T.C.

TRAKYA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

BEYİN DALGALARININ ÖLÇÜMÜNDE KULLANILAN ALGORİTMALARIN İNCELENMESİ VE GÜÇ ANALİZİ

Murat AKPINAR

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

Tez Danışmanı: Dr. Öğr. Üyesi FATMA BÜYÜKSARAÇOĞLU SAKALLI

EDİRNE-2023

Murat AKPINAR'ın hazırladığı "**BEYİN DALGALARININ ÖLÇÜMÜNDE KULLANILAN ALGORİTMALARIN İNCELENMESİ VE GÜÇ ANALİZİ**" başlıklı bu tez, tarafımızca okunmuş, kapsam ve niteliği açısından **Bilgisayar Mühendisliği Anabilim Dalında** bir **Yüksek lisans tezi** olarak kabul edilmiştir.

Jüri Üyeleri İmza

Dr. Öğr. Üyesi Fatma BÜYÜKSARAÇOĞLU SAKALLI

Dr. Öğr. Üyesi Özlem AYDIN

Dr. Öğr. Üyesi HALİL NUSRET BULUŞ

Tez Savunma Tarihi: 19/01/2023

Bu tezin Yüksek Lisans tezi olarak gerekli şartları sağladığını onaylarım.

İmza

Dr. Öğr. Üyesi Fatma BÜYÜKSARAÇOĞLU SAKALLI ……

Tez Danışmanı

Trakya Üniversitesi Fen Bilimleri Enstitüsü onayı

……

Prof. Dr. Hüseyin Rıza Ferhat KARABULUT

Fen Bilimleri Enstitüsü Müdürü

T.Ü.FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI

DOĞRULUK BEYANI

Trakya Üniversitesi Fen Bilimleri Enstitüsü, tez kurallarına uygun olarak hazırladığım bu tez çalışmasında, tüm verilerin bilimsel ve akademik kurallar çerçevesinde elde edildiğini, kullanılan verilerde tahrifat yapılmadığını, tezin akademik ve etik kurallara uygun olarak yazıldığını, kullanılan tüm literatür bilgilerinin bilimsel normlara uygun bir şekilde kaynak gösterilerek ilgili tezde yer aldığını ve bu tezin tamamı ya da herhangi bir bölümünün daha önceden Trakya Üniversitesi ya da farklı bir Üniversite tez çalışması olarak sunulmadığını beyan ederim.

19/01/2023
Murat AKPINAR
İmza

Yüksek Lisans Tezi

Beyin Dalgalarının Ölçümünde Kullanılan Algoritmaların İncelenmesi ve Güç Analizi

T.Ü. Fen Bilimler Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

ÖZET

Her yıl dünya da binlerce epilepsi teşhisi konulmaktadır. Teşhis konulan kişilerin çoğunda geç teşhis konulduğundan tedavi imkanları zorlaşmaktadır. Erken teşhis konulması durumunda tedaviye yanıt verme ihtimalleri yüksek olup iyileşme oranlarını ciddi derece de etkilemektedir.

Dünya çapında en çok ölümle sonuçlanan hastalıklar nörolojik hastalıklardır. Nörolojik bir hastalık olan epilepsinin erken teşhis konulan çocuklarda iyileşme durumlarının yüksek oranlarda başarılı olduğu gözlemlenmektedir. Epilepsi teşhisinde kullanılan EEG cihazları günümüzde doktorların en büyük yardımcısıdır. Bu nedenle belli bir kısmı değil de tüm insanlığa faydası olan bu cihazın doğru tespit ile erken tanı koyması hayati derecede önemlidir.

Bu tez çalışmasında EEG cihazlarında kullanılmakta olan algoritmaların çalışma şekli incelenecektir. İncelenen algoritmalar sonucunda cihaz kullanılması önerilen yeni bir süreç önerilecektir. Önerilecek olan algoritmanın diğer algoritmalara göre hata payının düşük ve performans olarak da yavaş olmaması beklenmektedir.

Yıl : 2023

Sayfa Sayısı : 53

Anahtar Kelimeler : EEG, Teşhis, Epilepsi, Nöroloji, Beyin Sinyali, Extreme Learning Machine, K-Nearest Neighbours

Master Thesis

Analysis of Algorithms Used in Measuring Brain Waves and Power Analysis

Trakya University Institute of Natural and Applied Science

Department Of Computer Engineering

ABSTRACT

Every year, thousands of epilepsy diagnoses are made around the world. Since most of the diagnosed people are diagnosed late, treatment opportunities become difficult. In case of early diagnosis, the probability of responding to treatment is high and it seriously affects the recovery rates.

The diseases that cause the most deaths worldwide are neurological diseases. It is observed that children with early diagnosis of epilepsy, a neurological disease, have high rates of recovery. EEG devices used in the diagnosis of epilepsy are the greatest assistants of doctors today. For this reason, it is vitally important that this device, which is beneficial to all humanity, not a certain part, makes an early diagnosis with accurate detection.

In this thesis, the working of the algorithms used in EEG devices will be examined. As a result of the analyzed algorithms, a new process that is recommended to use the device will be proposed. It is expected that the proposed algorithm will not have a low margin of error and slow performance compared to other algorithms.

Year **:** 2023

Number of Pages **:** 53

Keywords **:** EEG, Diagnosis, Epilepsy, Neurology, Brain Signaling, Extreme Learning Machine, K-Nearest Neighbors

TEŞEKKÜR

Başta aileme, çalışmalarımda emeği geçen ve yararlandığım kaynaklardaki tüm hocalarıma, meslektaşlarıma, çalışmam boyunca profesyonel olarak algoritmalar kullanarak yazılım işi yapan özel sektörde yer alan arkadaşlarıma, akademik ve insani açıdan bilgi-birikimi ile her zaman yanımda olup katkılarını, yardımlarını hiç esirgemeyen değerli tez danışmanım Dr. Öğr. Üyesi Fatma BÜYÜKSARAÇOĞLU SAKALLI hocama teşekkürü bir borç bilirim.

İÇİNDEKİLER

ÖZET.....iv

ABSTRACT.....v

TEŞEKKÜR.....vi

İÇİNDEKİLER..... vii

KISALTMALAR DİZİNİ.....x

ŞEKİLLER DİZİNİ.....xi

ÇİZELGELER DİZİNİ..... xii

BÖLÜM 1 GİRİŞ.....1

BÖLÜM 2 GENEL BİLGİLERİ.....3

2.1. Elektroensefalografi (EEG) 3

2.2. EEG Ölçümünün Zorluğu.....4

2.3. EEG Ölçümünün Önemi 4

2.4. EEG Dalga Boyları 5

2.5. Epilepsi 7

BÖLÜM 3 ÖLÇÜMDE KULLANILAN ALGORİTMALAR 8

3.1 Naive Bayes Algoritması 9

3.1.1. Naive Bayes Teoremi 9

3.1.2. Naive Bayes Uygulama Alanları 10

3.1.3. Naive Bayes Algoritması ile EEG Kullanımı 11
3.2. K-En Yakın Komşu Algoritması 11
3.2.1. Distance (Uzaklık) 12
3.2.1.1. Öklid Fonksiyonu 12
3.2.1.2. Minkowski Uzaklığı 13
3.2.2. KNN (K-Nearest Neighbors) Nasıl Çalışır? 14
3.2.3. KNN (K-Nearest Neighbors) Uygulama Alanları 14
3.2.4. KNN Algoritması ile EEG Kullanımı 14
3.3. Hidden Markov Model 15
3.3.1. Markov Chain (Markov Zinciri) 15
3.3.1.1. MC (Markov Chain) Uygulamaları 16
3.3.2. Hidden Markov Model Uygulama Alanları 17
3.3.3. Hidden Markov Model ile EEG Kullanımı 17
3.4. Extreme Learning Machine (ELM) 17
3.4.1. Extreme Learning Machine Uygulama Alanları 18
3.4.2. Extreme Learning Machine ile EEG Kullanımı 19
3.5. Bağımsız Bileşenler Analizi (Independent Component Analysis, ICA) 19
3.5.2. Bağımsız Bileşenler Analizi Uygulama Alanları 20
3.5.3. Bağımsız Bileşenler Analizi ile EEG Kullanımı 20
3.6. Çok Katmanlı Algılayıcılar (Multi-Layer Perceptron, MLP) 21
3.6.1. Çok Katmanlı Algılayıcıların Dezavantajları 22
3.6.2. Çok Katmanlı Algılayıcıların Uygulama Alanları 22
3.6.3. Çok Katmanlı Algılayıcılar ile EEG Kullanımı 23

3.7. Karar Ağaçları (DECISION TREES) 23
3.7.1. Karar Ağaçları (Desicion Trees) Yapısı 23
3.7.2. Karar Ağaçlarının Avantajları 24
3.7.3. Karar Ağaçlarının Dezavantajları 25
3.7.4. Karar Ağaçları Uygulama Alanları 25
3.7.5. Karar Ağaçları ile EEG Kullanımı 26
BÖLÜM 4 YENİ ALGORİTMA ÖNERİSİ 27
4.1 Veri Temizlemenin Önemi 28
4.2 Öznitelik Çıkartma 29
4.3 ELM İncelemesi 31
4.3.1 Araştırma Hakkında 31
4.3.2 Veri Seti Hakkında 31
4.3.3 ELM Algoritmasının Kullanma Nedenleri 32
4.3.4 Sonuç 33
4.4 KNN (K-En Yakın Komşu Algoritması) İncelemesi 33
4.4.1 Araştırma Hakkında 34
4.4.2 Veri Seti Hakkında 34
4.4.3 k Değerinin Belirlenmesi 36
4.4.4 KNN Algoritmasının Kullanma Nedenleri 36
4.4.5 Sonuç 36
BÖLÜM 5 SONUÇ 37
KAYNAKLAR 38
ÖZGEÇMİŞ 41

KISALTMALAR DİZİNİ

BIST : Borsa İstanbul

DVM : Destek Vektör Makineleri

ÇKA : Çok Katmanlı Algılayıcılar

EEG : Elektroensefalografi

ELM : Extreme Learning Machines (Aşırı Makine Öğrenmesi)

EKG : Elektrokardiyografi

EOG : Elektookülografi

fMRI : Functional Magnetic Resonance Imaging (Fonksiyonel Manyetik Rezonans Görüntüleme)

GHz : Gigahertz

HMM : Hidden Markov Model YSA (Yapay Sinir Ağları)

ICA : Independent Component Analysis (Bağımsız Bileşenler Analizi)

KNN : K-Nearest Neighbors (K-En Yakın Komşu)

LM : Levenberg-Marquardt

MC : Markov Chain (Markov Zinciri)

MEG : Manyetoensefalografi

MLP : Multi-Layer Perceptron

SLFN : Single hidden Layer Feed-forward Neural networks

SNR : Signal To Noise Ratio (Sinyal Gürültü Oranı)

YSA : Yapay Sinir Ağları

WHO : Dünya Sağlık Örgütü

XOR : Doğrusal olmayan

ŞEKİLLER DİZİNİ

Şekil 2. 1. Beyin Dalga Tipleri....3
Şekil 2. 2. Farklı durumlardaki EEG ölçümleri (Sivakumar, 2022)....4
Şekil 2. 3. Alfa ve Beta Dalga Boyu Ölçüm Örneği....6
Şekil 2. 4. Delta ve Teta Dalga Boyu Ölçüm Örneği....6
Şekil 3. 1. Naive Bayes Teoremi....10
Şekil 3. 2. Öklid Formülü....13
Şekil 3. 3. Öklid Uzaklık Gösterimi....13
Şekil 3. 4. Minkowski Uzaklığı Formülü....13
Şekil 3. 5. Markov Zinciri....15
Şekil 3. 6. Markov Zinciri Örnek Metin....16
Şekil 3. 7. Markov Zinciri Örnek Çıktılar....16
Şekil 3. 8. ELM Çalışma Mimarisi....18
Şekil 3. 9. Multi-Layer Perceptron Çalışma Mimarisi....22
Şekil 3. 10. Desicion Tree Gösterim 1....24
Şekil 3. 11. Desicion Tree Gösterim 2....24
Şekil 4. 1. Yeni Algoritma Süreci....27
Şekil 4. 2. Öznitelik Çıkarma Metodları....30
Şekil 4. 3. incelenen ELM Örneğinde Kullanılan Adımlar....31
Şekil 4. 4. ELM de İncelenen Araştırma da Kullanılan Algoritma Yapısı....33

ÇİZELGELER DİZİNİ

Çizelge 2. 1. EEG Kullanım Alanları 5

Çizelge 2. 2. Epilepsi vakaları ve buna bağlı ölüm oranları (Fatlawi & Kiss, 2022).....7

Çizelge 3. 1. Standart Bir Algoritma Süreci.....9

Çizelge 3. 2. K-NN Çalışma Mantığı.....12

Çizelge 4. 1. Bonn Üniversitesi Örnek EEG Verileri.....32

Çizelge 4. 2. AFYON Hisse Senedi Verileri 35

Çizelge 4. 3. k=1 için Test Sonuçları.....36

BÖLÜM 1

GİRİŞ

EEG işlemi, beyinde gerçekleşen elektriksel aktiviteleri ölçümlemek için yapılmaktadır. Başta epilepsi hastalığı olmak üzere bilinç ve algı bozuklukları, uyku bozuklukları gibi birçok sinirsel hastalığın tespitinde kullanılır. EEG ölçümü, nörolojik hastalıkların sebeplerinde ve teşhis belirlemeye önemli katkısı olan bir tetkik durumuna gelmiştir (Gül vd., 2021).

