

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



EEG İŞARETLERİ KULLANILARAK BAĞIMLILIĞA
YATKINLIĞIN MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE
DEĞERLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

VEYSEL YARĞI

BOLU, TEMMUZ - 2020

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI



EEG İŞARETLERİ KULLANILARAK BAĞIMLILIĞA
YATKINLIĞIN MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE
DEĞERLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

VEYSEL YARĞI

BOLU, TEMMUZ - 2020

KABUL VE ONAY SAYFASI

Veysel YARĞI tarafından hazırlanan EEG İŞARETLERİ KULLANILARAK BAĞIMLILIĞA YATKINLIĞIN MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE DEĞERLENDİRİLMESİ adlı tez çalışması Elektrik Elektronik Mühendisliği Anabilim Dalı'nda 17.08.2020 tarihinde savunularak **Bolu Abant İzzet Baysal Üniversitesi Lisansüstü Eğitim Enstitüsü** Yüksek Lisans Tezi olarak kabul edilmiştir.

Jüri Üyeleri

Danışman
Doç. Dr. Seda POSTALCIOĞLU
Bolu Abant İzzet Baysal Üniversitesi

İmza

.....

Üye
Prof. Dr. Kemal POLAT
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Abdullah Talha KABAKUŞ
Düzce Üniversitesi

.....

Prof. Dr. Osman GÖRÜR.....

Lisansüstü Eğitim Enstitüsü Müdürü



Aileme,

ETİK BEYAN

Bolu Abant İzzet Baysal Üniversitesi, Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Veysel YARĞI

ÖZET

**EEG İŞARETLERİ KULLANILARAK BAĞIMLILIĞA YATKINLIĞIN
MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE DEĞERLENDİRİLMESİ
YÜKSEK LİSANS TEZİ
VEYSEL YARĞI
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI
(TEZ DANIŞMANI: DOÇ.DR. SEDA POSTALCIOĞLU)**

BOLU, TEMMUZ - 2020

Alkol bağımlılığının Elektroensefalografi (EEG) sinyalleri ile teşhisi, hem kişisel açıdan hem de toplum açısından önemli bir konudur. Günümüzde birçok insan bu bağımlılıktan etkilenmektedir. Başta beyin, kalp ve bağışıklık sistemi olmak üzere fizyolojik etkileri olduğu gibi, psikolojik etkileri de söz konusudur. Bu etkileri gözlemleyebilmek için EEG sinyalleri etkin bir şekilde kullanılmaktadır. Bu çalışmada, alkolizme yatkınlığın EEG sinyalleri kullanılarak teşhisi yapılmıştır. Veritabanı aracılığı ile elde edilen EEG sinyal üzerinde öncelikle veri analizi yapılmıştır. Özyinelemeli öznitelik seçimi gerçekleştirilmiştir. Sınıflandırma için Çok Katmanlı Yapay Sinir Ağları (ÇKYSA), Evrimsel Sinir Ağları (ESA), XGBoost Algoritması (XGBA), Rassal Orman Algoritması (ROA), K-En Yakın Komşu Algoritması (K-EKA) kullanılmıştır. Python ortamında çalışılmıştır. Sınıflandırma başarımları için doğruluk, kesinlik, duyarlılık ve F1 Skor kullanılmıştır. Algoritmalar çalışma süreleri açısından karşılaştırılmıştır. Sınıflandırma başarımları arasında ÇKYSA ve ESA en iyi sonuçları vermiştir. Algoritmaların çalışma süreleri açısından bakıldığında XGBA en hızlı çalışan algoritma olduğu görülmüştür.

ANAHTAR KELİMELELER: Elektroensefalografi (EEG), Yapay Sinir Ağları, XGBoost, K-En Yakın Komşu, Rassal Orman Algoritması

ABSTRACT

EVALUATION OF PREDISPOSITION TO ADDICTION WITH MACHINE LEARNING TECHNIQUES USING EEG SIGNALS

MSC THESIS

VEYSEL YARĞI

BOLU ABANT İZZET BAYSAL UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

**DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING
(SUPERVISOR: ASSOC.PROF.DR. SEDA POSTALCIOĞLU)**

BOLU, JULY 2020

Diagnosis of alcohol dependence with Electroencephalography (EEG) signals is an important issue both personally and society. Today, Many people are affected by this addiction. It has physiological effects, especially the brain, heart and immune system, as well as psychological effects. EEG signals are used effectively to observe these effects. In this study, genetic predisposition to alcoholism is diagnosed using EEG signals. Firstly, data analysis was performed on the EEG signal data obtained through the database. Recursive feature selection is used. For the classification, Multilayer Artificial Neural Networks (MLPNN), 1D-Convolutional Neural Networks (CNN), XGBoost Algorithm (XGBA), Random Forest Algorithm (RFA), K-Nearest Neighbor Algorithm (K-NN) are used. It has been studied in Python environment. Accuracy, precision, sensitivity and F1 Score are used for classification performance criteria. Algorithms are also compared in terms of working time. In terms of classification success, MLPNN and CNN gave the best results. In terms of running time of algorithms, XGBA is the fastest running algorithm.

KEYWORDS: Elektroensefalografi (EEG), Neural Network, XGBoost, K-Nearest Neighbour, Random Forest Algorithm

İÇİNDEKİLER

Sayfa

ÖZET.....	v
ABSTRACT	vi
İÇİNDEKİLER.....	vii
ŞEKİL LİSTESİ.....	ix
ÇİZELGE LİSTESİ.....	x
KISALTMA VE SEMBOLLER LİSTESİ.....	xi
TEŞEKKÜR.....	xiii
1. GİRİŞ	1
1.1 Elektroensefalografi (EEG) Sinyalleri.....	1
1.1.1 EEG Dalgaları.....	1
1.1.1.1 Delta Dalgaları.....	2
1.1.1.2 Teta Dalgaları.....	2
1.1.1.3 Alfa Dalgaları.....	2
1.1.1.4 Beta Dalgaları.....	2
1.1.1.5 Gama Dalgaları.....	3
1.1.2 EEG Sinyallerinin Kullanıldığı Alanlar	3
1.1.3 EEG Sinyallerinin Ölçülmesi	3
1.2 Problemin Tanımlanması ve Tezin Amacı.....	5
1.3 Tezin Yapısı	6
2. KAYNAK ARAŞTIRMASI.....	7
3. MATERYAL VE YÖNTEM	9
3.1 Kullanılan EEG Veri Yapısı	9
3.2 Özyinelemeli Öznitelik Seçimi	15
3.3 Standardizasyon ve Normalizasyon	16
3.4 Çok Katmanlı Yapay Sinir Ağları	17
3.4.1 Aktivasyon Fonksiyonları	19
3.4.2 Çok Katmanlı Yapay Sinir Ağının Yapısı.....	20
3.4.3 Yapay Sinir Ağında Öğrenme	21
3.5 Evrimsel Sinir Ağları	22
3.6 K-En Yakın Komşu Algoritması.....	24
3.7 Karar Ağaçları.....	25
3.7.1 XGBoost Algoritması.....	27
3.7.2 Rassal Orman Algoritması	28
3.8 Sınıflandırma Başarım Ölçütleri	32
3.8.1 Karışıklık Matrisi	32
3.8.2 Doğruluk.....	32
3.8.3 Kesinlik	33
3.8.4 Duyarlık.....	33

3.8.5 F1-Score	33
4. BULGULAR.....	34
4.1 Çok Katmanlı Yapay Sinir Ağı Sınıflandırma Başarım Değerlendirmesi.....	35
4.2 Evrişimsel Sinir Ağı Sınıflandırma Başarım Değerlendirmesi	37
4.3 XGBoost Sınıflandırma Başarım Değerlendirmesi	39
4.4 Rassal Orman Algoritması Sınıflandırma Başarım Değerlendirmesi	40
4.5 K-En Yakın Komşuluk Sınıflandırma Başarım Değerlendirmesi	41
4.6 Sınıflandırma Başarımlarının Karşılaştırılması.....	42
5. SONUÇ VE ÖNERİLER	45
6. KAYNAKLAR	47
7. ÖZGEÇMİŞ	53



