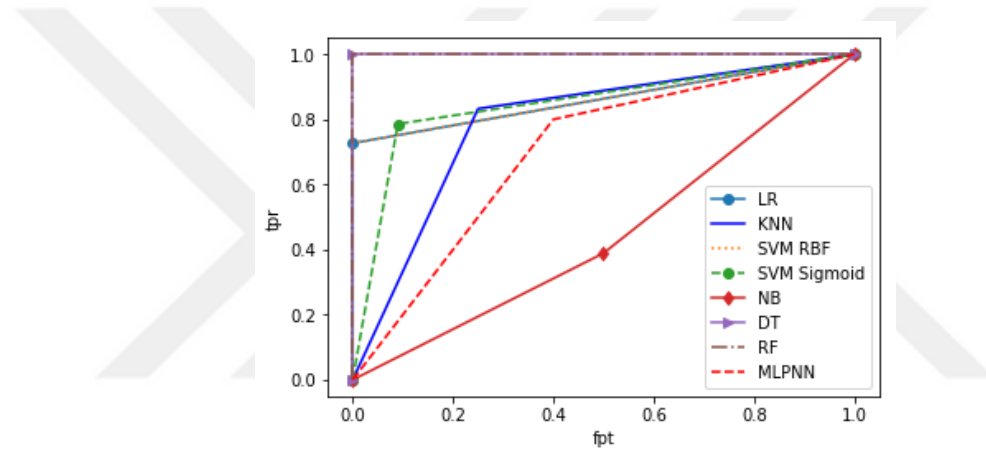
Şekil 4.2 Normalizesiz işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri

Normalizesiz işaretlerden oluşan veri kümesiyle yapılan deneysel çalışmalarda en yüksek başarı TCC=0.972 ile RF'dir. Çizelge 4.2'deki TCC oranlarına ve Şekil 4.2'deki ROC eğrilerinin altındaki alanlara bakılarak, normalizeli işaretlerden oluşan veri kümesiyle yapılan deneylere göre başarının genel olarak düştüğü gözlemlenmektedir. Başarısı yüksek olduğundan, bundan sonraki denemeler normalize edilmiş ses işaretleri üzerinde yapılmıştır.

Normalizeli ses işaretlerinin eğimi alınarak elde edilen yeni işarete Fourier ve 3 seviyeli dalgacık dönüşümü uygulanmıştır. Elde edilen yeni işaretlere temel istatistiksel parametreler uygulanıp yeni veri kümesi oluşturulmuştur. Bu veri kümesinde 3. seviye dalgacık dönüşümünün D katsayısının harmonik ortalamasının ( $D3h$ ), sanal ve reel kısımların varyanslarının ( $Ivar$ ,  $Rvar$ ) korelasyon katsayıları %0'a yakın olduğundan veri kümesinden çıkarılabilir. Oluşturulan yeni veri kümesi ML sınıflayıcıların girişine verilmiştir. Elde edilen sonuçlar Çizelge 4.3 ile verilmiştir.

Çizelge 4.3 Normalizeli eğimli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların başarı ölçütleri

	LR	KNN ( <i>Minkowski</i> )	SVM ( <i>RBF</i> )	SVM ( <i>Sigmoid</i> )	NB	DT	RF	MLPNN
TCC	0.833	0.778	0.833	0.806	0.44	1.0	1.0	0.75
TPR	0.727	0.833	0.727	0.786	0.389	1.0	1.0	0.8
FPR	0.0	0.25	0.0	0.09	0.5	0.0	0.0	0.4
CM	$\begin{bmatrix} 16 & 0 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 10 & 6 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 11 & 5 \\ 3 & 17 \end{bmatrix}$	$\begin{bmatrix} 7 & 9 \\ 11 & 9 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 9 & 2 \\ 3 & 2 \end{bmatrix}$