Dünya Sağlık Örgütünün açıkladığı istatiksel sonuçlara göre ölümlere yol açan hastalıklar da en büyük orana sahip hastalık nörolojik hastalıklardır. Türkiye de 750.000 civarında epilepsi rahatsızlığı bulunan insan olduğu düşünülmektedir. Bu sayı Dünya genelinde ise 65.000.000 olduğu tahmin edilmektedir (Türk Nöroloji Derneği, 2022).

Belirtilen bu ciddi sayılar her gün gittikçe artmaktadır. Bu hastalıktaki ölüm oranı da göz ardı edilmeyecek derecede yüksektir. Bu da epilepsiyi nörolojik hastalıklardan en yaygın olanı olarak karşımıza çıkarmaktadır.

Epilepsi teşhisinin konulmasından çok teşhisin ne zaman ve hangi evrede konulduğu önemlidir. Erken teşhis ile geç teşhis konulan hastaların tedavi süreci ve hastanın tedaviye verdiği cevap ciddi derece de fark göstermektedir. Erken teşhis konulan hastalar hızlı şekilde normal yaşantılarına uyum sağlayabilmekte ve kendi yaşamlarını bir düzen içerisine getirebilmektedir.

Doktorların epilepsi teşhisi koyarken başta gelen yardımcıları EEG cihazlarıdır. EEG cihazlarında ölçümlerinin yorumlanması ve EEG cihazının vardığı sonuç doktorun görüşünü önemli derece de etkilemektedir.

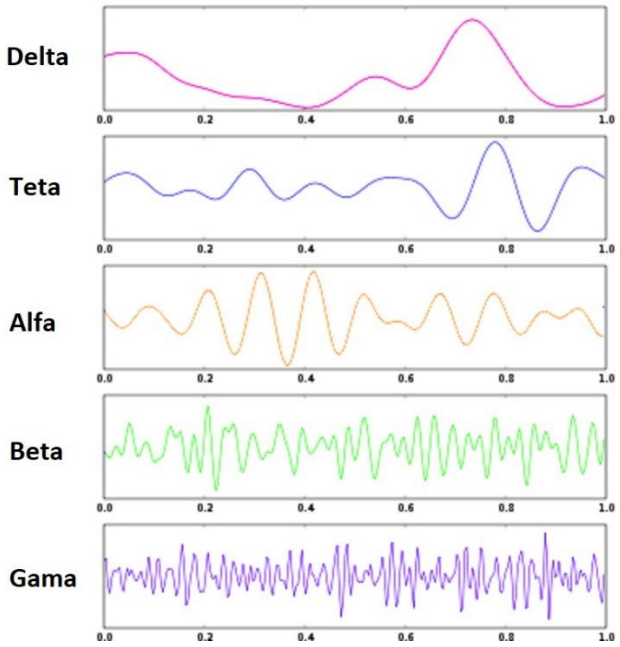
Bu yüzden EEG cihazının çıkarım yapacağı sonuca ulaşırken kullandığı algoritmalar konusunda ufak bir değişiklik yapmak bile doktorun teşhisini büyük ölçüde etkileyecektir.

BÖLÜM 2

GENEL BİLGİLERİ

2.1. Elektroensefalografi (EEG)

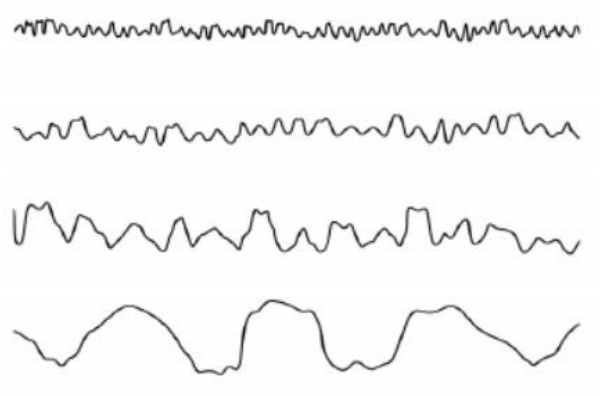
EEG işlemi, beyinde gerçekleşen elektriksel aktiviteleri ölçümlemek için yapılmaktadır. Bu işlem de beyinde yer alan sinir hücrelerinin vücut durumuna bağlı olarak ürettiği elektriksel aktiviteleri kâğıda yazarak beyin dalgalarını temsil etmektedir. Beyindeki sinyallerin dalga tipleri farklıdır (Şekil 2.1).



Şekil 2. 1. Beyin Dalga Tipleri

2.2. EEG Ölçümünün Zorluğu

EEG ölçümündeki sinyallerin dalga boyları çok yüksek değildir. Ölçüm sırasında dışarıdan çok fazla farklı sinyaller gelebilmektedir. Bu etkenlerin yanı sıra hastanın fiziksel hareketliliği de ayrı sinyaller üretmektedir. Şebeke, elektriksel gürültüler, EKG, EOG vb. gibi dış faktörler olabilmektedir. Bu faktörlerin etkisi ölçümü bozma yönünde olup EEG sinyalinde yer alan bilginin kaybolmasına kadar ciddi sonuçlara varmaktadır. Kişilerin durumlarına göre bu ölçümler farklılık gösterebilir (Şekil 2.2).



Şekil 2. 2. Farklı durumlardaki EEG ölçümleri (Sivakumar, 2022)

2.3. EEG Ölçümünün Önemi

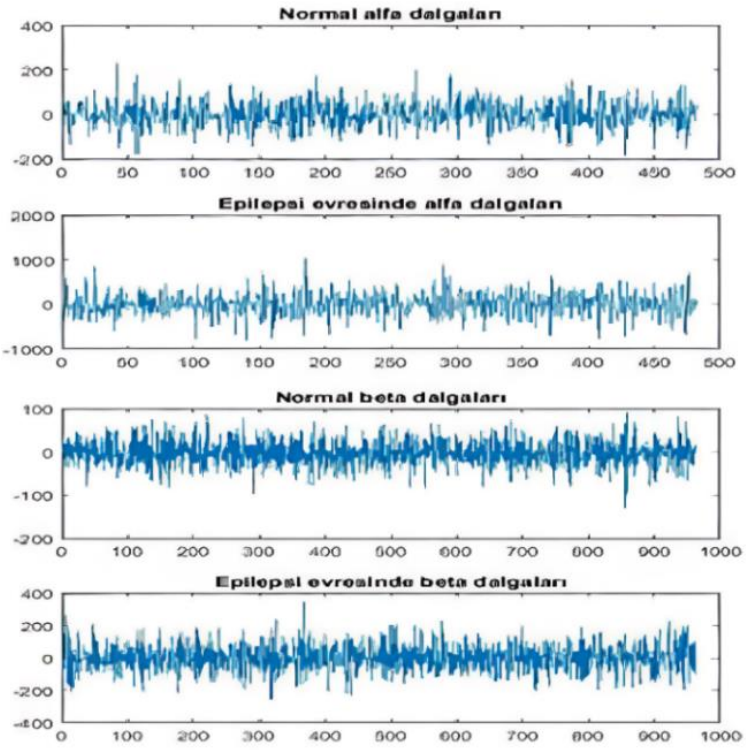
Başta epilepsi hastalığı olmak üzere bilinç ve algı bozuklukları, uyku bozuklukları gibi bir çok sinirsel hastalığın tespitinde kullanılır. EEG ölçümü, nörolojik hastalıkların sebeplerinde ve teşhis belirlemeye önemli katkısı olan bir tetkik durumuna gelmiştir. EEG cihazının farklı alanlarda ilgili alanlara özgü şekilde kullanımları bulunmaktadır (Çizelge 2.1).

Çizelge 2. 1. EEG Kullanım Alanları

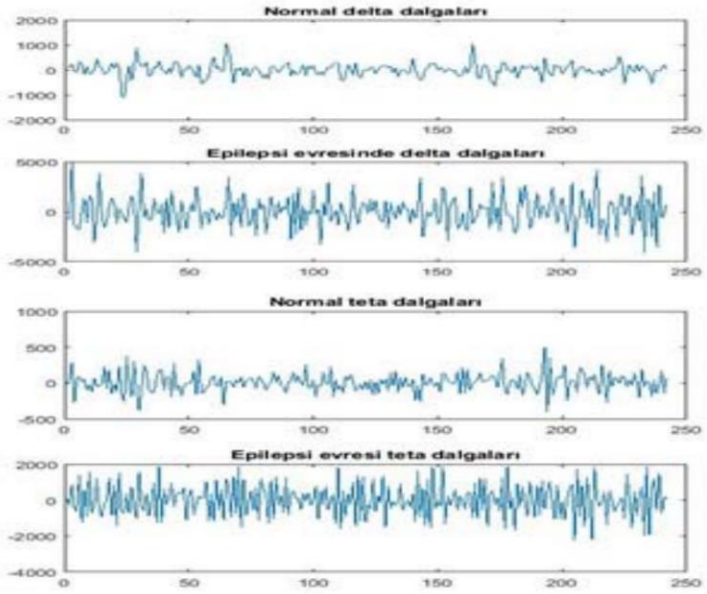
Alanı	Kullanım Şekli
Nöroloji	Hastanın beyin patolojisinin belirlenmesinde
Beyin Cerrahisi	Beyin ameliyatla çıkartılacak tümör gibi anormal patolojik dokuların yerinin belirlenmesinde
Anestezi	Anestesi altındaki hastanın anestezi seviyesinin belirlenmesinde
Pediatri	Yeni doğmuş çocukların görme ve duyma problemlerinin belirlenmesinde
Psikiyatri	Zihinsel bir bozuklukta beyinde hastalığında var olup olmadığının belirlenmesinde

2.4. EEG Dalga Boyları

EEG işlemleri esnasında beyinden alınan dalga boyları ölçüm yapılan kişinin durumuna göre değişiklik göstermektedir. Normal, nöbet durumu, uyku hali vs. durumlarında dalga boyları kişinin beyninin durumunu yansıtmaktadır. Şekil 2.3 ve Şekil 2.4'te örnek dalga boyları değinilmektedir.