ŞEKİL LİSTESİ

Sayfa

Şekil 1.1.	Uluslararası 10-20 sistemine göre 19 kanallı elektrot yerleşimi.....	4
Şekil 3.1.	Sensör konumları.	9
Şekil 3.2.	Eğitim veri setindeki alkol ve kontrol grubu dağılımı	10
Şekil 3.3.	Alkolik (a) gruba ait sensör değerlerinin korelasyon ilişkisi.....	11
Şekil 3.4.	Kontrol (c) grubuna ait sensör verilerinin korelasyon ilişkisi	12
Şekil 3.5.	İlk 16 sensör pozisyonu arasındaki korelasyon	13
Şekil 3.6.	16 ile 32 no'lu sensör pozisyonları arasındaki korelasyon.....	13
Şekil 3.7.	32 ile 48 arasındaki sensör özellikleri arasındaki korelasyon	14
Şekil 3.8.	48 ile 64 arasındaki sensör özellikleri arasındaki ilişki	14
Şekil 3.9.	Özellik sayısına göre çapraz doğrulama puanı	15
Şekil 3.10.	Biyolojik nöron.....	18
Şekil 3.11.	Temel bir YSA örneği.....	18
Şekil 3.12.	Sigmoid aktivasyon fonksiyonu.....	19
Şekil 3.13.	ReLU aktivasyon fonksiyonu	20
Şekil 3.14.	Çok katmanlı yapay sinir ağı yapısı.....	20
Şekil 3.15.	Genel ESA yapısı.....	23
Şekil 3.16.	Havuzlama işlemi.....	24
Şekil 3.17.	Karar ağacının genel yapısı	26
Şekil 3.18.	Rassal Orman Algoritmasının yapısı	30
Şekil 4.1.	Çalışmanın blok diyagramı.....	34
Şekil 4.2.	ÇKYSA için karışıklık matrisi.....	36
Şekil 4.3.	ÇKYSA için modelin doğruluk eğrisi	36
Şekil 4.4.	ÇKYSA için modelin kayıp eğrisi	37
Şekil 4.5.	ESA için karışıklık matrisi	38
Şekil 4.6.	ESA için modelin doğruluk eğrisi.....	38
Şekil 4.7.	ESA için modelin kayıp eğrisi	39
Şekil 4.8.	XGBoost için Karışıklık matrisi	40
Şekil 4.9.	ROA için karışıklık matrisi.....	41
Şekil 4.10.	K-EKA için karışıklık matrisi.....	42
Şekil 4.11.	Kesinlik başarımları sonuçları	42
Şekil 4.12.	Duyarlılık başarımları	43
Şekil 4.13.	F1 Skor başarımları	43
Şekil 4.14.	Doğruluk başarımları.....	44
Şekil 4.15.	Süre açısından başarımları.....	44

ÇİZELGE LİSTESİ

Sayfa

Çizelge 3.1. Veri seti özellikleri..	16
Çizelge 3.2. Yapay sinir ağları ve biyolojik sinir ağlarının karşılaştırılması.....	17
Çizelge 3.3. Karışıklık matrisi.....	32
Çizelge 4.1. Kullanılan kütüphane ve versiyonları	35
Çizelge 4.2. ÇKYSA için başarımlar sonuçları.....	35
Çizelge 4.3. ESA için başarımlar sonuçları	37
Çizelge 4.4. XGBoost için başarımlar sonuçları	39
Çizelge 4.5. Rassal Orman algoritması için başarımlar sonuçları.....	40
Çizelge 4.6. K-en yakın komşu algoritması için başarımlar sonuçları	41

KISALTMA VE SEMBOLLER LİSTESİ

$\hat{y}$	: Karar Ağacında Tahmin Edilen Değer
Σ	: Toplam Sembolü
B	: Yapay Sinir Ağında Bias Ağırlığı
C	: Merkez bölge (Central)
ÇKYSA	: Çok Katmanlı Yapay Sinir Ağları
d(x,y)	: Öklit Mesafesi
EEG	: Elektroensefalografi
ESA	: Evrişimsel Sinir Ağları
F	: Ön lob (Frontal)
f(x_i)	: L1 ve L2 Norm Tahmini Değer
GN	: Gerçek Negatif
GP	: Gerçek Pozitif
Hm	: Çok Katmanlı Yapay Sinir Ağlarında Gizli Katmanlar
Hz	: Hertz
In	: Çok Katmanlı Yapay Sinir Ağlarında Girdi Katmanı
K	: En Yakın Komşu Sayısı
K-EKA	: K-En Yakın Komşu Algoritması
O	: Arka bölge (Oksitipital)
OOB	: Out Of Bag
ÖÖS	: Özyinelemeli Öznitelik Seçimi
P	: Yan loblar (Pariyetal)
Q	: Karar Ağaçlarındaki Ağaçların Yapısı
ReLu	: Rectified Lineer Unit
ROA	: Rassal Orman Algoritması
T	: Karar Ağacındaki İterasyon Sayısı
T	: Şakak bölgesi (Temporal)
w_i	: Yapay Sinir Ağında Giriş Bilgilerinin Ağırlığı
w_q	: Karar Ağaçlarındaki Yaprak Düğümlerinin Ağırlığı
XGBA	: Extreme Gradient Boosting Algoritması
x_i	: Yapay Sinir Ağında Giriş Bilgisi

y_i	: Karar Ağacında Gerçekte Ölçülen Değer
Y_i	: L1 ve L2 Norm Hedef Değer
y_k	: Yapay Sinir Ağında Çıktı Bilgisi
YN	: Yanlış Negatif
YP	: Yanlış Pozitif
YSA	: Yapay Sinir Ağları
φ	: Yapay Sinir Ağında Aktivasyon Fonksiyonu



TEŞEKKÜR

Bu tezin oluşmasında, her türlü desteği veren çok kıymetli hocam Doç. Dr. Seda POSTALCIOĞLU'na, manevi telkinleriyle destek veren anne ve babama, Python programlama dilini bana sevdirmeye çalışan ve bu tezde kullanmam için beni gayrete getiren ve yardımlarını esirgemeyen değerli arkadaşım Ömer ULUTAŞ'a teşekkürlerimi sunarım.



1.GİRİŞ

1.1 Elektroensefalografi (EEG) Sinyalleri

Elektriksel sinyallerin varlığı, ilk 1875 yılında Richard Caton tarafından tavşan ve maymunların beyinleriyle ilgili yaptığı çalışmalar ve deneyler sonucu ortaya çıkmıştır. Hans Berger ise 1929 yılında savaşta ağır yaralanan insanların beyinin iç kısmına elektrotlar yerleştirip, galvanometre ile beyindeki elektriksel sinyallerin varlığını tespit etmiştir. İlerleyen yıllarda gelişen teknolojiyle beraber saçlı deri üzerinde de ölçümler yapılabilmektedir (İasemidis, 2003). EEG, beyindeki elektriksel aktiviteleri ölçen ve kayıt altına alan bir sinyal ölçme tekniğidir. Kafatasındaki belli başlı bölgeler esas alınmak suretiyle, saçlı deri üzerine yerleştirilen elektrotlar aracılığıyla bu sinyaller ölçülür. Bu teknik beyin bölgesine veya herhangi bir bölgeye elektrik aktarmadığı için zararsızdır.

1.1.1 EEG Dalgaları

Beyindeki psikolojik ve fiziksel aktivitelerin durumuna göre EEG dalgaları şekil, genlik ve frekans değerleri açısından farklılıklar gösterir. Sinyaller kaotiktir ve zaman serilerine bağlı olarak örneklendirilir. Beyindeki hücreler arası iyonik akış yoğunluğundan dolayı volt değerlerinde dalgalanmalar görülebilir (Schomer ve Silva, 2017). EEG sinyalleri; uyku sırasında, uyanıklık durumunda, gözler açık ve kapalı gibi farklı durumlar için ölçüldüğünde farklı genlik ve frekanslarda sonuçlar elde edilir. Periyodik olmayan bu ritmik sinyaller 5 farklı frekans bandı adı altında incelenebilir. Bu frekans bandları; Delta (0.2-3 Hz), Teta (3-8 Hz), Alfa (8-12 Hz), Beta (12-27 Hz), Gamma (27 Hz ve üzeri) olarak literatüre geçmiştir (Kruizinga,2016).

1.1.1.1 Delta Dalgaları

Walter tarafından 1936'da tespit edilmiştir. Beyin faaliyetlerinin en az seviyede olduğu, uyku veya anestezi gibi bilincin olmadığı koşullarda ortaya çıkan dalgalardır. Kafatasındaki frontal ve oksipital bölgelerinde daha sık görülür. Beyin dalgaları içinde 0.2 Hz-3 Hz bandında olup en düşük genlik ve frekansa sahip olan dalgadır (Tülay, 2009).

1.1.1.2 Teta Dalgaları

1944 yılında Doney ve Walter tarafından incelenmiştir. 3-8 Hz aralığında olan teta dalgası beyin aktivitelerinin düşük olduğu dalgınlık, derin huzur, rahavet, uykuya geçiş ve duygusal anlarda belirgin bir şekilde gözlemlenmiştir (Tülay, 2009).

1.1.1.3 Alfa Dalgaları

Berger tarafından 1929 yılında tespit edilmiştir. 8-12 Hz aralığında olup, uyanıkken odaklanmama halinde, rahatlama ve gevşeme anlarında belirgin olarak görülür. En iyi gözlemlendiği yer kafatasının frontal korteks bölgeleridir. Uyku durumunda sinyaller kaybolur. Beyin kanaması gibi durumlarda anormal olarak gözlemlenir. Konsantrasyon bozukluğu ve depresyon gibi durumların düzeltilmesinde alfa dalga boyuna sahip ses dosyaları dinlenilmesi tavsiye edilmiş ve başarı elde edilmiştir (Tülay, 2009; Akman, 2011).

1.1.1.4 Beta Dalgaları

12-27 Hz aralığında olup düşük genliğe sahiptirler. Beyin aktivitelerinin arttığı durumlarda gözlemlenir. Stresli ve gergin anlarda, korku ve heyecan durumlarında görülür. Beta dalgaları aktivitesi uyarıcı gıda veya maddeler alındığında artış gösterir (Akman, 2011).