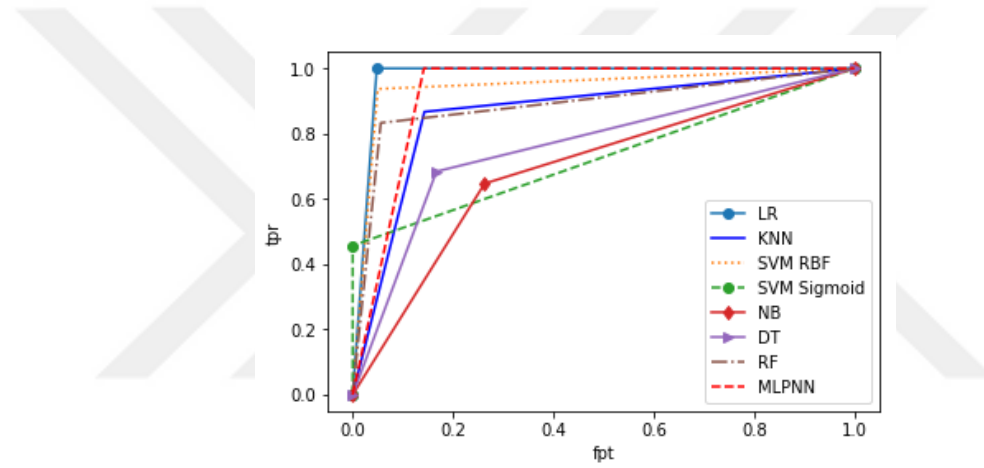
Şekil 4.3 Normalizeli Eğimli işaretlerde istatistiksel parametrelere dayalı ML sınıflayıcıların ROC eğrileri

Normalizeli işaretlerin eğimi alınarak oluşturulan yeni veri kümesiyle yapılan deneysel çalışmalarda TCC=1.0 ile DT ve RF, TCC=0.833 ile LR ve SVM (*RBF*)'dir. Çizelge 4.3'teki doğruluk oranlarına ve Şekil 4.3'teki ROC eğrileri altında kalan alanlara bakılarak MLPNN, SVM kernel *Sigmoid* seçildiğinde ve KNN de genel olarak yüksek doğruluk oranlarına sahiptir.

Normalizeli işaretlerin zaman domeninde noktalar arasındaki eğimi alınarak elde edilen işarete temel istatistiksel parametrelerin yanı sıra, entropi ve Ki-Karesi'nin entropisi alınarak başarımın artırılması için yeni veri kümesi oluşturulmuştur. Yeni veri kümesi PD teşhisi için ML sınıflayıcının girişine verilmiştir. Elde edilen sonuçlar Çizelge 4.4'te verilmiştir.

Çizelge 4.4 ZD’de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların başarı ölçütleri

	LR	KNN ( <i>Minkowski</i> )	SVM ( <i>RBF</i> )	SVM ( <i>Sigmoid</i> )	NB	DT	RF	MLPNN
<b>TCC</b>	<b>0.972</b>	0.861	0.944	0.472	0.694	0.778	0.888	0.93
<b>TPR</b>	1.0	0.867	0.937	0.457	0.647	0.684	0.833	1.0
<b>FPR</b>	0.048	0.143	0.05	0.0	0.263	0.167	0.056	0.142
<b>CM</b>	$\begin{bmatrix} 15 & 1 \\ 0 & 20 \end{bmatrix}$	$\begin{bmatrix} 13 & 3 \\ 2 & 18 \end{bmatrix}$	$\begin{bmatrix} 15 & 1 \\ 1 & 19 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 19 & 1 \end{bmatrix}$	$\begin{bmatrix} 11 & 5 \\ 6 & 14 \end{bmatrix}$	$\begin{bmatrix} 13 & 3 \\ 5 & 15 \end{bmatrix}$	$\begin{bmatrix} 15 & 1 \\ 3 & 17 \end{bmatrix}$	$\begin{bmatrix} 9 & 1 \\ 0 & 6 \end{bmatrix}$



Şekil 4.4 ZD’de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların ROC eğrileri

ZD’de eğimi alınan normalizeli işaretin Ki-Kare ve entropiye dayalı veri kümesi ML sınıflayıcıların girişine verildiğinde en yüksek başarı TCC=0.972 ile LR’dır. Çizelge 4.4’teki doğruluk oranlarına ve Şekil 4.4’teki ROC eğrilerin altına bakılarak SVM (*RBF*) ve MLPNN’nin de yüksek doğruluk oranına sahip olduğu gözlenmiştir. Yine KNN, SVM kernel *sigmoid*, DT ve RF de yüksek başarıya sahiptir.

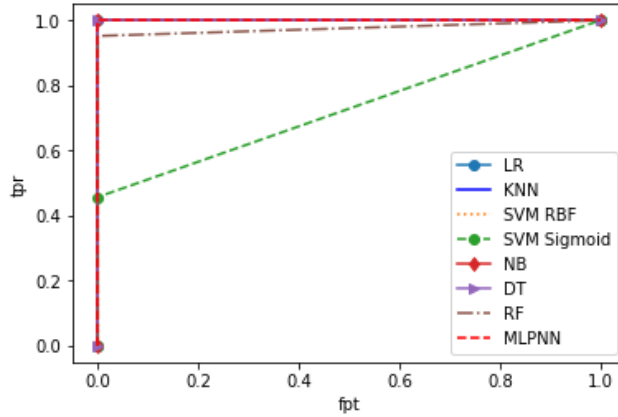
ZD’de eğimi alınan normalizeli işaretin Ki-Kare ve entropiye dayalı veri kümesi genel olarak bakıldığında normalizeli işaretlerin eğimine istatistiksel parametreler uygulanıp oluşturulan veri kümesine göre başarısı oldukça yüksektir. Eğimi alınan normalizeli işarete entropi ve Ki-Kare uygulamak başarıyı arttırmıştır.