Şekil 2. 3. Alfa ve Beta Dalga Boyu Ölçüm Örneği

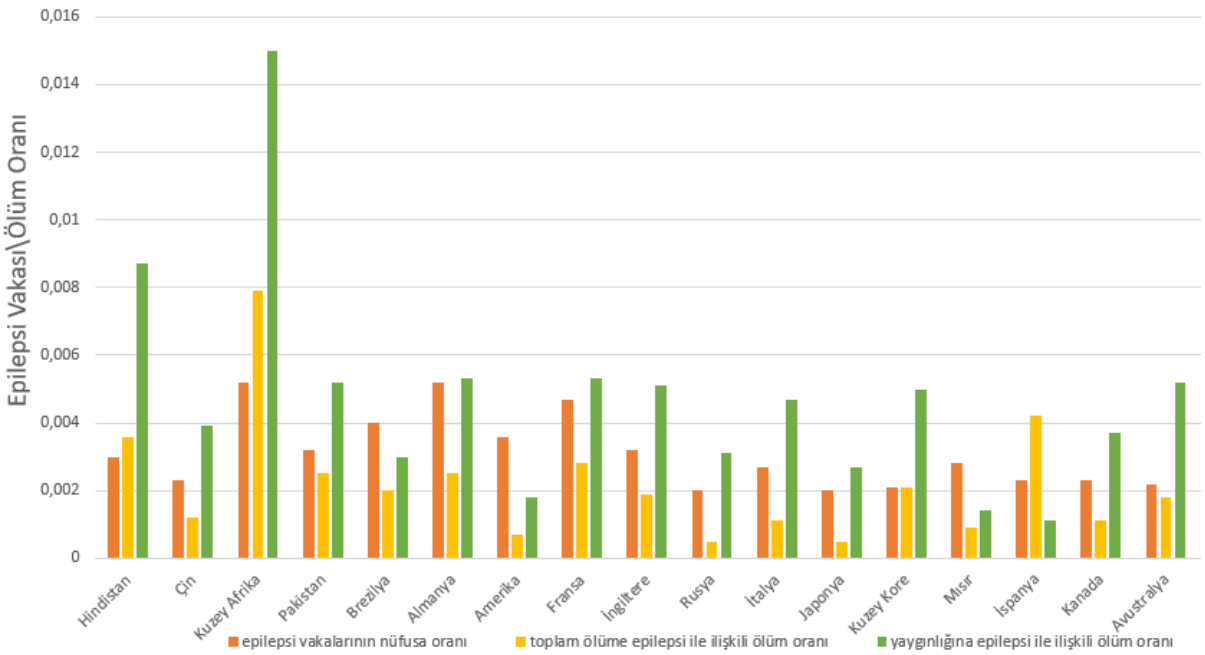


Şekil 2. 4. Delta ve Teta Dalga Boyu Ölçüm Örneği

2.5. Epilepsi

Bulaşma özelliği olmayan ve beyinde etkisini gösteren bir hastalıktır. Vücudun tamamına veya bellir bir bölümüne etki edebilmektedir. Hasta, hastalığın vücudunda kalması boyunca nöbetler geçirmektedir. Bu nöbetlerde kısa süreli de olsa mesane veya bağırsak fonksiyonu bazen de bilinç kaybı gibi ataklar yer almaktadır. Nöbetler yılda bir defa da olabilir günde birden fazla kez de gerçekleşebilmektedir. Her yıl dünyada yaklaşık 5 Milyon kişiye bu teşhis konulmaktadır. Teşhis konusunda ise nöbetlerin etiyolojik hazırlanmış belgeleri ve EEG çizelgesi en tutarlı belirleyicileridir. Epilepsi hastalığı ile ilgili bazı ülkelerdeki ölüm oranı ile ilgili durumu Çizelge 2.2 de görülmektedir.

Çizelge 2. 2. Epilepsi vakaları ve buna bağlı ölüm oranları (Fatlawi & Kiss, 2022)



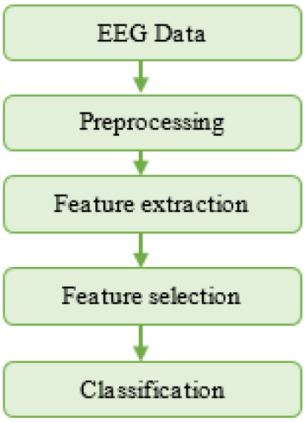
BÖLÜM 3

ÖLÇÜMDE KULLANILAN ALGORİTMALAR

Bu bölümde EEG cihazlarında kullanılan aşağıda listenen algoritmalar incelenecektir. Standart bir algoritma süreci Çizelge 3.1 de yer almaktadır.

- Naive Bayes Algoritması
- K-En Yakın Komşu Algoritması
- Hidden Markov Model
- Extreme Learning Machine (ELM)
- Independent Component Analysis (ICA)
- Multi-Layer Perceptron
- Decision Trees

Çizelge 3. 1. Standart Bir Algoritma Süreci



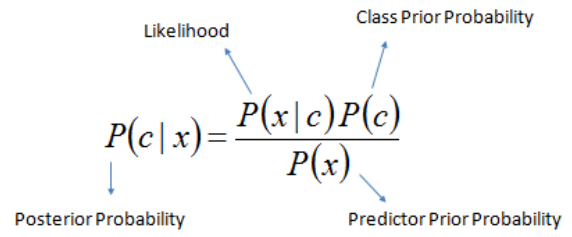
3.1 Naive Bayes Algoritması

Thomas Bayes'ten adını alan bu sınıflandırıcı, herhangi bir özelliğin başka bir özellikle ilişkili olmadığını kabullenmektedir.

Bayes modelini oluşturmak veri büyüklüklerinden bağımsız olduğundan kolaydır. Veri setinde az sayıda hatalı verinin olması veya gürültünün olması algoritma sonucuna etki etmemektedir. Gelen verinin boyutu çok önemli değildir çünkü bayes sınıflandırıcısı gruplayarak işlem yapar. Naive Bayes'in son derece karmaşık diğer algoritmalardan bile daha iyi performans ortaya koyulduğu gözlenmiştir.

3.1.1. Naive Bayes Teoremi

Bayes sınıflandırılmasında kullanılan Bayes Teoreminin matematiksel gösterimi şu şekildedir.



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Şekil 3. 1. Naive Bayes Teoremi

Teoremdeki değişkenleri inceleyecek olursak;

P(c|x); x olayı gerçekleştiği durumda c olayının meydana gelme olasılığıdır (koşullu olasılık).
P(x|c); c olayı gerçekleştiği durumda x olayının meydana gelme olasılığıdır.
P(c) ve **P(x)**; c ve x olaylarının önsel olasılıklarıdır.

3.1.2. Naive Bayes Uygulama Alanları

- Veri madenciliğinde,
- Biyomedikal mühendisliği alanında,
- Hastalıkların ya da anormalliklerin tıbbi tanımlanmasında (otomatik olarak mühendislik ürünü tıbbi cihazlar tarafından tanı konulması),
- Elektrokardiyografi (EKG) grafiğinin sınıflandırılmasında,
- Elektroansefalografi (EEG) grafiklerinin ayrıştırılmasında,
- Genetik araştırmalarında,
- Yığın mesaj tanımlanmasında,
- Metin ayrıştırılmasında,
- Ürün sınıflandırma ve diğer bazı alanlarda kullanılır.

3.1.3. Naive Bayes Algoritması ile EEG Kullanımı

Bayes algoritması, kullanılan veri setinin doğruluğunu bulmak için tercih edilen popüler sınıflandırma yöntemidir.

Literatür taraması sonucunda Bayes algoritmasının performans değerleri ortalama düzeyde olduğu diğer algoritmalara kıyasla doğruluk değerinin ciddi fark alarak sonuçlandığı gözlemlenmemiştir.

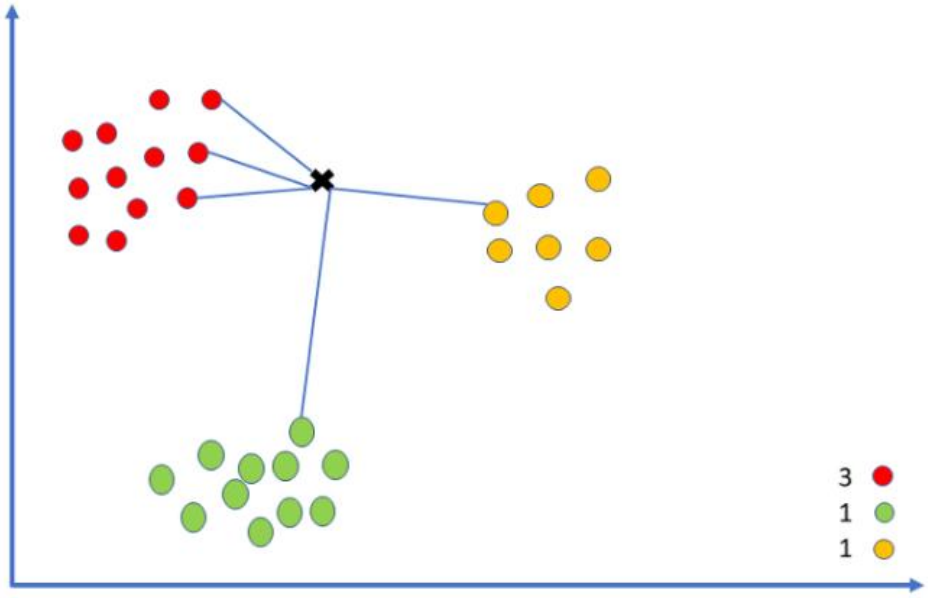
3.2. K-En Yakın Komşu Algoritması

KNN (K-Nearest Neighbors), komşuların arasında durumu baz alarak yakınlık derecesine göre çalışma mantığı yer almaktadır. k-en yakın komşu isminde yer alan k parametresi de komşu sayısı değerini barındırmaktadır. Yakınlık hesaplandıktan sonra k parametresine göre işlem yapılacak komşular belirlenmektedir.

Bu algoritma parametrelerle çalışmamaktadır. Çalışması için hazırlanan veri kümesini inceler ve bu eğitim verisini ezberlemektedir. Tahmin işlemi yapılacağı zaman tüm veri setine bakarak tahmin edilecek verinin komşularını arayıp sonrasında bu komşularla işlem yapmaktadır. Çalışma şekli ve yapısı Çizelge 3.2'de belirtilmiştir.

K (komşuluk sayısı): yakınlık derecesine göre kaç komşu üzerine hesaplama gerçekleştireceğini bu parametre ile belirlemekteyiz. Hesaplamayı direkt etkileyen parametre olduğu için sonucun üzerinde çok büyük etkisi vardır. O yüzden en uygun k değerini belirlemek için test verileri üzerinde denemeler yapılıp hata oranına göre seçmek gerekmektedir.

Çizelge 3. 2. K-NN Çalışma Mantığı



3.2.1. Distance (Uzaklık)

Hesaplanacak noktanın veri setinde yer alan noktalarla arasındaki uzaklık hesaplanmaktadır. Bu hesaplama için birden çok fonksiyonlar yer almaktadır.

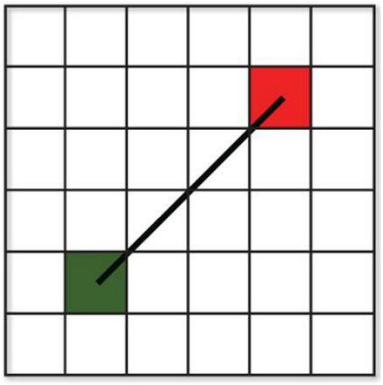
3.2.1.1. Öklid Fonksiyonu

KNN Algoritmasında en çok tercih edilen uzaklık hesaplama fonksiyonudur.

Öklid fonksiyonun çalışma mantığına göre uzaklığı hesaplanacak noktalar arasında doğrusal bir yol olduğu varsayılmaktadır. Bu doğrusal yolu hesaplamak aradaki uzaklığı bulmamızı sağlayacaktır. Uzaklık hesaplama için aşağıda yer alan Öklid Formülü (Şekil 3.2) kullanılmaktadır (Şekil 3.3).