1.1.1.5 Gama Dalgaları

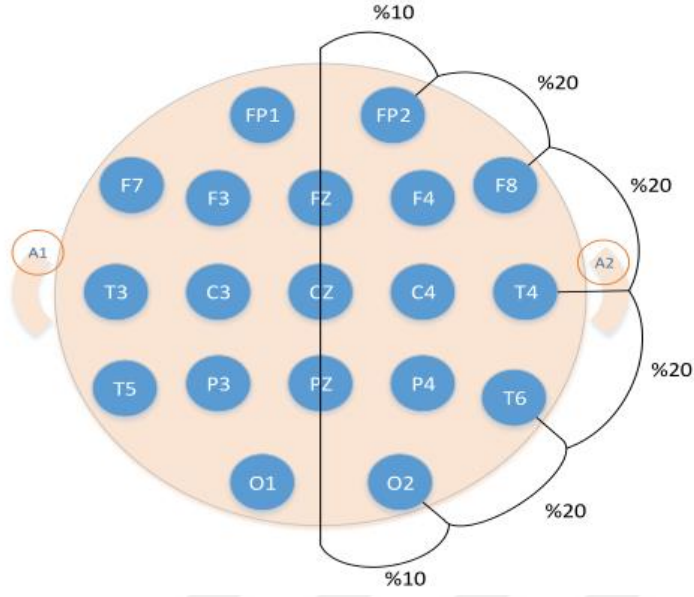
Jasper ve Adrew tarafından ilk olarak 1938 yılında incelenmiştir. Beyin dalgaları içerisinde 27 Hz ve üzeri frekansa sahip olan en hızlı dalgalardır. Konsantrasyon gerektiren faaliyetlerde ve beynin herhangi bir veriyi hızlı okuyup karar vermesi gerektiği durumlarda ortaya çıkar. Ayrıca uyku ile ilgili çalışmalarda da, uykunun özelliklerini ortaya çıkartmakta kullanılabilmektedir (Tülay, 2009; Akman, 2011).

1.1.2 EEG Sinyallerinin Kullanıldığı Alanlar

Beyin ile ilgili tüm rahatsızlıkların teşhis edilmesinde bu sinyallerden faydalanılır. Özellikle uyku ile ilgili fonksiyonların analizinde, beyin tümörü etkinliğinin tespitinde, epilepsi ile ilgili çalışmalarda; nöbet öncesi, sonrası ve nöbet anındaki sinyallerin ölçülüp değerlendirilmesinde kullanılır. Ayrıca odaklanma, algı bozukluğu ve duygu analizi gibi bilişsel aktivitelerin analizinde bu sinyallerden faydalanılabilir. Bu tez çalışmasında ise alkole olan bağımlılık ile ilgili EEG sinyalleri analiz edilerek, farklı makine öğrenmesi algoritmaları ile sınıflandırma yapılmıştır.

1.1.3 EEG Sinyallerinin Ölçülmesi

EEG sinyallerinin ölçülmesi elektrotların kafatası yüzeyine yerleştirilmesi aracılığıyla yapılır. Bu işlem uygulanırken Uluslararası EEG Federasyon Birliği tarafından belirlenen bir standart kullanılır. Uluslararası 10-20 sistemine göre en az 19 kanallı EEG kullanılır. Elektrotlar bu sisteme göre kafatası bölgelerine yerleştirilir. Tek numaralı elektrotlar beynin sol yarım küresine yerleştirilirken, çift numaralı elektrotlar sağ yarım küreye yerleştirilir. Şekil 1.1’de Uluslararası 10-20 sistemine göre 19 kanallı elektrot yerleşimi gösterilmiştir (Tuncer, 2015).



Şekil 1. 1. Uluslararası 10-20 sistemine göre 19 kanallı elektrot yerleşimi (Tuncer ,2015).

Elektrotların yerleştirildiği bölgeler harflerle gösterilmiştir. Bu harfler beyin loblarını temsil eder. Buna göre:

F: Ön lob (Frontal)

C: Merkez bölge (Central)

T: Şakak bölgesi (Temporal)

P: Yan loblar (Pariyetal)

O: Arka bölge (Oksipital)

İki harfin birleşmesi ise ara elektrot yerleşimini belirtir. Örneğin;

FC: Elektrot ön ve merkez bölgesi arasında konumlandırılır.

PO: Elektrot yan ve arka bölgeler arasında konumlandırılır.

Beyin ile ilgili yapılan çalışmalar neticesinde elde edilen bilgilere göre, beyindeki her bir bölge farklı aktiviteler gerçekleştiğinde etkin olmaktadır. Örneğin;

CZ ve C3 bölgelerine yerleştirilen elektrotlardan duyusal ve korteks aktiviteleri esnasında daha belirgin ölçümler alınır. FZ bölgesine yerleştirilen elektrotlardan bilinçli ve motivasyon aktiviteleri sırasında belirgin ölçümler elde edildiği görülmüştür (İmak, 2019).

1.2 Problemin Tanımlanması ve Tezin Amacı

Alkol bağımlılığından her yıl milyonlarca insan maddi ve manevi etkilenmektedir. Hem bağımlı olanın kendi fiziksel ve psikolojik zarar görmesinin yanı sıra başta ailesi ve yakın çevresi olmak üzere tüm toplumu etkileyen büyük bir sorun olarak göze çarpmaktadır. Bu yüzden hastalığın teşhisi bu noktada çok önemlidir. Alkol bağımlılığından en çok beyin fonksiyonları etkilenmektedir. Beyin görüntüleme ve sinyalleri sayesinde bu bağımlılığın beyne ne denli zarar verdiği günümüzde daha çok anlaşılmaya başlanmıştır.

Bu tez çalışmasında EEG işaretleri ile alkole bağımlılığın tespiti üzerinde durulmuştur. Beş farklı makine öğrenmesi teknikleri ile sınıflandırma başarımlarının analizi ve bu başarımlar doğrultusunda algoritmaların çalışma süreleri değerlendirilmiştir.

1.3 Tezin Yapısı

Bu tez çalışması giriş, kaynak araştırması, materyal ve yöntem, bulgular, sonuç ve öneriler olmak üzere beş bölümden oluşmaktadır.

Giriş bölümünde EEG sinyallerinin kullanıldığı alanlar, ölçümü gibi genel bilgiler verilmiştir. Ayrıca problemin tanımlanması ve tezin amacı belirtilmiştir.

Kaynak Araştırması bölümü çalışma ile ilgili kaynak araştırmasını içermektedir.

Materyal ve Yöntem bölümünde veri analizi, öznitelik seçimi, Çok Katmanlı Yapay Sinir Ağları, Evrimsel Sinir Ağları, K-En Yakın Komşu Algoritması, Xgboost Algoritması, Rassal Orman Algoritması ve sınıflandırma başarımlarına yer verilmiştir.

Bulgular bölümünde kullanılan algoritmaların başarımları ayrı ayrı verilip, sınıflandırma ve çalışma süreleri açısından başarımların karşılaştırılması ve değerlendirilmesi yapılmıştır.

Sonuç ve Öneriler bölümünde tez çalışmasında elde edilen sonuçlar özetlenmekte ve ileride konuyla ilgili yapılabilecek çalışmalar verilmektedir.

2. KAYNAK ARAŞTIRMASI

Güler ve arkadaşları, çalışmalarında ağır beyin yaralanmalarında travmanın şiddetinin tespit edilmesi için EEG sinyallerine ayrık zamanlı fourier dönüşümü uygulayıp, yapay sinir ağlarıyla sınıflandırma yapmıştır. %87 oranında bir sınıflandırma performansı elde edilmiştir (Güler vd., 2009). Ersöz ve Özşen, yaptıkları çalışmada EEG sinyallerini yapay sinir ağları yöntemiyle analiz edip uyku evrelerini tespit etmişlerdir. Neticede dalgacık dönüşümü ve hızlı fourier dönüşümü için yakın sonuçlar elde edilmiştir (Ersöz ve Özşen, 2011). Sharanreddy ve Kulkarni, beyin tümörü olan hastalara ait EEG sinyallerini ve epilepsi nöbeti esnasında oluşan EEG sinyallerini analiz etmişlerdir. Yapay sinir ağları ve çoklu dalgacık tekniği kullanılmıştır (Sharanreddy ve Kulkarni, 2013). Kalaivani ve arkadaşları, yaptıkları çalışmada beyindeki anormal aktiviteleri tespit etmek için, EEG sinyallerinden ayrık dalgacık dönüşümü tekniği ile özellik çıkartımı yapıp kümeleme algoritmasıyla sınıflandırma işlemi yapmıştır (Kalaivani vd.,2014). Gajic ve arkadaşları, epileptik nöbetlerin teşhis edilmesi için istatistiksel örüntü tanıma ve dalgacık dönüşümü teknikleriyle EEG sinyalleri sınıflandırmıştır (Gajic vd., 2014.) O'Shea ve arkadaşları, yaptıkları çalışmada konvolüsyonel sinir ağları ve destek vektör makineleri algoritmalarını yenidoğan nöbet teşhisinde kullanmışlardır. İki algoritmada da benzer başarımlar elde etmişlerdir (O'Shea vd., 2017). Vieira ve arkadaşları, beyin ile ilgili hastalıkların sınıflandırılmasında derin sinir ağlarıyla ilgili yapılan çalışmaları incelemişlerdir. Neticede derin sinir ağlarının psikolojik ve nörolojik hastalıkların teşhis edilmesinde kullanılabileceğini belirtmişlerdir (Vieira vd.,2017). Chambon ve arkadaşları, uyku karakteristiklerini sınıflandırmak için, uyku anında ölçülen EEG sinyallerini derin sinir ağlarıyla analiz etmişlerdir (Chambon vd, 2018). İpek, yaptığı çalışmasında, EEG sinyallerini konvolüsyonel sinir ağları ve destek vektörleri makinesi algoritmalarını sentezleyerek incelemiştir (İpek,2018). Gül, çalışmasında lösemili ve sağlıklı bireylere ait EEG sinyallerini dalgacık dönüşümü tekniklerini ve ileri beslemeli yapay sinir ağlarını kullanarak sınıflandırmıştır. Bu çalışmada dalgacık ailelerinin sınıflandırmaya etkisi ortaya koyulmuştur (Gül, 2018). Edenberg ve arkadaşları, genetik varyasyonların alkol bağımlılığını ve aşırı tüketimini nasıl etkilediğini incelemişlerdir (Edenberg vd, 2019). Rieg ve arkadaşları alkol