ZD’de eğimi alınan normalizeli işaretin Ki-Kare ve entropiye dayalı veri kümesi normalizeli işaretlerin noktalar arası eğiminin frekans domeninde başarısının ölçülmesi

için Fourier dönüşümünde sonra bu işarete temel istatistiksel parametreler, entropi ve Ki-Karesi'nin entropisi uygulanarak yeni veri kümesi oluşturulmuştur. Yeni veri kümesi ML sınıflayıcının girişine verilmiştir. Elde edilen sonuçlar Çizelge 4.5 ile verilmiştir.

Çizelge 4.5 FD'de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların başarı ölçütleri

	LR	KNN (Minkowski)	SVM (RBF)	SVM (Sigmoid)	NB	DT	RF	MLPNN
<b>TCC</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	0.472	<b>1.0</b>	<b>1.0</b>	0.9722	<b>1.0</b>
<b>TPR</b>	1.0	1.0	1.0	0.457	1.0	1.0	0.952	1.0
<b>FPR</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>CM</b>	$\begin{bmatrix} 20 & 0 \\ 0 & 16 \end{bmatrix}$	$\begin{bmatrix} 20 & 0 \\ 0 & 16 \end{bmatrix}$	$\begin{bmatrix} 20 & 0 \\ 0 & 16 \end{bmatrix}$	$\begin{bmatrix} 16 & 0 \\ 19 & 1 \end{bmatrix}$	$\begin{bmatrix} 20 & 0 \\ 0 & 16 \end{bmatrix}$	$\begin{bmatrix} 20 & 0 \\ 0 & 16 \end{bmatrix}$	$\begin{bmatrix} 20 & 0 \\ 1 & 15 \end{bmatrix}$	$\begin{bmatrix} 10 & 0 \\ 0 & 6 \end{bmatrix}$



Şekil 4.5 FD'de eğimi alınan işaretin Ki-Kare ve entropiye dayalı ML sınıflayıcıların ROC eğrileri

FD'de eğimi alınan normalizeli işarete Ki-Kare ve entropi uygulanarak yapılan deneysel çalışmalarda SVM kernel *sigmoid* hariç tüm yöntemler yüksek doğruluk oranlarına sahiptir. Çizelge 4.5 ve Şekil 4.5'e bakılarak LR, KNN, SVM kernel *RBF*, NB, DT ve MLPNN sınıflayıcıların TCC'si 1.0'dır. FD'de eğimi alınan normalizeli işarete Ki-Kare ve entropi uygulanarak oluşturulan veri kümesiyle yapılan deneyler tüm veri kümeleriyle yapılan deneylere kıyasla en yüksek başarıya sahiptir.

### BFS Kullanılarak Gerçekleştirilen Deneyler

BFS kullanılarak özellik vektörlerinin seçimi *Anaconda-Spyder* programında yapılmıştır. BFS veri kümesindeki en uygun özelliğin bulunması ve bellek açısından tasarruf edilmesi için uygulanmıştır. Tüm veri kümesindeki özellikler başlangıçta veri kümesine dâhil edilerek  $P$  değeri (olasılık değeri) hesaplanmıştır. Daha sonra adım adım en büyük  $P$  değerine sahip özellikler çıkarılmış ve en sonunda veri kümesinde sadece  $P$  değerleri 0 ve 0'a yakın olan özellikler kalmıştır. Çizelge 4.6'da çıkarılan öznelikler verilmiştir.

Çizelge 4.6 BFS'den sonra veri kümesinden çıkarılan öznelikler

Temel istatistiksel parametrelerden oluşan veri kümesindeki P-value değeri 0 ve 0'a yakın olan veriler	1. dalgacık dönüşümü sonrasında elde edilen $D$ katsayılarının harmonik ortalaması
	FFT sonrasında elde edilen imajiner katsayıların varyansları
	FFT sonrasında elde edilen imajiner kısımların, 3. Dalgacık dönüşümü sonrasında elde edilen $D$ katsayılarının ortalaması

Normalizeli işaretlere temel istatistiksel parametreler uygulandığında en yüksek sonucu 0.889 ile DT, normalizesiz işaretlerde temel istatistiksel parametreler uygulandığında 0.95 ile MLPNN elde etmiştir.