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Şekil 3. 2. Öklid Formülü



Şekil 3. 3. Öklid Uzaklık Gösterimi

3.2.1.2. Minkowski Uzaklığı

Hesaplanacak iki nokta için Minkowski Uzaklığı aşağıda yer alan fomüle göre hesaplanmaktadır. İki noktanın temsili x ve y değişkenleri kullanılmaktadır. İndis olarak da i harfi kullanılmaktadır (Şekil 3.4).

p iki nokta arası yerine Minkowski mesafesi(uzalığı)

$$P = (x_1, x_2, \ldots, x_n) \text{ and } Q = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$$

aşağıdaki şekilde tanımlanır:

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{1/p}.$$

Şekil 3. 4. Minkowski Uzaklığı Formülü

3.2.2. KNN (K-Nearest Neighbors) Nasıl Çalışır?

KNN çalışma mantığına göre ilk ve önemli konularından biri k değerinin hesaplanmasıdır. Bir nokta geldiğinde k değerine göre komşular seçilerek ilgili nokta ile arasındaki uzaklık hesaplaması yapılmaktadır.

Literatür taramasına göre KNN algoritmalarında uzaklık hesaplaması olarak yaygın şekilde Öklid Fonksiyonu kullanılmaktadır. Öklid fonksiyonuna alternatif olarak aşağıdaki uzaklık hesaplama fonksiyonları da kullanılmaktadır.

- Minkowski
- Manhattan
- Hamming

3.2.3. KNN (K-Nearest Neighbors) Uygulama Alanları

- Yüz Tanıma
- El Yazısı Tanıma
- Hastalık Belirleme
- E-Posta Spam Filtrelemesi
- Müşteri Segmentasyonu
- Kişiye Özel İlaç veya İhtiyaç Tespiti

3.2.4. KNN Algoritması ile EEG Kullanımı

Literatür taraması sonucunda KNN algoritmasının doğruluk değerlerinin diğer algoritmalara göre daha yüksek çıktığı gözlemlenmiştir.

İncelenen bir araştırmada Bayes algoritması %62 civarında doğruluk değeri alırken KNN ise %81 civarında doğruluk değeri sergilemektedir (Özkan, 2019).

Literatür taraması sonucunda KNN algoritmasının EEG veri seti üzerinde eğitim süresinin diğer algoritmalara göre daha az sürdüğü görülmektedir (Dursun & Akın, 2021).

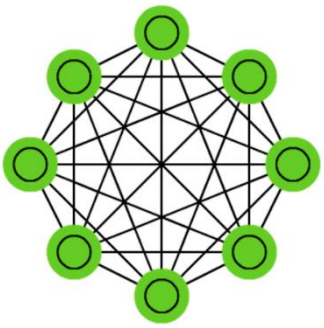
Literatür araştırması esnasında incelenen bir makale de EEG veri seti üzerinde sınıflandırma yapılırken Yapay Sinir Ağları ve Lineer Diskriminant Analiz sınıflandırmasına göre KNN algoritması daha iyi sonuç vermiştir (Çağlıyan & Köse, 2021).

3.3. Hidden Markov Model

Bu sınıflandırma modelini 1960 yıllarında LE Baum ve meslektaşları ile birlikte ortaya sunuldu. Modelin temel dayanağı mevcut ve geçmiş durumdaki olasılıkları kullanarak bir değişkenin gelecekteki durumunu tahmin etmektir. Bir Markov zinciri, Markov modeline göre zincirdeki ara çıktıların gözlemci tarafından görülememesidir. Hidden kelimesinden de anlaşılacağı üzere model içerisinde gözlemleyici tarafından belirlenmiş saklı durumlar vardır. Buradaki kritik durum ise gözlemleyicinin bu durumları bir mantığa göre oluşturmasıdır. Gizli durumlar sonucu oldukça etkileyebilmektedir.

3.3.1. Markov Chain (Markov Zinciri)

Stokastik (rastlantısal) ve dinamik sistemlerin analizinde kullanılmaktadır. Ayrıca bir sistemin zaman boyunca verilen girdilere göre verdiği çıktıların incelenmesinde kullanılır. Günümüzde yaygın olarak kullanıldığı yerlere örnek olarak otomatları verebiliriz. Yapısal olarak gösterimi olarak aşağıdaki gibi temsil edilmektedir (Şekil 3.5).



Şekil 3. 5. Markov Zinciri

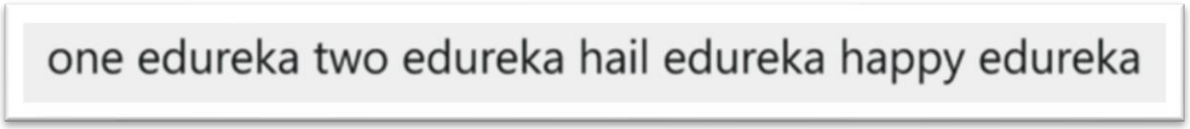
3.3.1.1. MC (Markov Chain) Uygulamaları

MC, istatiksel olarak rastlantısal olmayan süreçlerin modellerinin çıkarılmasında kullanılmaktadır. Yazı oluşturmadan ekonomiksek durumların modellemeye kadar kullanılmaktadır.

Literatür taramasına bakıldığında MC kullanımın en çok kullanıldığı sektör olarak yazı oluşturma ve otomatik tamamlama olduğu görülmektedir. Bu konuda aşağıdaki örneği inceleyebilirsiniz (Şekil 3.6, Şekil 3.7).

Uygulama Örneği

Tahmin edilecek metin:



Şekil 3. 6. Markov Zinciri Örnek Metin

Tahmin Edilen Çıktılar:



Şekil 3. 7. Markov Zinciri Örnek Çıktılar

3.3.2. Hidden Markov Model Uygulama Alanları

- Yüz Tanıma
- El Yazısı Tanıma
- Hastalık Belirleme
- Konuşma Tanıma
- Termodinamik
- Ekonomi, Finans
- Mimik Tanıma
- Konuşmadan Kelime Seçme

3.3.3. Hidden Markov Model ile EEG Kullanımı

Literatür taramasında karşılaşın bir çalışmadaki alttaki yorumdan yola çıkarak veri setindeki gürültü konusunda kullanılabileceğini düşünülebilir.

EEG işaretleri değişkenlik gösterebilir ve ölçüm esnasındaki dış faktörler gibi sinyal içindeki gürültüleri temizleyip SNR oranını azaltmaktadır (Yanar & Mishchenko, 2016).

Literatür taramasında HMM ile yeni bir algoritma öneren diğer bir çalışmada incelendiğinde yeni oluşturulan algoritmanın diğer algoritmalara göre önemli ölçüde daha performanslı çalıştığı görülmektedir (Wang vd., 2018).

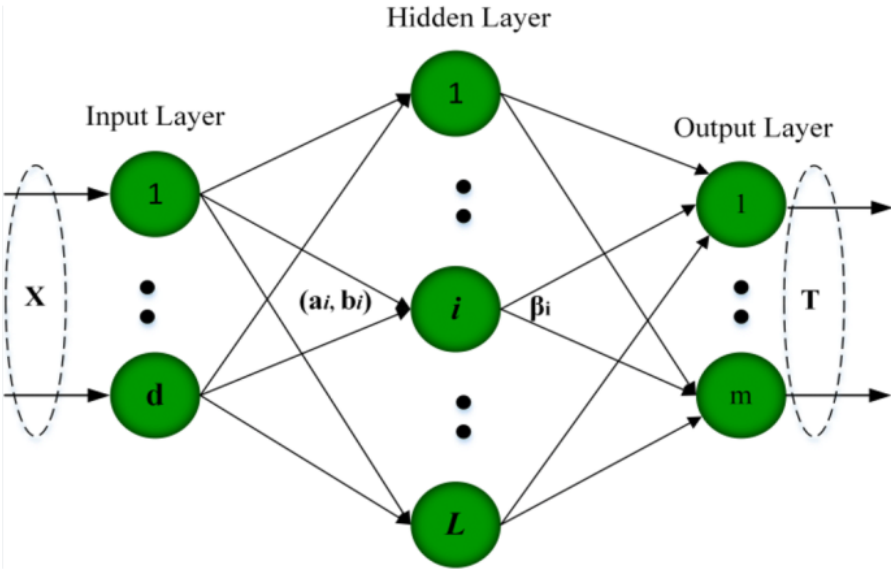
3.4. Extreme Learning Machine (ELM)

EEG işaretlerinin formundaki değişim gözle analiz edilebilmektedir. Gözle yorumlanmayacak durumların olduğu da bilinmektedir. Bu insan gözü ile yorumlanmayan bölümler için detaylı bilgi verilmesi ve teşhisin erken konulması için bilişim sektöründen faydalanması gerekmektedir. Bilişim sektöründe makine öğrenmesinin her sektördeki

başarısı göz önüne alındığında doktorlara tanı koyma sırasında önemli avantaj sağlayabilecektir.

DVM, YSA, Karar Ağaçları ve kaba küme yaklaşımı gibi yöntemlerle EEG sinyallerini yapay zeka işlemleri analiz edebilimektedirler.

ELM, giriş değerlerinin belirli olmayan çıkış değerlerinin ise analitiğe dayalı hesaplanarak yapan tek gizli katmanı olan ileri beslemeli (Single hidden Layer Feed-forward Neural networks, SLFN) bir YSA (Yapay Sinir Ağları) modelidir (Şekil 3.8).



Şekil 3. 8. ELM Çalışma Mimarisi

3.4.1. Extreme Learning Machine Uygulama Alanları

- Tıp
- Kimya
- Ulaşım
- Ekonomi
- Robotik

- IoT
- El Yazısı Tanıma
- Yüz Tanıma
- 3D Nesne Tanıma
- Parmak İzi Eşleştirme

3.4.2. Extreme Learning Machine ile EEG Kullanımı

Literatürde incelenen bir çalışmaya göre bir kullanım açıklaması yapılabilmektedir. Bu açıklamayı şu şekilde özetlenebilir.

Farklı şekilde çalışan ve farklı nöron sayıları olan iki veri kümesi için 2 ELM modeli bulunmuştur. Birinci veri kümesiyle ilgili olan model de yer alan gizli katmanın içinde "Tringular basis" aktivasyon fonksiyonu kullanılmaktadır. 57 nöron kullanılarak bu model %100 başarı sağlanmıştır. Diğer veri kümesi ile ilgili olan modelde ise farklı aktivasyon fonksiyonu ile diğer test verisinden de farklı olarak farklı nöronlar kullanılarak %100 başarı gözlemlenmiştir (Kaya &Tekin, 2012).

3.5. Bağımsız Bileşenler Analizi (Independent Component Analysis, ICA)

BBA/ICA, çeşitli değişkenelere sahip sayısal bilgilerin özgün özniteliklerini (feature) çıkarmaya yarayan bir yöntemdir. Bağımsız kaynaklardan gelen sinyallerin birbirinden ayrıştırılmasıdır. Temel olarak farklı sinyallerin farklı fiziksel kaynaklar tarafından oluşturulduğu düşüncesiyle çalışır. Örneğin iki tür canlının çıkardığı sesler, iki farklı sinyal olarak düşünülür. Bu iki sinyalin kaynakları farklı olduğundan birbirinden bağımsızdır.

Benzer şekilde insan beyninin belli bölgelerinden elde edilmek istenilen sinyaller çevre etkenlerinin kolay bir şekilde engellenememesi sebebiyle farklı kaynakların bileşimi şeklinde elde edilir.

ICA algoritması modeli, ‘önişleme’ ve ‘tahmin’ olmak üzere iki aşamada gerçekleştirilir.

Önişleme: ICA’nın ilk aşamasıdır ve bu kısımda dönüşüm yapılacak veri türleri varsa burada o işlem yapılır. Bu dönüşümler bir sonraki aşama için hazırlık olarak görülebilir.

Tahmin: ICA’nın diğer aşamasıdır. Bu kısımda ise farklı kaynak işaretleri belirlenmeye çalışılır ve bağımsız öznitelikler tespiti yapılır.

3.5.2. Bağımsız Bileşenler Analizi Uygulama Alanları

- **Biyomedikal uygulamalar**: EEG (Elektroensefalogram), fMRI (fonksiyonel manyetik rezonans görüntüleme) ve MEG (Manyetoensefalografi) sinyalleri gibi farklı biyomedikal sinyallerle karışan artefaktları gidermek için kullanıldı.
- **Ses sinyali işleme**: gürültüyü gidermek için ses sinyallerinde yaygın olarak kullanılmaktadır.
- **Biyometri**: yüz tanıma gibi farklı biyometrilerdeki ayırt edici özellikleri çıkarmak ve parmak izi konularında kullanıldı.
- **Görüntü işleme**: orijinal görüntüden farklı katmanları çıkarmak için görüntü işlemede kullanılır.