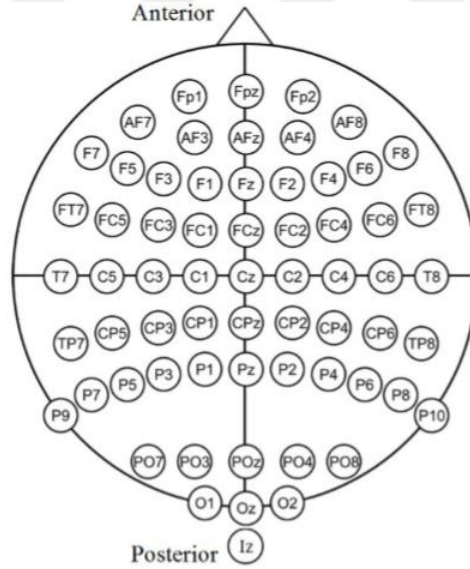
bağımlılığını tespit eden bir araştırma yapmışlardır. Rassal Orman algoritması ile sınıflandırılmıştır (Rieg vd., 2019). Bavkar ve arkadaşları, yaptıkları çalışmada alkolizm tespiti için EEG sinyallerini dar bant geçiren ve butterworth filtreleriyle daha kullanışlı hale getirip özellikle alfa ve gamma bandlarındaki sinyallerin özelliklerinden faydalanarak hibrit bir özellik çıkarım metodu geliştirmiştir (Bavkar vd., 2019). Wang ve arkadaşları, yaptıkları çalışmada alkol bağımlılığını tespit etmek için konvolüsyonel sinir ağlarını kullanarak başarıyı artırmaya yönelik çalışmalar yapmışlardır (Wang vd., 2020).



3. MATERYAL VE YÖNTEM

3.1 Kullanılan EEG Veri Yapısı

Çalışmada kullanılan Veri seti UCI veri tabanı üzerinden sağlanmıştır (UCI,2020). Bu veriler University of New York Health Center’da bulunan Neurodynamics Laboratory bölümünden elde edilmiştir. Deneklerin kafatası yüzeyine yerleştirilen elektrotların pozisyonu (uluslararası 10-20 sistemine göre) Şekil 3.1’teki gibi konumlandırılmıştır (Ge ve ark., 2017). Ölçüm değerleri 256 Hz’de örneklenmiştir.

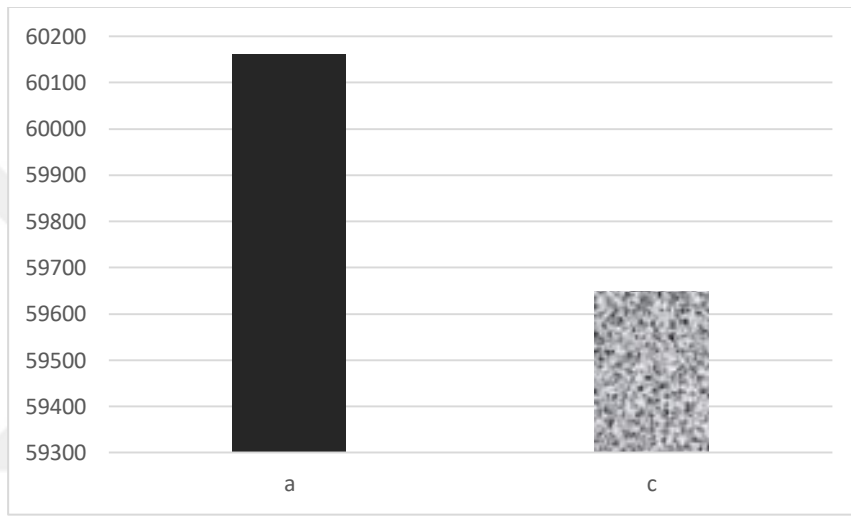


Şekil 3.1. Sensör konumları (Ge ve ark., 2017).

Alkolik grup (a) ve kontrol grubu (c) olmak üzere iki grup bulunmaktadır. EEG verileri için deneklere farklı resimler göstererek beyin dalgaları incelenmiştir. Uyarıların gösterilme durumuna göre üç farklı durum mevcuttur. Sadece S1 uyarısına maruz bırakılan durum (tek bir nesne gösteriliyor), S1’e benzer başka bir nesne gösterilmesi, S1’den farklı bir başka resim gösterilmesi şeklindedir.

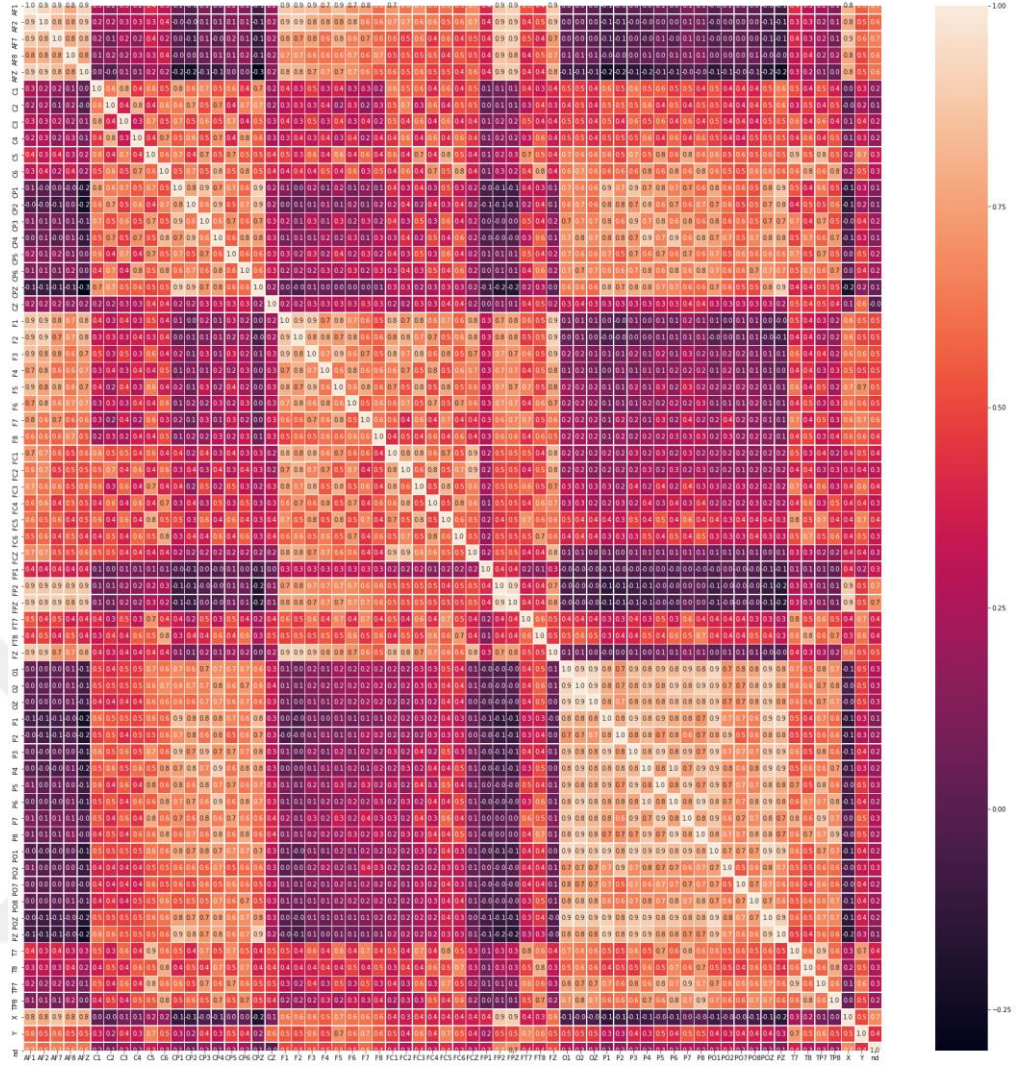
Veriler eğitim ve test klasörleri şeklindedir. Eğitim klasöründe 468 csv uzantılı dosya, test klasöründe ise 480 csv uzantılı dosya bulunmaktadır. Her bir csv uzantılı

dosya ise kendi içinde ölçüm sonuçları içermektedir. Dağınık halde olan veriler her bir denek için atanmış seri kod, deney sayısı ve uyarıların gösterilme durumları temel alınarak veri yeniden düzenlenmiştir. Düzenleme sonucunda eğitim klasöründe 119808x69 veri boyutu, test klasöründe ise 122880x69 veri boyutu elde edilmiştir. 64 sensör bulunmaktadır. 64 sensörün 61 tanesi aktif kanal, 3 tanesi ise referans kanalıdır. Bu sensörlerden gelen veriler kullanılmıştır. Eğitim verisindeki alkol grubu (a) ve kontrol grubu (c) sayıları Şekil 3.2’de gösterilmiştir. Alkolik özellik gösteren veri sayısı 60160 ve kontrol grubu olan veri sayısı ise 59648’dir.