## 5. SONUÇ

Bu çalışmada, hastalardan ve sağlıklı gönüllülerden kaydedilen vokal kord işaretlerinin zaman, frekans ve zaman-frekans domenlerinde istatistiksel parametreler yardımıyla özellik vektörleri çıkarılmış ve bu vektörler PD tanısı için makine öğrenmesi (ML) temelli sınıflayıcı modellere girdi olarak uygulanmıştır. Bu amaç için, vokal kord işaretlerine DFT kullanılarak elde edilen frekans spektrumu üzerinde ve DWT uygulanarak ayrıştırılan frekans alt-bantları üzerinde istatistiksel parametreler ve bu çalışmada PD teşhisinde ilk defa noktalar-arası eğim kullanılarak elde edilen özellik vektörleri ve ML sınıflayıcı modellerin sınıflandırma başarılarına önemli katkılar sağlamıştır.

Tüm veri kümeleri içinde, genel olarak uygulanan tüm ML sınıflayıcı modellerinde FD'de işaretin eğimi alınıp elde edilen işarete ve işaretin kendisine Ki-Karesi ve entropisi alınan veri kümesiyle yapılan deneyler en başarılı olmaktadır. LR, KNN, SVM kernel *RBF*, NB, DT ve MLPNN sınıflayıcıların TCC'si 1.0'dır. ZD'de LR TCC=0.972 ve MLPNN TCC=0.93'tür. İşareti frekans düzlemine taşımak başarıyı arttırmıştır. Ancak veri kümesinde verilerin dağılımından dolayı hem ZD'de hem de FD'de SVM sınıflayıcı kullanılırken kernel olarak *sigmoid* seçmek başarıyı azaltmıştır. Bu veri kümesinde SVM sınıflayıcıda kernel olarak *RBF* seçmek uygundur. Normalizeli işarete istatistiksel parametrelerin uygulandığı veri kümesiyle yapılan deneylerde ise MLPNN, NB TCC=1.0, LR TCC ise 0.972'dir. Normalizesiz işarete temel istatistiksel parametreler uygulanarak oluşturulan veri kümesiyle yapılan deneylerde RF TCC=0.972 ve LR TCC=0.916'dır. Normalizasyon başarıyı arttırmıştır.

BFS uygulandıktan sonra tüm veri kümeleri ve tüm yöntemler kıyaslandığında ise eğim uygulanarak elde edilen özellik vektörlerinin kullanımı MLPNN modeli ile en yüksek başarı sağlamıştır. Ayrıca, düşük belleğe ihtiyaç duyması ve hızlı olmasından dolayı da tercih edilebilir. Normalizeli işaretlerde istatistiksel parametrelerin oluşturduğu veri kümesinin MLPNN ye uygulanması yüksek başarı vermiştir.

Sonuç olarak, vokal kord işaretlerinden makine öğrenmesi kullanılarak PD teşhisinin yüksek başarı oranlarında gerçekleştirilmesi sağlanmıştır. Özellikle bu çalışmada ilk defa önerilen noktalar-arası eğimin PD teşhisinde sınıflayıcıların doğru sınıflandırma başarı oranlarına katkı yaptığı gösterilmiştir.