3.5.3. Bağımsız Bileşenler Analizi ile EEG Kullanımı

Literatür incelemesi yapıldığında aşağıdaki sonuca ulaşabiliyoruz.

Beynin çalışması gereği bölgelerin işaretleri durumlarına göre değişiklik göstermektedir. Yani hasarlı bölge ile hasarsız bölge arasında sinyal farkı olacağından hasarlı bölge tespit edilmesi mümkün olabilecektir. Araştırmalar da yaygın olarak kullanımı sinyallerin gürültüden temizlenmesinde olduğu gözlemlenmiştir (Sezgin vd., 2012).

3.6. Çok Katmanlı Algılayıcılar (Multi-Layer Perceptron, MLP)

Bilişim sektörünün gelişmesi ile birlikte çözülemeyen problemler için sinir ağları kullanımı artmaktadır. Ayrıca medikal uygulamalarda başarıyla kullanılmıştır.

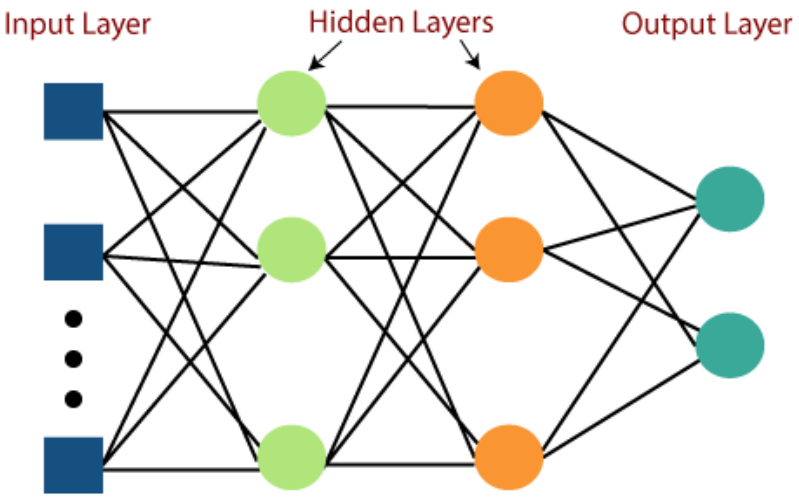
Doğrusal olmayan (XOR) problemleriçin basit algılayıcı modeller yetersiz olduğu bilinmektedir. 1969 yılında Papert ve Minsk'in basit algılayıcı modeli geliştirerek iki katmanlı ileri beslemeli bir ağ yapısı sunmuşlardır. Ancak geliştirdikleri çözüm hala giriş katmanından ara (gizli) katmana olan ağırlıkların nasıl belirleneceğine ilişkin bir çözüm olmamıştır. Burada eksik olan gizli katmanların ağırlıklarını belirlemek için kullanılabilecek öğrenme kuralının olmamasıdır.

1986 yılında Hinton ve Williams adında araştırmacılar YSA ile geri yayılım işlemi kullanarak XOR probelimini çözmüşlerdir. Buradan da devam ederek YSA konusunda araştırmalar ivme kazanmıştır.

İnsan beyni baz alınarak öğrenme süreci konusunda detaylı incelemeler yapılmıştır. Bu incelemeler sonrasında matematiksel bir model ortaya sunulmuştur. Bu model için Çok Katmanlı Yapay Sinir Ağları isimlendirilmesi yapılmıştır.

Çok Katmanlı Yapay Sinir Ağları çalışma mantığı olarak beyindeki sinir ağlarına benzetilmektedir. Sinir ağları gibi hatırlama, genelleme ve öğrenme güçleri vardır. Bu güçler çok benzer yapıya sahip olunduğundan taklit edildiği düşünülmektedir.

Aynı anda birden fazla işlem yapma durumunda işin içine paralellik girdiği için birden fazla katman anlamına gelmektedir. Single Perceptron Model'den farklı bir yapıya sahip olan Çok Katmanlı Yapay Sinir Ağlarında arada gizli katmanlar yer almaktadır (Şekil 3.9).



Şekil 3. 9. Multi-Layer Perceptron Çalışma Mimarisi

3.6.1. Çok Katmanlı Algılayıcıların Dezavantajları

- **Donanım Bağımlılığı**: Paralel işlemlerin yapılması için yüksek işlemcilere ihtiyaç duymaktadır.
- **Ağ Güvensizliği**: Ağ da yapılan işleme dair detaylı bilgi olmadığı için ağa güvensizlik vardır.
- **Ağ Yapısının Belirlenmesi**: Uygun ağ yapısı deneme yanılma ile bulunabilmektedir.
- **Verilerin Dönüştürülmesi**: Ağlar nümerik değerler ile çalıştığı için problemler dönüştürülmek zorundadır. Bu da kullanıcıya bir yük bindirmekte ve kullanıcının veriyi dönüştürürken hata yapma olasılığı ortaya çıkmaktadır.

3.6.2. Çok Katmanlı Algılayıcıların Uygulama Alanları

- Otomotiv alanında yol izleme
- Bankacılıkta kredi kartı suçu tespiti
- Uzay sanayinde uçuş simülasyonu
- Otomatik pilot uygulamaları
- Finans sektöründe ise döviz kuru tahminleri

3.6.3. Çok Katmanlı Algılayıcılar ile EEG Kullanımı

Yapılan bir çalışmadaki kullanım şekline örnek vermek gerekirse;

İncelenen bir araştırma da verilerin toplanmasında toplam katılanların yarısı sağlıklı bireylerdir. Bu bireylerden alınan veriler de 3200 segment bulunmaktadır. Kurulan mimari de 1 veya 2 gizli katman, 16 giriş ve 2 çıkıştan oluşmaktadır. Ara katmanlar da sigmoid fonkisyonu kullanırken çıkış da doğrusal fonksiyon, aktivasyon fonksiyonu olarak tasarlanmıştır. Sınıflandırma işleminde %99.62 başarını oranı tespit edilmiştir (Toprak & Caglar, 2007).

3.7. Karar Ağaçları (DECISION TREES)

Araştırmalar da gözetimli öğrenme olarak en çok tercih edilen algoritma Karar Ağaçlarıdır. Genel itibariyle ele alınan bütün problemler (sınıflandırma ve regression) için uygun bir formu bulunabilmektedir. Ayrıca bu algoritmanın veri madenciliği sınıflandırma konularında da yeteneği bulunmaktadır.

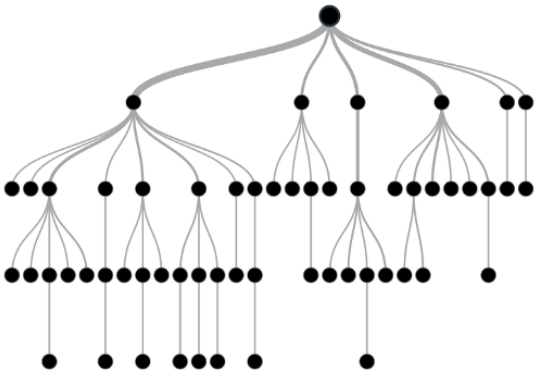
Mimari olarak ağaç yapısına benzemektedir. Karar düğümleri yapı içerisinde önemli rol oynamaktadır. Sınıflama, hedef ve özelliğe göre Decision Nodes (karar düğümleri) bir yanıt vermektedir. Bu modelin diğer bir yapı öğesi olarak da yaprak düğümleri yer almaktadır.

3.7.1. Karar Ağaçları (Desicion Trees) Yapısı

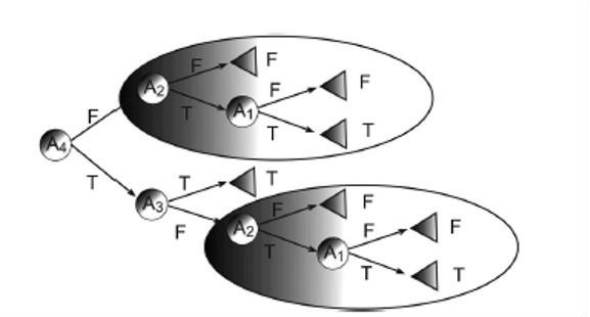
Karar ağaçlarının modelindeki ilk düğüm için kök düğüm adı verilir.

Karar ağaçları veri türü olarak sayısal ve kategorik verilerini kullanmaktadırlar.

Yapısındaki temel ilke küçük gruplar oluşturmaktır. Verinin boyutunun büyüklüğünden çok küçük gruplara ayırırken kullanabilecek karar mekanizmalarına odaklıdır. Basit kararlar halinde bölerek küçük grupları elde etmektedir (Şekil 3.10, Şekil 3.11).



Şekil 3. 10. Desicion Tree Gösterim 1



Şekil 3. 11. Desicion Tree Gösterim 2

3.7.2. Karar Ağaçlarının Avantajları

- Anlaması ve yorumlaması kolaydır.
- Az oranda bir veri hazırlığına ihtiyaç duyar.
- Hem sayısal hem de kategorik verileri işleyebilir.
- Çok çıktılı problemleri ele alabilmektedirler.
- İstatistiksel testler kullanılarak bir modelin doğrulanması mümkündür.

3.7.3. Karar Ağaçlarının Dezavantajları

- Veriyi iyi bir şekilde açıklamayan aşırı karmaşık ağaçlar üretilebilir. Bu durumda ağaç dallanması takip edilemeyebilir.
- Kullanılan ağacın maliyeti, ağacı eğitmek için kullanılan veri noktalarının sayısına bağlı olarak kat ve kat artabilmektedir.
- Ezbere öğrenme yaşanabilir ("over-fitting"). Bu problemin çözümü için model parametrelere kısıtlamalar ve budama gibi yöntemler kullanılabilir. Budama işlemi, az sayıda nesneyi barındıran yaprak düğümlerin karar ağacı grafiğinden atılmasını ifade etmektedir

3.7.4. Karar Ağaçları Uygulama Alanları

- Geçmiş satış verilerine baz alarak genişleme fırsatlarını değerlendiren teknoloji işletmesi.
- Sınırlı reklam bütçesini nereye hedefleyeceğine karar veren bir oyuncak şirketi, müşterilerin satın alma olasılığının ne olduğunu gösteren demografik verilere dayanarak.
- Bankalar ve ipotek sağlayıcıları, bir borçlunun ödemelerinde temerrüde düşme olasılığını tahmin etmek için geçmiş verileri kullanır.
- Acil servis triyajı, hasta bakımına öncelik vermek için (yaş, cinsiyet, semptomlar vb. faktörlere dayalı olarak) karar ağaçlarını kullanabilir.
- Sizi ihtiyacınız olan sonuca yönlendiren otomatik telefon sistemleri,

3.7.5. Karar Ağaçları ile EEG Kullanımı

Yapılan bir çalışmadaki kullanım şekline örnek vermek gerekirse;

Farklı epileptik EEG veri setleri için karar ağaçları ve karar kuralları yöntemleri ile elde edilen tanı performans değerleri %95.3 ile %99.7 arasında değişim göstermiştir. (Kaya, Ertuğrul, Tekin, 2012)

3.8. ELM ve K-NN Algoritmalarının Literatündeki Önemi

ELM ve KNN Algoritması literatür taraması sonucunda çeşitli alanlarda çok fazla kullanımı görülmektedir.

Bu bölümde örnek verilen çalışmalar 2021-2022 yılında yapılan çalışmalardan sadece küçük bir kısmıdır. Bu çalışmalar gibi çok daha fazlası temel algoritma olarak ELM ve KNN algoritmasını baz almıştır. Kendilerine göre kimisi farklı bir yaklaşım sunarken kimisi de başka bir algoritma ile birleştirmiştir.

Literatür taramasından ELM için aşağıdakileri örnek olarak verebiliriz.