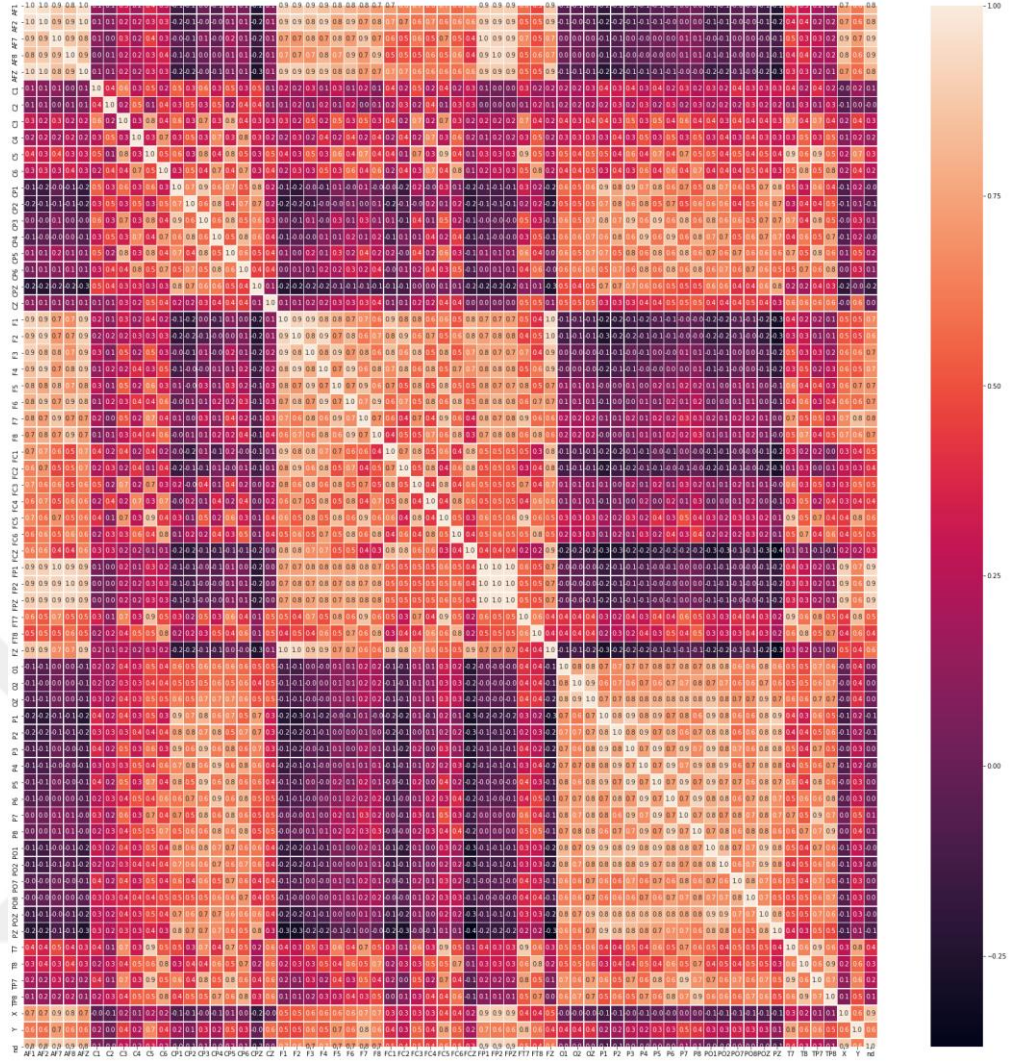


Şekil 3. 2. Eğitim veri setindeki alkol ve kontrol grubu dağılımı

Sadece Alkolik (a) gruba ait sensör değerlerinin korelasyon ilişkisi şekil 3.3’te verilmiştir. Şekil 3.4’te kontrol (c) grubuna ait sensör verilerinin korelasyon ilişkisi sunulmuştur. Şekil 3.3 ve şekil 3.4’e bakıldığında birbirine yakın bölgeler arasında yüksek korelasyon çıktığını söyleyebiliriz. Daha detaylı incelendiğinde FP1 kontrol için yüksek korelasyon içerirken, alkolde oldukça düşük korelasyon içermektedir. Alkolik grup için FP1, (AF1, AF2, AF3, AF8, AFZ) ile korelasyonu oldukça düşük iken, kontrol grubu için oldukça yüksektir. Fakat beynin, tepki değerlerinin iki grup arasında önemli ölçüde farklı olduğu belirli bir bölgenin olduğunu söylemek zor görünmektedir.

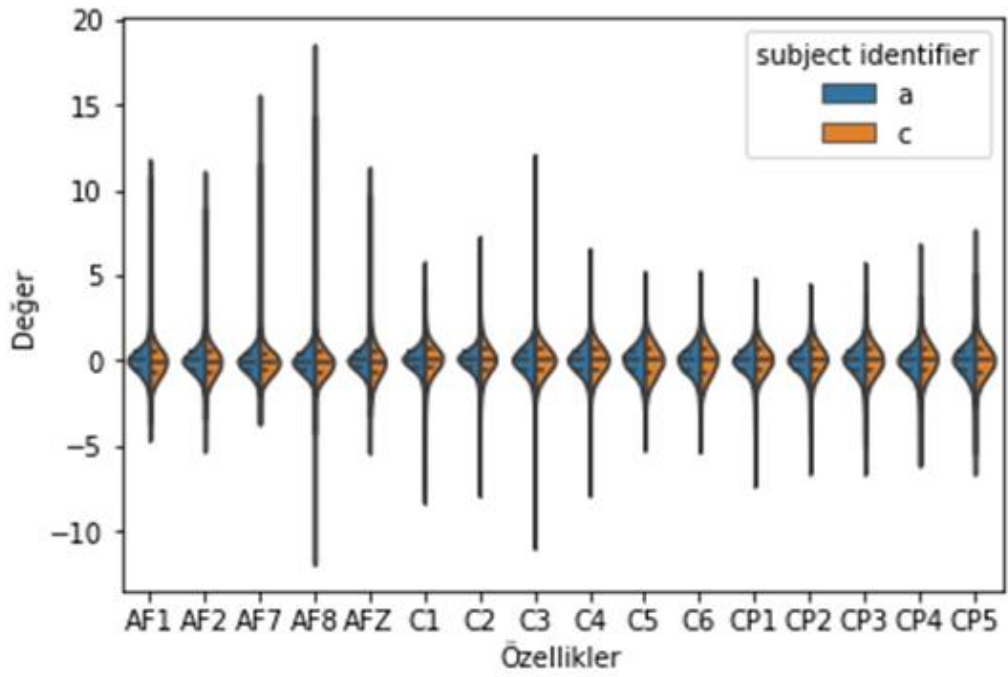


Şekil 3.3. Alkolik (a) gruba ait sensör değerlerinin korelasyon ilişkisi



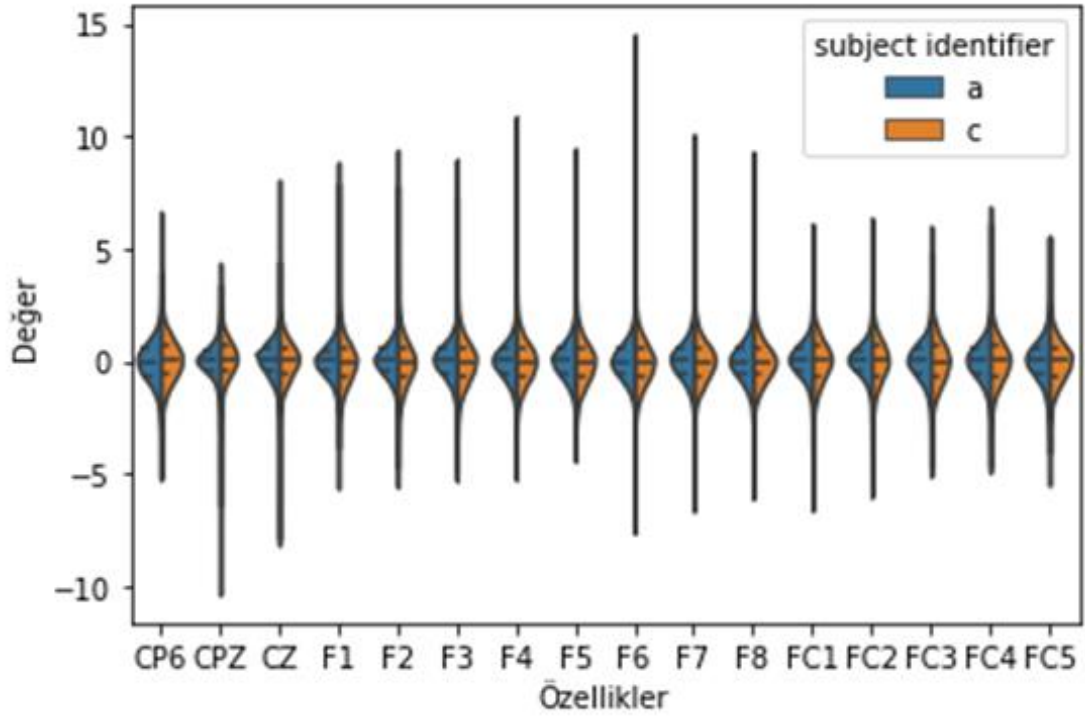
Şekil 3.4. Kontrol (c) grubuna ait sensör verilerinin korelasyon ilişkisi