## 6. KAYNAKLAR

- Ai, L., Wang, J. ve Wang, X., 2008. Multi-features fusion diagnosis of tremor based on artificial neural network and d-s evidence theory. *Signal Processing*, 88(12), 2927-2935.
- Anonim, Parkinson's Data Set, <https://archive.ics.uci.edu/ml/datasets/parkinsons> (21.01.2018).
- Anonim, Parkinsons telemonitoring data set, <https://archive.ics.uci.edu/ml/datasets/parkinsons+telemonitoring> (21.01.2018).
- Ayhan, Ş. ve Erdoğan Ş., 2014. Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 9(1), 175- 198.
- Ayhan, S., 2006. Sıralı Lojistik Regresyon Analiziyle Türkiye'deki Hemşirelerin İş Bırakma Niyetini Etkileyen Faktörlerin Belirlenmesi. Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Eskişehir Osmangazi Üniversitesi.
- Baby, M.S., Saji, A.J. ve Kumar, C.S., 2017. Parkinson's disease classification using wavelet transform based feature extraction of gait data. *International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Kollam India.
- Chandrashekar, P., 2014. Mathematically modeling the GPE/STN neuronal cluster to account for Parkinsonian tremor and developing a novel method to accurately diagnose Parkinson's disease using speech measurements and an artificial neural network. *Quantitative Biology, Neurons and Cognition*, <https://arxiv.org/abs/1405.1314>.
- Chen, H.L., Huang, C. C., Yu, X.G., Xu, X., Sun, X., Wang, G. ve Wang, S.J., 2013. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. *Expert Systems with Applications*, 40(1), 263-271.
- Erdogdu Sakar, B., Isenkul, M., Sakar, C.O., Sertbas, A., Gurgun, F., Delil, S., Apaydin ve H., Kursun, O., 2013. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings, *IEEE Journal of Biomedical and Health Informatics*, 17(4), 828-834.
- Guruler, H., 2017. A novel diagnosis system for Parkinson's disease using complex-valued artificial neural network with k-means clustering feature weighting method. *Neural Computing & Applications*, 28(7), 1657.
- Hamilton, D., List, A., Butler, T., Hogg, S. ve Cawley, M., 2006. Discrimination between parkinsonian syndrome and essential tremor using artificial neural network classification of quantified DaTSCAN data. *Nuclear Medicine Communications*, 27(12), 939-944.
- Hlavica J., Prauzek M., Peterek, T. ve Musilek, P., 2016. Assessment of Parkinson's disease progression using neural network and ANFIS models. *Neural Network World*, 26(2), 111-128.
- Jankovic, J., 2007. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry*, 79(4), 368-376.
- Khan, M.M., Mendes, A., Zhang, P. ve Chalup, S., 2017. Evolving multi-dimensional wavelet neural networks for classification using Cartesian Genetic Programming. *Neurocomputing*, 247, 39-58.

- Kılınç D., Borandağ E., Tunalı V., Şimşek M. ve Özçift A., 2016. KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi* 2016, 3: 89-94.
- Langston, J.W., 2002. Parkinson's disease: current and future challenges. *Neurotoxicology*, 23(4-5), 443-450.
- Muniz, A., Liu, W., Liu, H., Lyons, K.E., Pahwa, R., Nobre, F.F. ve Nadal, J., 2009. Assessment of the effects of subthalamic stimulation in Parkinson disease patients by artificial neural network. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society Engineering in Medicine and Biology Society*, 5673-5676.
- Nilashi, M., Ibrahim, O. ve Ahani, A., 2016. Accuracy improvement for predicting Parkinson's disease progression. *Scientific Reports*, 6(34181). Doi:10.1038/Srep34181.
- Ozcift, A., 2012. SVM feature selection based rotation forest ensemble classifiers to improve computer-aided diagnosis of Parkinson disease. *Journal of Medical Systems*, 36(4), 2141-2147.
- Özdamar, K., 2009. *Paket Programlar ve İstatistiksel Veri Analizi*, Kaan Kitabevi, Eskişehir.
- Soner V., 2011. *Veri Madenciliğinde Kullanılan Yöntemler*. <https://volkansoner.wordpress.com/2011/09/15/veri-madenciliginde-kullanilan-yontemler/>; (23.06.2019).
- Tsanas, A., Little, M.A., Mcsharry, P.E. ve Ramig, L.O., 2010. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE Transactions on Biomedical Engineering*, 57(4):884-893.
- Yılancıoğlu K. 2017. Vocal cord measure based ANN approach for prediction of Parkinson's diseases status. *SDU Journal of Health Science Institute / SDÜ Sağlık Bilimleri Enstitüsü Dergisi*, 1, 139-142.
- Zhang, H., Gu, C., Gu, L. ve Zhang, Y. (2011). The evaluation of tourism destination competitiveness by TOPSIS & information entropy a case in the Yangtze river delta of China, *Tourism Management*, 32(2), 443-451.



## 7. ÖZGEÇMİŞ

Büşra Zeynep GÜREL, 8 Ekim 1993 yılında Samsun'da doğdu. Okul hayatına sırası ile Bayındır İlköğretim Okulu'nda başladı. Ortaöğretimini Samsun Tülay Başaran Anadolu Lisesi Fen Bilimleri bölümünden 83.10/(100) ortalamayla ile tamamladı. Lisans eğitiminden Samsun Ondokuz Mayıs Üniversitesi Elektrik Elektronik Mühendisliği bölümünden 2015 tarihinde 74.77 / 100 ortalama ile mezun oldu. Eylül 2019 yılında Erzurum İspir Mimar Sinan Mesleki ve Teknik Anadolu Lisesi'nde Elektrik Elektronik Teknolojileri Öğretmeni olarak göreve başladı. Yüksek Lisans eğitimine Gaziosmanpaşa Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Elektrik-Elektronik Mühendisliği öğrencisi olarak devam etmektedir.