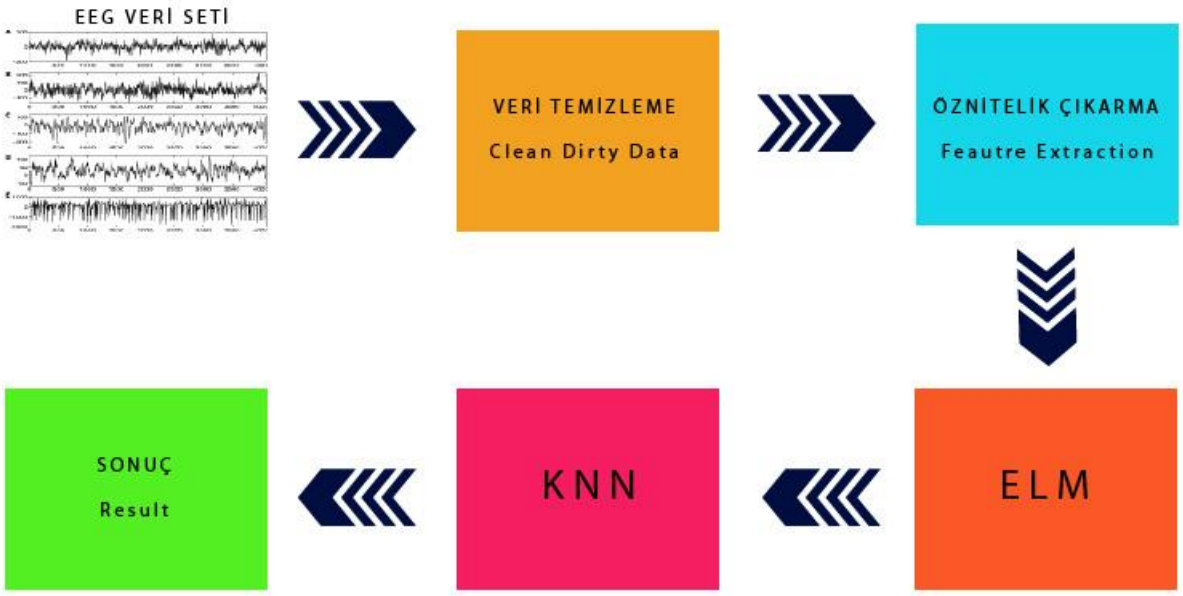
1- Bütünleme sınavına girecek öğrenci sayısının tahmini (Kıran, Sıramkaya, Eşme, 2022)
2- Diyabet Hastalıkları için tanı sistemi (Doğantekin, Bal, 2022)
3- Akciğer Kanser tespiti (Arı, Alçin, Şengür, 2022)

KNN için ise;

1- Öksürük Sesi Kayıtlarından COVID-19 Tespiti (%99,39) (Demircan, 2022)
2- Öğrencilerin Dersteki Niteliklerinin Sınıflandırılması (%97,66) (Güvenç, Sakal, Çetin, Özkaraca , 2022)
3- Diyabet Riski Tahmini (Özsezer, Mermer, 2022)

BÖLÜM 4

YENİ ALGORİTMA ÖNERİSİ



Şekil 4. 1. Yeni Algoritma Süreci

Bu bölümde önerilecek süreç için gösterim yukarıdaki akıştadır (Şekil 4.1).
Algoritma Adımları:

1. Veri Temizleme (Clean Dirty Data)
2. Öznitelik Çıkarma (Feature Extraction)
3. ELM
4. KNN
5. Sonuç (Result)

4.1 Veri Temizlemenin Önemi

EEG ölçümleri esnasında cihazlar çok fazla gürültüye maruz kalabilmektedir. Fiziksel aktiviteler ve çevresel etkenler bu gürültülerin büyük kısmını oluşturmaktadır. Sağlıklı ve başarılı sınıflandırma için elimizdeki veri setinin temizlenmesi ciddi anlamda önemlidir.

Independent Component Analysis (ICA) algoritmasının gücüyle veri setindeki gürültü tanımına uyan verilerin çıkarılması bu adımda uygulanabilir. Ölçüm esnasında fiziksel hareketler, dış etkenlere karşı beynin tepkisi gibi normal halinin dışına çıkan beyin sinyalleri gürültü niteliği taşıyabilmektedir. ICA algoritmalarında genellikle önişlem basamağı olarak merkezleme ve ağartma gibi methodlar ile problemin karmaşıklığını basitleştirmek veya boyut indirgemek için kullanılmaktadırlar.

Temizleme performansını etkileyen faktörler arasında eşik değerinin doğru belirlenmesi böylece gürültü veya orijinal sinyal bileşenlerinin hatasız ayrılması yer almaktadır. Belirlenecek eşik değeri kritik düzeyde önemlidir. Diğer adımlara iletilecek veri bu adımda belirlenmektedir. Önemli olan ilgili veri setinde temizlenecek verilerin doğruluğudur.

Bu aşamada ICA analizi kullanılarak gürültü tespiti yapılır ve eliminasyon işlemleri gerçekleştirilebilir.

4.2 Öznitelik Çıkartma

Öznitelik çıkartma ile ilgili birden fazla metot bulunmaktadır.

Literatür taraması sonucunda zaman ve frekans verileri ile çalışan Ayrık Dalgacık Dönüşümü metodu kullanılabilir. Dalgacık dönüşümünde seviye-seviye sınıflandırılmış frekanslar bulunmaktadır. Literatür de sık kullanılan çeşidi ise 4.seviyeye kadar ayrım yapılmaktadır.

Diğer bir öznitelik çıkartma metodu ise Yerel İkili Örüntü'dür. Bu metot her bir pikselin 8 komşusuna göre bir değer ortaya çıkarır. Bu değer hesaplandıktan sonra pikseldeki değerlerle kıyaslanarak küçük ve büyüklüklerine göre 1 ve 0 değerlerine dönüştürülür. Bu metot bu şekilde temizlenen veri setine uygulanabilir.

Mevcut kanalların her birisi için 256 özellik elde edilebilir. Tek bir kanal için bu sayının fazla olmasından dolayı veri boyutu düşürülmesi uygundur. Bunun içinde Yerel İkili Örüntü veri boyutunun düşürülmesi faydalı olabilir. Literatürde bu metot için sinyal verisi 8 bitlik alınıp 1 ve 0 değerlerine dönüştürülmüştür. Boyutu düşen veride de doğruluk oranının arttığı gözlemlenmiştir.

Diğer bir öznitelik çıkarma metodu ise Diferansiyel evrim algoritmasıdır. Son zamanlarda kullanımı artan bu algoritma temelinde Destek Vektör Makineleri (DVM) ile kullanıldığı için sınıflandırma da en yüksek performanslı öznitelikleri tespit etmede oldukça başarılı olacaktır. Bunu Literatür taraması sonuçları ortaya koymaktadır.

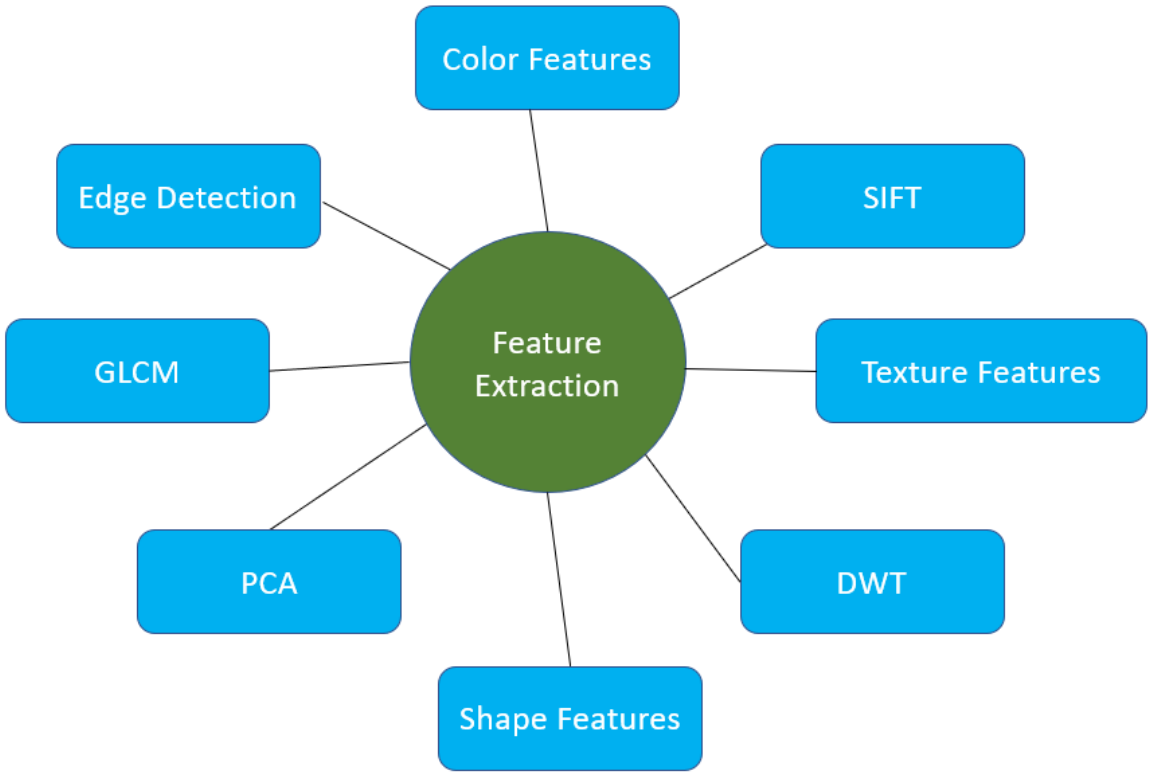
Öznitelik çıkarma da bu 3 metot denenerek çıkan öznitelikler yorumlanarak diğer aşama için karar verilmelidir. Literatür taraması yapıldığında Diferansiyel evrim algoritması ön plana çıkmaktadır.

Öznitelikler belirlenirken temel model olarak Diferansiyel evrim algoritması ile temizlenen veri seti işleme alınarak öznitelikler belirlenir. Ardından diğer öznitelik çıkartma işlemleri ile de bu işlem yapılır. Ortak öznitelikler kesin olarak öznitelik kümesine eklenir. Diferansiyel evrim algoritmasında elde edilip diğer algoritmalarda çıkmayan

öznitelik olursa bunlar fark yaratan öznitelik olma olasılığı olduğu için öznitelik kümesinin sonuna eklenir.

Diferansiyel evrim algoritması dışındaki algoritmalarda ortak olan öznitelikler de kümeye eklenir ve ardından en sonda olan fark yaratması muhtemel olan öznitelikler yorumlanarak karar verilmelidir.

Öznitelik çıkarma konusunda birden fazla metod bulunmaktadır. Bunlardan işe yarayacağı düşülen metodlar gözlemci tarafından belirlenmelidir (Şekil 4.1).



Şekil 4. 2. Öznitelik Çıkarma Metodları

4.3 ELM İncelemesi

Literatür de EEG ile yapılmış bir araştırmayı inceleyelim.

İncelenecek araştırma olarak "Epileptic EEG classification based on extreme learning machine and nonlinear features" isimli araştırma ele alanacaktır (Yuan vd., 2011).

4.3.1 Araştırma Hakkında

İncelenen araştırma EEG verilerinin ELM yaklaşımı ile incelenmesi üzerine bir yol göstermekte ve tatmin edici bir başarı oranı sağladığını ortaya sunmaktadır (Şekil 4.3).

The procedure of the ELM can be described as follows (Huang et al., 2004; Liang et al., 2006):

Suppose that there are N training samples $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{N}$, hidden layer activation function $g(x)$, and hidden neuron number M, Step 1: Input weights $\mathbf{w}_i$ and hidden layer biases b_i are randomly evaluated.

Step 2: Hidden layer output matrix $\mathbf{H}$ is computed.

Step 3: Output weight $\hat{\beta}$ is computed according to $\hat{\beta} = \mathbf{H}^{+}\mathbf{T}$.

Şekil 4. 3. incelenen ELM Örneğinde Kullanılan Adımlar

4.3.2 Veri Seti Hakkında

500 Kişiye ait 4097 veri noktasına ait ortalama 23.6 sn'lik EEG ölçümüne ait veriler ile yapılan araştırma ele alınacaktır.