Veriler arasındaki ilişkiyi daha iyi görebilmek için her bir şekilde 16 sensör pozisyonu özelliği olacak dört adet grafik çizdirilmiştir. Şekiller her bir özellik için a ve c sınıfları arasındaki ilişkiyi göstermektedir. Şekil 3.5'te ilk 16 özellik arasındaki korelasyon gösterilmiştir:



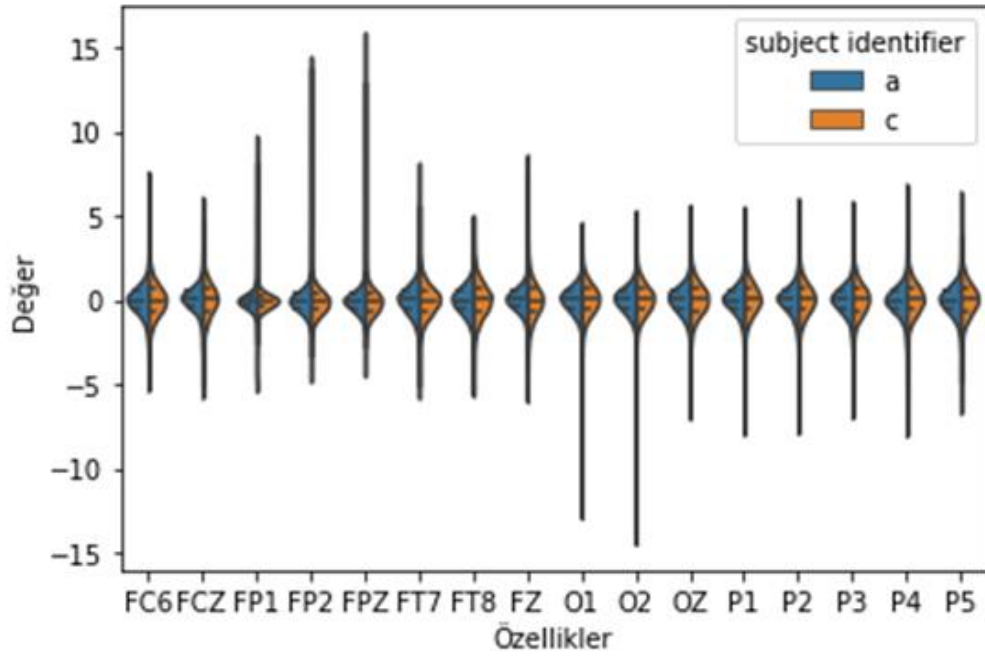
Şekil 3. 5. İlk 16 sensör pozisyonu arasındaki korelasyon

Şekil 3.6’te 16-32 aralığındaki sensör pozisyonları arasındaki korelasyon gösterilmiştir.



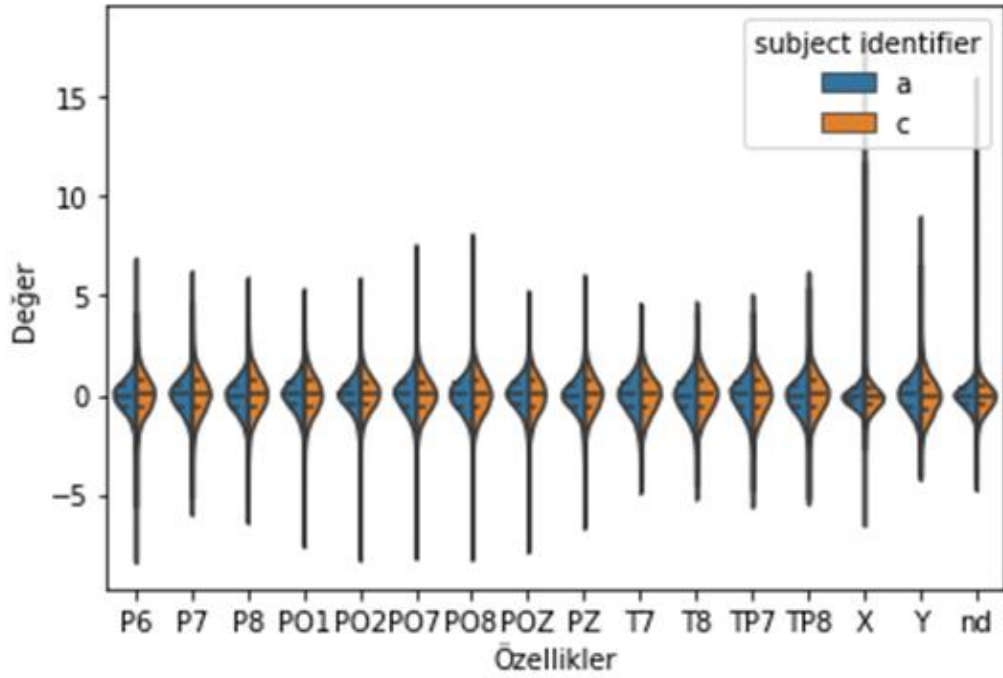
Şekil 3.6. 16 ile 32 no’lu sensör pozisyonları arasındaki korelasyon

Şekil 3.7’de 32-48 aralığındaki özellikler arasındaki korelasyon gösterilmiştir:



Şekil 3.7. 32 ile 48 arasındaki sensör özellikleri arasındaki korelasyon

Şekil 3.8’de ise 48-64 arasındaki özelliklerin ilişkisi gösterilmiştir.

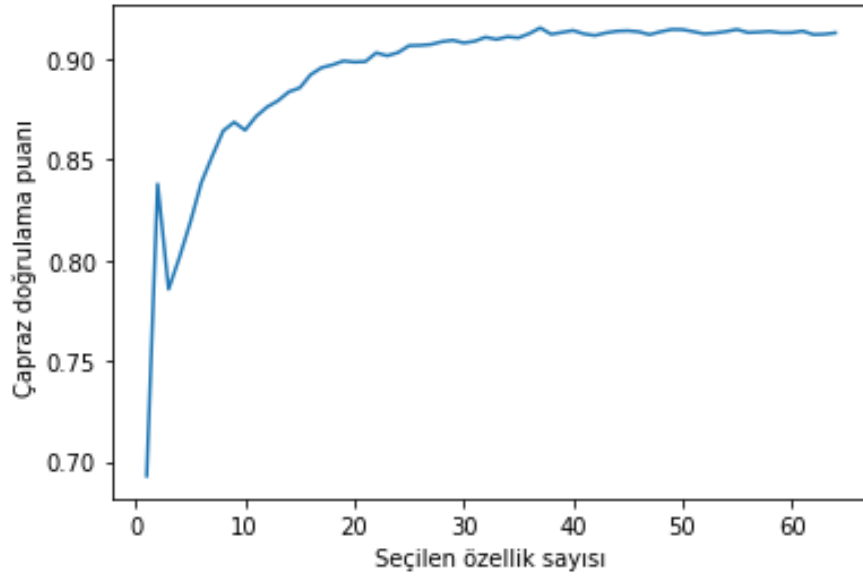


Şekil 3. 8. 48 ile 64 arasındaki sensör özellikleri arasındaki ilişki

3.2 Özyinelemeli Öznitelik Seçimi

Özyinelemeli Öznitelik Seçimi (ÖÖS) yöntemi oldukça popülerdir. ÖÖS sarmal (wrapper) tipi bir özellik seçim algoritmasıdır. Kullanımı ve konfigürasyonu basittir. Eğitim verilerinde bulunan sütunlardaki özelliklerden en önemlileri seçmede ve hedef değişkeni tahmin etmede oldukça etkili olan bir tekniktir. ÖÖS, eğitim verilerindeki tüm özellikleri analiz ederek, gereksiz tüm özellikleri kaldırır ve sadece en önemli özellikler kalana kadar devam eder. Böylece kullanılan makine öğrenme algoritmalarının daha başarılı bir sınıflandırma yapmasını sağlar. (Brownlee, 2020).