Her bir 4097 veri noktasını 23 parçaya ayrılıp karıştırılmış. 1sn veride 178 veri noktası (sütun) bulunmaktadır. 23 x 500 = 11500 adet bilgi(satır) vardır (Çizelge 4.1).

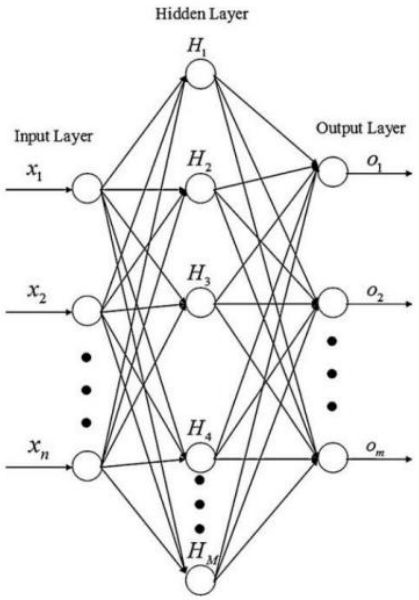
Tüm deneyleri, 3.00 GHz Intel Pentium işlemcide çalışan MATLAB7.0 ortamında gerçekleştirdiler.

Çizelge 4. 1. Bonn Üniversitesi Örnek EEG Verileri

Veri 1	Veri 2	Veri 3	Veri 4	Veri 5	Veri 6	Veri 7	Veri 8	Veri 9	Veri 10	Veri 11	Veri 12	Veri 13	Veri 14	Veri 15	Veri 16	Veri 17	Veri 18	Veri 19
12	-56	-37	-31	14	-87	-2	-31	8	-41	14	-64	10	31	30	34	-62	-18	0
22	-50	-22	-43	26	-89	20	-16	17	-38	25	-72	-6	25	19	34	-55	-9	2
35	-64	-17	-39	32	-73	42	10	29	-31	31	-52	-14	17	-3	31	-44	9	-1
45	-91	-24	-39	25	-69	48	28	46	-25	36	-24	-14	7	-14	4	-42	33	14
69	-135	-31	-9	16	-51	27	31	50	-32	33	4	-9	4	-12	-15	-33	52	36
74	-140	-20	-5	8	-48	11	8	44	-27	33	21	-7	4	-5	-41	-31	69	60
79	-134	-5	18	8	-51	9	-21	28	-29	27	18	1	4	10	-53	-24	56	80
78	-114	14	7	12	-47	48	-50	19	-20	27	-16	4	6	20	-63	-14	57	84
66	-115	31	-12	11	-30	82	-67	17	1	21	-53	0	-2	19	-49	-5	59	86
43	-126	31	-42	19	-6	83	-59	15	3	24	-64	-8	7	1	-25	9	57	78
33	-138	20	-53	23	12	50	-44	32	-9	32	-63	-21	14	-6	-2	17	46	77
36	-143	2	-62	24	7	12	-34	43	-29	39	-60	-15	26	-16	11	18	39	79
34	-126	0	-31	31	-9	-17	-35	36	-43	26	-61	-7	31	-18	-1	10	50	78
38	-91	2	-19	38	-30	-16	-29	14	-50	0	-48	17	38	-15	-35	-9	59	71
36	-57	12	-14	35	-42	-1	-22	8	-50	-11	-34	20	43	-4	-53	-27	66	63
28	-62	23	5	41	-27	11	-13	8	-43	-13	-18	15	35	-7	-54	-28	85	42
6	-91	20	-2	44	-8	-1	-6	16	-32	-3	5	11	30	5	-52	-30	78	10
-19	-125	18	-31	44	-3	-21	3	24	-22	6	12	21	28	10	-48	-37	69	-16
-38	-153	17	-68	26	-14	-28	-2	30	-1	16	4	33	40	0	-45	-27	73	-25
-45	-154	34	-108	11	-20	-7	-13	27	11	14	-7	49	44	-11	-38	-17	84	-30
-35	-147	43	-92	-4	-29	36	-15	33	0	28	-15	63	64	-23	-53	-2	95	-46
-11	-136	47	-76	-25	-10	86	-12	29	-22	25	-31	68	72	-42	-62	11	88	-54
6	-143	30	-41	-32	13	110	4	12	-41	1	-48	63	72	-52	-80	12	57	-66
7	-147	5	-41	-27	31	97	12	-4	-43	-15	-48	51	64	-62	-70	7	27	-75
7	-153	-28	15	-32	26	67	17	-17	-35	-23	-59	43	60	-52	-44	0	0	-74
-2	-171	-27	59	-31	6	36	8	-31	-14	-28	-74	26	55	-42	-26	-5	9	-69
-3	-187	-19	99	-23	-25	23	-8	-38	2	-31	-92	7	59	-44	-13	-7	26	-60
5	-175	-24	114	-19	-52	31	-5	-53	7	-36	-111	-10	73	-43	-12	-12	43	-60
19	-141	-27	75	3	-67	53	0	-42	-1	-20	-119	-27	81	-43	5	-7	25	-57
26	-96	-35	25	24	-56	74	13	-30	-12	1	-121	-39	85	-47	22	5	-1	-57
25	-47	-40	-13	44	-40	74	32	-26	-31	25	-99	-48	76	-44	12	2	-35	-57
41	-40	-42	-2	31	-12	63	52	-15	-49	45	-96	-43	73	-60	-15	2	-39	-53
46	-55	-46	16	11	4	51	57	-7	-46	57	-94	-39	63	-80	-54	-7	-20	-43
52	-74	-41	33	16	21	39	63	-4	-40	66	-78	-27	46	-73	-73	-14	9	-37

4.3.3 ELM Algoritmasının Kullanma Nedenleri

ELM yaklaşımının eğitim sürecinin yavaş olmaması ve hata payının düşük olması model eğitiminde ön plana çıkmaktadır. Öğrenme algoritmalarına göre daha hızlı öğrenme kapasitesine sahiptir. İncelenen araştırma da literatüre göre ELM algoritmasının sınıflandırma gücünün diğer algoritmalara kıyasla daha iyi olduğu belirtilmiştir. Şekil 4.4'te kullanılan algoritma yapısı gösterilmektedir.



Şekil 4. 4. ELM de İncelenen Araştırma da Kullanılan Algoritma Yapısı

4.3.4 Sonuç

ELM'nin performans ve doğruluk olarak denenen diğer algoritmalara göre daha iyi olduğu gözlemlediler. ELM algoritması eğitimdeki hata payını en küçüğe indirir ve sınıflandırma başarısı olarak diğer algoritmalara göre doğruluk oranı yüksek bir başarı sergilediği ortaya konulmuştur.

4.4 KNN (K-En Yakın Komşu Algoritması) İncelemesi

Literatür de KNN ile yapılmış bir araştırmayı inceleyelim.

İncelenecek araştırma olarak "K-En Yakın Komsu (K-Nn) Algoritması ile Hisse Senedi Fiyatlarının Tahmin Edilmesi: BIST'den Örnek Bir Uygulama" isimli araştırma ele alanacaktır (İlarslan, 2016).

4.4.1 Araştırma Hakkında

İncelenen araştırma BIST verilerinin KNN yaklaşımı ile incelenmesi üzerine bir yol göstermekte ve tatmin edici bir başarı oranı sağladığını ortaya sunmaktadır. İncelenen konu çok yoğun olarak kullanılan yatırım işlemi yapan bir çok kişiye hitap etmektedir. Sayısal verilerin olması tahmin işleminin başarılı olmasında büyük payı vardır.

4.4.2 Veri Seti Hakkında

BIST'in 2014 yılına ait veriler ile yapılan araştırma ele alınacaktır.

İlgili tahmin sistemi 2015 yılının ilk çeyreğinde tahmin edilecek hisse senedinin değerleri üzerinde test edilecektir.

2014 Yılının Son 1 aylık örnek verileri aşağıdaki gibidir. Tüm yıla ait verileri hisse senedi platformlarında yer almaktadır (investing, 2022).

Kullanılan verilere örnek olarak aşağıdaki gibi gösterilebilir (Çizelge 4.2).

Çizelge 4. 2. AFYON Hisse Senedi Verileri

Tarih	Açılış	Yüksek	Düşük	Hac.	Fark %
31.12.2014	2,3	2,32	2,22	35,50M	-0,88%
30.12.2014	2,21	2,3	2,18	74,03M	3,64%
29.12.2014	2,05	2,22	1,96	114,61M	8,37%
26.12.2014	2,11	2,24	1,95	91,11M	-2,87%
25.12.2014	1,96	2,11	1,94	67,08M	7,73%
24.12.2014	1,84	1,98	1,79	59,04M	6,01%
23.12.2014	1,9	1,9	1,83	21,80M	-2,14%
22.12.2014	1,89	1,91	1,78	36,90M	0,54%
19.12.2014	1,84	1,94	1,84	61,58M	2,20%
18.12.2014	1,64	1,85	1,64	60,64M	12,35%
17.12.2014	1,6	1,65	1,55	34,91M	1,89%
16.12.2014	1,65	1,69	1,55	34,87M	-3,64%
15.12.2014	1,68	1,71	1,58	51,93M	-1,79%
12.12.2014	1,79	1,8	1,68	43,00M	-6,15%
11.12.2014	1,78	1,83	1,68	70,48M	1,70%
10.12.2014	1,7	1,78	1,68	54,97M	6,67%
9.12.2014	1,63	1,74	1,55	108,52M	1,85%
8.12.2014	1,4	1,63	1,4	78,15M	16,55%
5.12.2014	1,35	1,45	1,25	116,99M	2,96%
4.12.2014	1,61	1,67	1,35	86,65M	-16,15%
3.12.2014	1,42	1,61	1,42	65,21M	18,38%
2.12.2014	1,19	1,36	1,19	47,68M	16,24%
1.12.2014	0,99	1,17	0,99	74,21M	18,18%
28.11.2014	0,9	1,02	0,9	64,84M	10,00%
27.11.2014	0,84	0,92	0,84	71,52M	12,50%
26.11.2014	0,72	0,8	0,72	36,66M	11,11%
25.11.2014	0,74	0,74	0,72	14,40M	-2,70%
24.11.2014	0,71	0,75	0,68	56,52M	1,37%
21.11.2014	0,74	0,74	0,73	1,78M	0,00%
20.11.2014	0,74	0,74	0,73	2,89M	0,00%
19.11.2014	0,74	0,74	0,73	4,37M	-1,35%
18.11.2014	0,74	0,74	0,73	7,23M	1,37%
17.11.2014	0,73	0,75	0,73	24,13M	0,00%
14.11.2014	0,73	0,73	0,73	2,07M	0,00%
13.11.2014	0,73	0,74	0,72	2,28M	0,00%
12.11.2014	0,72	0,73	0,72	3,81M	2,82%
11.11.2014	0,71	0,72	0,71	4,28M	0,00%
10.11.2014	0,71	0,72	0,71	1,30M	0,00%
7.11.2014	0,71	0,72	0,71	1,72M	0,00%
6.11.2014	0,72	0,72	0,71	2,62M	-1,39%
5.11.2014	0,74	0,74	0,72	6,77M	-2,70%
4.11.2014	0,73	0,74	0,73	8,45M	1,37%
3.11.2014	0,74	0,75	0,73	6,96M	-1,35%
31.10.2014	0,77	0,77	0,74	10,24M	-1,33%
30.10.2014	0,71	0,76	0,7	21,63M	5,63%

4.4.3 k Değerinin Belirlenmesi

Algoritma da kullanılacka k değerinin tespiti için 1'den 10'a kadar k değerleri ile test verisi üzerinde deneme yapılmıştır. K=1 test verisi için örnek çıktılar aşağıdaki sonucu çıkarılmıştır (Çizelge 4.3).

Çizelge 4. 3. k=1 için Test Sonuçları

1. Test kümesi hata oranı = $e_i^2 = (48,8 - 50,3)^2 = 2,25$
2. Test kümesi hata oranı = $e_i^2 = (48,8 - 48,85)^2 = 0,0025$
3. Test kümesi hata oranı = $e_i^2 = (54,35 - 55,2)^2 = 0,7225$
4. Test kümesi hata oranı = $e_i^2 = (54,5 - 53,85)^2 = 0,4225$
5. Test kümesi hata oranı = $e_i^2 = (52,25 - 52,1)^2 = 0,0225$
6. Test kümesi hata oranı = $e_i^2 = (53,6 - 53)^2 = 0,36$
7. Test kümesi hata oranı = $e_i^2 = (52,8 - 51,8)^2 = 1$
8. Test kümesi hata oranı = $e_i^2 = (48,95 - 50,1)^2 = 1,3225$
9. Test kümesi hata oranı = $e_i^2 = (53,65 - 58,4)^2 = 22,5625$
10. Test kümesi hata oranı = $e_i^2 = (165 - 153,5)^2 = 132,25$

Testler sonucunda k=6 değeri için en küçük hata oranı tespit edilmiştir. Bu yüzden sınıflandırma işleminde k=6 olarak ele alınmıştır.

4.4.4 KNN Algoritmasının Kullanma Nedenleri

Literatürdeki yapılan araştırmaları örneklendirerek KNN algoritmasının oldukça başarılı olduğu gözükmektedir. Doğrusal olmayan dinamikleri modellemede olan başarısı da özellikle bu algoritmayı seçmelerinde büyük etkenlerden birisidir.

4.4.5 Sonuç

2015 yılı verilerindeki tahminleri oldukça başarılı çıkmıştır. KNN başarısı 97% oranında doğruluğa ulaşmıştır.

Sonuç itibariyle sayısal verilerdeki çıktılar büyük miktar öngörülebilir olduğu bilindiğinden bu öngörü de KNN'in yüksek yeteneği ile birleştiğinde başarı kaçınılmaz olmuştur.

BÖLÜM 5

SONUÇ

Ham (RAW) verilerin gürültülerden arındırılmasıyla elde edilen veri seti işlemlerin başlangıcıdır. Bu veri seti incelenmesi sonucunda çıkarılan öznitelikler ile ELM ve KNN algoritmaları sınıflandırılma işlemi yapılması optimum bir sonuç elde edilmesi beklenilmektedir. ELM ve KNN sınıflandırmaları da kendilerini kanıtlamış ve sınıflandırma işlemlerinde yüksek oranı veren algoritmalardır. İki önemli algoritmanın güçlerinin birleştirilmesiyle birlikte hata oranı çok düşük seviyelere düşmesi ön görülmektedir. Her adım bir sonraki adımı ciddi derecede etkilediğinden dolayı yapılan işlemler farklı şekilde denenerek karşılaştırma yapılabilir. Algoritma içerisinde ufak dokunuşlar yapılarak veri seti ve ortama göre revizeler yapılabilir.

KNN algoritmasındaki uygun k değerinin bulunması için temizlenen veri setinin yalnız KNN ile çalıştırılması uygundur. Optimum k değeri bu denemeler sonucunda bulunduktan sonra önerilen algoritma baştan uca ilgili veri seti ile çalıştırılabilir.

Algoritma aşamaları teorik açıdan ele alındığından başarılı olması sadece ön görülmektedir. EEG veri setinin oluşturulması, ölçümlerin yapılması ile önerilen algoritmanın ilgili veri seti üzerinde çalıştırılması ve bu işlemin farklı birkaç şekilde denenmesi gerçek sonuçları gösterecektir.

Yapılacak olan çalışmalara katkı sağlaması amacıyla önerilen sürecin adımlarının detaylarına inilip yapılandırmalar yapılabilir. En önemli adımlardan biri olan özellik çıkarma adıma öncelikle ele alınabilir.

KAYNAKLAR

Arı, B., Alçin, Ö. F. & Şengür, A. (2022). Akciğer Kanser Tipi Tespitinde Etkili Bir Görüntü Çoğullama Tekniği. Computer Science, IDAP-2022 : International Artificial Intelligence and Data Processing Symposium , 12-20 . DOI: 10.53070/bbd.1173074

Aydemir, O., & Kayikcioglu, T. (2014). Decision tree structure based classification of EEG signals recorded during two dimensional cursor movement imagery. *Journal of neuroscience methods*, 229, 68-75.

Bablani, A., Edla, D.R., & Dodia, S. (2018). Classification of EEG data using k-nearest neighbor approach for concealed information test. *Procedia computer science*, 143, 242-249.

Çağlıyan, B., & Köse, U. (2021). Epilepsi eeg verilerinin makine öğrenmesi teknikleriyle sınıflandırılması. *Avrupa bilim ve teknoloji dergisi*, 23 , 163-172

Demircan, S. (2022). Öksürük sesi kayıtlarından spektral özellikler ile otomatik COVID-19 tespiti. *Avrupa bilim ve teknoloji dergisi*, 492-495

Doğantekin, A., & Bal, C. (2022). Diyabet hastalıkları için ga-dçf-uöm tabanlı uzman tanı sistemi. *Fırat üniversitesi mühendislik bilimleri dergisi*, 34, 473-484

Duan, L., Ge, H., ma, W., & Miao, J. (2015). EEG feature selection method based on decision tree. *Bio-medical materials and engineering*. 26. S1019-S1025

Dursun, İ., & Akın, M. (2021). Eeg verilerine uygulanan senkrosıkıştırma yönteminin migren teşhisinde kullanılması. *Dicle üniversitesi mühendislik fakültesi mühendislik dergisi*, 12, 767-774

Erdoğan, F. (2015, Temmuz 21). 22 Temmuz Dünya Beyin Günü: Epilepsi; "Bizim Beynimiz – Bizim Geleceğimiz". Erişim Tarihi 15 Ekim 2022 https://www.noroloji.org.tr/haber/106/22-temmuz-dunya-beyin-gunu-epilepsi-bizim-beynimiz-bizim-gelecegimiz

Gül, A., & Bayraktar, B. (2021). *Electroencephalography*. (Yayımlanmamış Bitirme Projesi) Trakya Üniversitesi/Bilgisayar Mühendisliği Bölümü Bitirme Projesi, Edirne

Güvenç, E., Sakal, M., Çetin, G., & Özkaraca, O. (2022). Öğrencilerin dersteki niteliklerinin makine öğrenmesi teknikleri kullanılarak sınıflandırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 10, 1359-1371

Guang-Bin, H., Qin-Yu, Z., & Chee-Kheong, S., Extreme learning machine: Theory and applications. *Neurocomputing*, 70, 489-501

Investing (2022, Ekim 15). Afyon Çimento Sanayi T.A.Ş (AFYON). https://tr.investing.com/equities/afyon-cimento-historical-data

Işıkhan, Elif. (2022, Ekim 5). Multi Layer Perceptron (MLP) Nedir? https://isikhanelif.medium.com/multi-layer-perceptron-mlp-nedir-4758285a7f15

İlkarslan, K. (2016). K-en yakın komşu (k-nn) algoritması ile hisse senedi fiyatlarının tahmin edilmesi BIST'den örnek bir uygulama. *The Journal of Academic Social Science*, 30, 375-392

Kaya, Y., & Tekin, R. (2012). Epileptik nöbetlerin tespiti için aşırı öğrenme makinesi tabanlı uzman bir sistem. *Bilişim Teknolojileri Dergisi*, 5, 33-40

Kaya, Y., Ertuğrul, Ö. F., & Tekin, R. (2012). Epileptik eeg işaretlerinin sınıflandırılmasında karar kuralları ve karar ağaçlarının kullanılması. *Batman Üniversitesi Yaşam Bilimleri Dergisi*, 1, 403-413

Kıran, M. S., Sıramkaya, E., & Eşme, E. (2021). Bütünleme sınavına girecek öğrenci sayısının tahmini için k-ELM yaklaşımı. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 37, 295-304

Novak, D., Al-Ani, T., Hamam, A., & Lhotska, L., (2004). Electroencephalogram processing using hidden markov models abstract. *5th eurosım congress on modelling and simulation*

Özkan, G. (2019). Videolara saklanmış 25. kare etkisinin eeg sinyalleri yardımı ile tespit edilmesi. *İskenderun Teknik Üniversitesi / Mühendislik ve Fen Bilimleri Enstitüsü*, Hatay

Özsezer, G. & Mermer, G. (2022). Diabetes risk prediction with machine learning Models. *Artificial Intelligence Theory and Applications,* 2, 1-9.

Öztürk, Ş. (2022, Şubat 10). Dünya Epilepsi Günü: Türk Nöroloji Derneği Basın Açıklaması. Erişim Tarihi 5 Ekim 2022 https://www.noroloji.org.tr/haber/353/dunya-epilepsi-gunu-turk-noroloji-dernegi-basin-aciklamasi

Reçber, D., & Akşahin, M. F. (2019). Epilepsy classification in ıctal stage from eeg signals for children. *2019 Medical Technologies Congress (TIPTEKNO)* 1-4

Sezgin, N., Tağluk, M. E., & Tekin, R. (2012). Separation of EEG signals by using independent component analysis. *2012 20th Signal Processing and Communications Applications Conference (SIU),* 1-4

Sivakumar, S. (2011). The Wave- The characteristics of an EEG, Erişim tarihi 10 Ekim 2022, https://www.firstclassmed.com/articles/2017/eeg-waves

Sohrabian, B. (2013). *Bağımsız Bileşenler Analizi ile Çok Değişkenli Jeoistatistiksel Kestirim*. (Yayımlanmamış doktora tezi). Hacettepe Üniversitesi/Fen Bilimleri Enstitüsü, Ankara

Toprak, I. B., Çağlar M. F., & Merdan, M. (2007, Eylül 4). Automatic recognition of epilepsy from eeg using artificial neural network and discrete wavelet transform. *2007 IEEE 15th Signal Processing and Communications Applications,* 1-4

Tuncer, E. & Doğru Bolat, E. (2022). Destek vektör makinaları ile eeg sinyallerinden epileptik nöbet sınıflandırması. *Politeknik Dergisi,* 25, 239-249

Ulgen, E. K. (2017, Kasım 12), Makine Öğrenimi Bölüm-5 (Karar Ağaçları), Erişim tarihi 10 Eylül 2022, https://medium.com/@k.ulgen90/makine-%C3%B6%C4%9Frenimi-b%C3%B6l%C3%BCm-5-karar-a%C4%9Fa%C3%A7lar%C4%B1-c90bd7593010

Wang, M., Abdelfattah, S., Moustafa, N., & Hu, J. (2018, Haziran 20) Deep gaussian mixture-hidden markov model for classification of eeg signals. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 278-287

World Health Organization. (2022, Şubat 9), Epilepsy, Erişim tarihi 10 Eylül 2022, https://www.who.int/news-room/fact-sheets/detail/epilepsy

Yuan, Q., Zhou, W., Li, S., & Cai, D. (2011, Eylül). Epileptic EEG classification based on extreme learning machine and nonlinear features, *Epilepsy Research,* 96, 29-38

Yanar, H., & Mishchenko, Y. (2016). Gelişmiş EEG tabanlı beyin bilgisayar arayüzleri için elektroensefalografik beyin aktivitesinin saklı Markov modeli. *Mersin üniversitesi,* 78-86

ÖZGEÇMİŞ

Sivas'ta doğdum. 1997-98 yıllarında Çorlu/Tekirdağ şehrine taşındım. Çorlu Şahinler İlköğretim okulunda okudum. Lise eğitimi için de Mimar Sinan Anadolu Lisesi'ne gittim. 2012 yılında Trakya Üniversitesi'nde Bilgisayar Mühendisliği bölümünde lisans eğitimime başladım. 2016 yılında mezuniyetimle beraber yazılım sektöründe çalışmaya başladım. 2018 yılında uzun dönem askerliğim de Milli Savunma Üniversitesi'nin İzmir şehrindeki Meslek Yüksek Okulu'nda 1 yıl boyunca yazılım eğitimi verdim. Askerlikten döndükten sonra 2019 yılında Trakya Üniversitesi'nde Bilgisayar Mühendisliği Anabilim Dalı'nda yüksek lisans eğitimime başladım.