Çalışmada, çapraz doğrulama ile özyinelemeli öznitelik seçim yöntemi kullanılmıştır. Puanlama olarak doğruluk oranı temel alınmıştır. Şekil 3.9’da seçilen özellik sayısına göre çapraz doğrulama puanı görülmektedir.



Şekil 3.9. Özellik sayısına göre çapraz doğrulama puanı

Ham veri, öznitelik seçim öncesi 61,34 MB iken, öznitelik seçim sonrası 35,46 MB’a düşmüştür. Çizelge 3.1’de veri setinin özellikleri verilmiştir. Özyinelemeli öznitelik seçimine göre ‘AF1’, ‘AF2’, ‘AF7’, ‘C5’, ‘CP3’, ‘CP4’, ‘CP5’, ‘CP6’, ‘F1’, ‘F3’, ‘F4’, ‘F5’, ‘F6’, ‘F7’, ‘FC4’, ‘FC6’, ‘FPZ’, ‘FT8’, ‘OZ’, ‘P2’, ‘P6’, ‘P7’, ‘PO1’, ‘PO7’, ‘T7’, ‘T8’, ‘TP7’ en kötü özelliğe sahip kolonlar kaldırılmıştır.

Çizelge 3. 1. Veri seti özellikleri

	Eğitim Verisi	Test Verisi	Veriseti Boyutu
Ham Veri Seti	119808x64	122880x64	61.34 MB
Öznitelik seçimli veri seti	119808x37	122880x37	35.46 MB

3.3 Standardizasyon ve Normalizasyon

Sınıflandırmak istediğimiz verilerde farklı türde değişkenlerle karşılaşabiliriz. Bu değişkenlerin bulunduğu aralıklar çok farklı olabilmektedir. Bu sorunla başa çıkmak için veri ön işleme aşamasında verileri yeniden ölçeklendirme gereği duyarız. Bunu da standardizasyon ve normalizasyon işlemleri adı altında gerçekleştirebiliriz. Bu işlemlerin uygulanmasının amacı veride bulunan tüm özellikleri aynı ölçek seviyesine getirmektir. Böylece makine öğrenme algoritmaları veriyi daha kolay ve doğru bir şekilde sınıflandırabilir.

Normalizasyon tekniğinin; verilerin dağılımının nasıl olduğu bilinmediği veya Gaussian (çan eğrisi) tarzı dağılım olmadığı bilindiğinde kullanılması avantajlıdır. Veri setinde farklı ölçeklemeler varsa ve beraber kullanılan algoritmalar veri hakkında varsayımlarda bulunamayan k-en yakın komşu ve yapay sinir ağları gibi yöntemlerse kullanılması yararlı olabilir. Standardizasyon tekniğinde; ise durum tersidir. Gaussian dağılımı olan ve veri hakkında varsayımlar yapabilen lineer regresyon, lojistik regresyon ve lineer ayırık analiz gibi algoritmalarla beraber kullanılabilir. (Lakshmanan, 2019). Bu çalışmada L2 normalizasyon tekniği kullanılmıştır. Normalizasyon işlemi özel bir sınırlandırma olarak da yorumlanabilir. Aşırı öğrenme durumunu önlemek için ve makine öğrenme modellerinin performanslarını artırmak için kullanılır.

Son yıllarda birçok regülasyon yöntemi geliştirilmiştir. Bunlardan bazıları da L1 ve L2 normlarıdır. L2 regülasyonu özellikle, sinir ağındaki ağırlık değerlerinin 0'a yakın olmasını sağlamaktadır. Böylece giriş sinyallerindeki küçük değişiklikler çıkışta büyük farklılıklar doğurmayacaktır (Zhao vd., 2020).

L1 normu hedef değeri (Y_i) ve tahmini değerler ($f(x_i)$) arasındaki mutlak farklılıkların toplamını en aza indirir. Eşitlik 3.1’de bu ifade verilmiştir (Gadade, 2019).

$$S = \sum_{i=1}^n |y_i - f(x_i)| \quad (3.1)$$

L2 norm eğitim sırasında aykırı değerleri dikkate alır. Verilerdeki aykırı değerlere karşı dayanıklıdır. İfadesi eşitlik 3.2’de verilmiştir (Gadade, 2019) .

$$S = \sum_{i=1}^n |y_i - f(x_i)|^2 \quad (3.2)$$

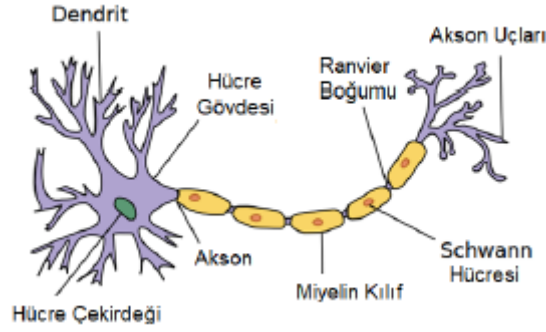
3.4 Çok Katmanlı Yapay Sinir Ağları

Yapay sinir ağları, insan beyninin çalışma sisteminden esinlenerek geliştirilmiş bir yapay zeka algoritmasıdır. Yapay sinir ağları, yazılımsal ve donanımsal olarak farklı şekillerde modellenir (Öztemel, 2006; Tosun, 2004). Çizelge 3.2’de yapay sinir ağları ve biyolojik sinir ağlarının karşılaştırılması verilmiştir.

Çizelge 3. 2. Yapay sinir ağları ve biyolojik sinir ağlarının karşılaştırılması (Elmas, 2016)

Beyin Sinir Sistemi	Yapay Sinir Sistemi
Nöron	Sinir, ara Düğüm
Dentrit	Toplama Fonksiyonu
Hücre Gövdesi	Aktivasyon Fonksiyonu
Akson	Sinir çıkış düğümleri
Sinapslar	Sinirler arası bağlantı ağırlıkları

Şekil 3.10'da biyolojik bir nöronun yapısını görülmektedir (Zhang vd., 2018).

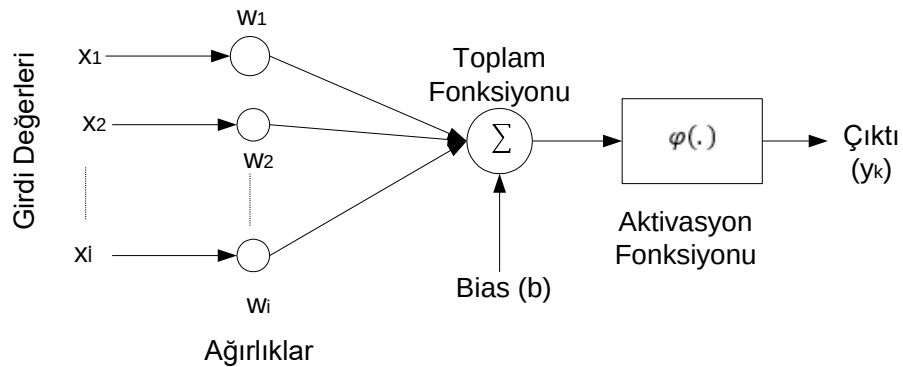


Şekil 3 .10. Biyolojik nöron

Dış ortam sensörlerinden veya diğer nöronlardan gelen x_i bilgisi, dendritler aracılığıyla giriş olarak alınır. Girişlerin önem değeri w_i ağırlıklarıyla belirlenir. Daha sonra bu bilgilerin toplamına b (bias) ağırlığı eklenir. Eşitlik 3.3'de ifade verilmiştir (Zhang vd., 2018).

$$z = b + \sum_{i=1}^n x_i w_i \quad (3.3)$$

İnsan beynindeki sinir ağı sisteminde nöronlar görev yaptığı gibi yapay sinir ağında da yapay nöronlar yer alır. Bunlar bir araya getirilerek ağ yapısına bağlanır. Bu ağ yapısı, her bir input değerinin bir veya daha fazla çıkış değişkenine bağlandığı bir ağ oluşturur. Temel bir yapay sinir ağında, girişler, ağırlıklar, toplam fonksiyonu, aktivasyon fonksiyonu ve çıkış olmak üzere 5 ana birimden oluşur. Şekil 3.11'de temel yapay sinir ağının modeli gösterilmiştir. (Gül, 2018).



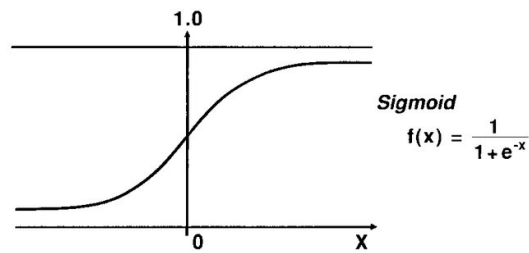
Şekil 3.11. Temel bir YSA örneği (Calp, 2019)

Girişler, çevreden gelen bilgileri sinirlere aktarır. Ağırlıklar, girişlerin yapay sinir üzerindeki önemini belirleyen katsayılardır. Önem değerinin büyük veya küçük olması ilgili sinirin yapay sinir ağına güçlü ya da zayıf bağlanması anlamına gelir. Toplam fonksiyonu, her bir giriş değeri ile ilgili ağırlığın çarpımlarının toplamalarını, eşik değeri ile toplayan fonksiyondur. Aktivasyon fonksiyonu, net girdiyi işleyerek sinir ağının buna karşılık üreteceği çıktıyı hesaplar. Aktivasyon fonksiyonu olarak genellikle doğrusal olmayan bir fonksiyon seçilir. Fonksiyon seçilirken fonksiyonun türevinin kolaylıkla elde edilebilir olması önemlidir. YSA ile yapılan çalışmalarda sigmoid, lineer, birim basamak, sinüs, hiperbolik tanjant fonksiyonları kullanılabilmektedir. Çıktı, aktivasyon işlemi sonucunun dış ortama iletildiği yerdir. Sinirin çıkışı, kendinden sonra gelen diğer sinirlere giriş olabilir. (Elmas, 2016.)

Yapay sinir ağlarında, katmanlar arasında nöronlar konumlandırılmıştır. Bu katmanlar girdi katmanı, gizli katman ve çıkış katmanı olarak birimlere ayrılmıştır. Bu katmanlar birbirlerinden girdileri alır ve çıktı üretilip bir sonraki katmana iletir. Girdi verilerinin boyut miktarı girdi katmanındaki düğüm sayısını belirler. Gizli katman ise sinir ağı içindeki hesaplamaları yapar. Bu katmandaki nöron sayısı arttıkça ağın derinliği de artar (Fedjaev, 2017).

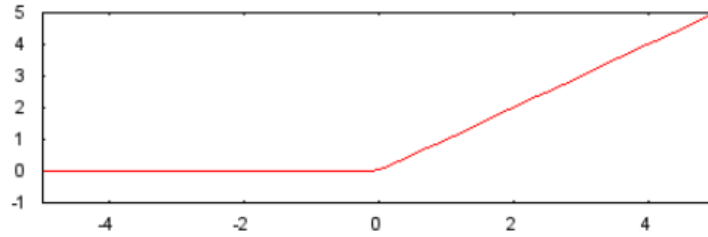
3.4.1 Aktivasyon Fonksiyonları

Çok katmanlı sinir ağlarında, ağa giren bilginin işlenmesini ve buna göre çıktı oluşturmalarını sağlayan lineer olmayan fonksiyonlardır. Çalışmada kullanılan aktivasyon fonksiyonları Şekil 3.12 ve Şekil 3.13’de gösterildiği gibi Sigmoid ve ReLu aktivasyon fonksiyonlarıdır.



Şekil 3.12. Sigmoid aktivasyon fonksiyonu (Næs ve ark., 1993)

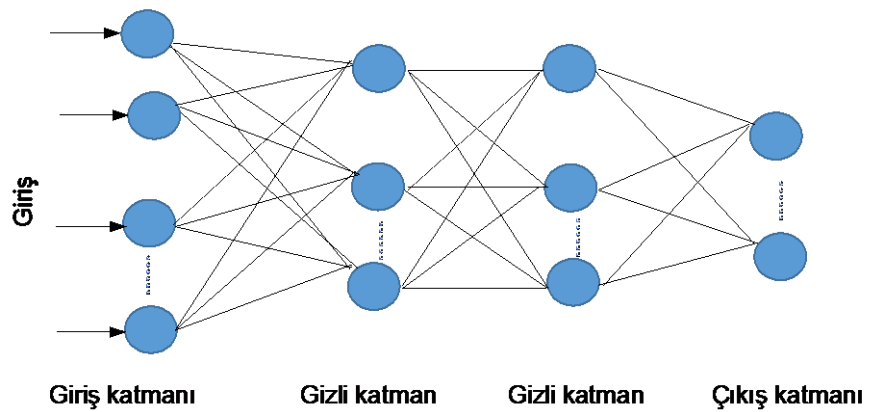
Sigmoid fonksiyonu, Şekil 3.12’de görüldüğü üzere lineer olmayan ve giriş değerleri için 0 ile 1 arasında değerler alır. Rectified linear unit (ReLU), özellikle gizli katmanda çok sık kullanılan bir aktivasyon fonksiyonudur. Girdi bilgisinin değeri 0’dan büyük ise aynı kendi değerini çıkışta verir, 0’dan küçük ise çıkışta 0 değerini üretir (Jarrett vd., 2009).



Şekil 3.13. ReLu aktivasyon fonksiyonu

3.4.2 Çok Katmanlı Yapay Sinir Ağının Yapısı

Çok katmanlı yapay sinir ağlarında, katmanlar ağırlıklarla önceki katmana bağlıdır. Rosentblatt’ın Perceptronu (tek sinir ağları) ortaya atmasından sonra çok katmanlı ileri beslemeli ağların birçok çeşidi önerilmiştir. Geri yayılım algoritması ile ağlar daha uygulanabilir olmuştur (Elmas, 2016). Şekil 3.14’de çok katmanlı yapay sinir ağ yapısı görülmektedir.



Şekil 3.14. Çok katmanlı yapay sinir ağ yapısı

Tipik bir geri yayılım algoritmasında bir giriş katmanı, bir çıkış katmanı ve en az bir gizli katman vardır. Problemin karmaşıklığı arttıkça ağıdaki gizli katman sayısı da artış gösterecektir (Elmas, 2016).

YSA modelleri nöronlar arasındaki işleyiş yönüne göre ileri beslemeli ve geri beslemeli olmak üzere ikiye ayrılır. İleri beslemeli ağlarda hücreler katman tabakaları şeklinde düzenlenip, giriş katman birimleri çevreden aldığı bilgileri direk gizli katman birimlerine iletir. Bu ağlar çok katmanlı ileri beslemeli ağlar olarak da isimlendirilirler. Geri beslemeli ağlarda ise en az bir hücrenin çıktısı, diğer hücrelere veya kendine girdi olarak verilir. Bu ağ yapıları özellikle gözetimsiz öğrenme modelinde kullanılan yapay sinir ağlarında uygulanmaktadır (Gürsoy, 2009).

Daha önceki yapılan çalışmalarda birçok ağ modeli önerilmesine karşın, geri beslemeli öğrenme sistemine sahip çok katmanlı ileri beslemeli ağ, pratik uygulamalarda en yaygın kullanılan modeldir (Kantarzdic, 2011). Hücreler katman tabakaları şeklinde oluşturulur ve bir katmandaki hücrelerin çıktıları bir sonraki katmana ağırlık değerleri aracılığıyla girdi olarak verilir (Murugesan ve Sukanesh, 2009).

3.4.3 Yapay Sinir Ağında Öğrenme

Sinir ağında düğüm noktalarındaki ağırlık değerlerinin belirlenmesine ağı eğitilmesi denilir. Başlangıçta bu değerler rasgele seçilirken daha sonraki eğitimlerde çıktı değerlerine göre tekrar tekrar değerler değiştirilerek eğitim gerçekleştirilir buna sinir ağının öğrenmesi denir (Dandıl ve Çevik, 2011).

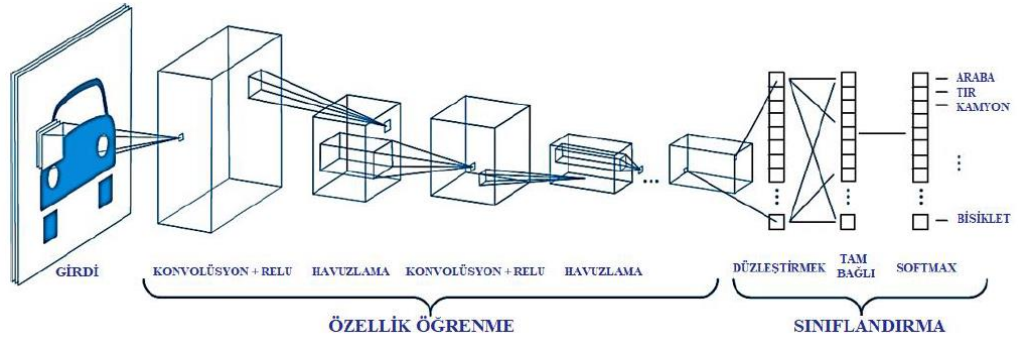
Yapay sinir ağına girdi değerleri verildikten sonra ilk eğitim gerçekleşir ve daha sonra ağırlık değerlerinin yenilenmesiyle çıktı değerleri ile beklenen değerler arasındaki hata karelerinin ortalamasının yapılan çalışma için en küçük değeri olması istenir. Karelerinin ortalaması hedeflenen değerden daha küçük değilse, öğrenme ağırlık değerlerinin değiştirilmesiyle devam eder (Şen, 2004).

3.5 Evrişimsel Sinir Ağları

Evrişimsel Sinir Ağları (ESA) çok katmanlı sinir ağlarının bir çeşididir. Hayvanların görme sistemlerinden esinlenilmiştir. Dil işleme, ses ve görüntü tanıma ve biyomedikal gibi alanlarda bu ağlardan faydalanılmakta olup, en iyi sonuçlar görüntü tanıma alanında alınmıştır (Tüfekçi ve Karpat, 2019).

ESA, klasik yapay sinir ağları modelinde kullanılan vektörler yerine matrisleri kullanır. Bunun en önemli avantajı ise, komşu tabakalar arasındaki bağlantıyı ve iletilen bilgilerin kaybolmasını engellemektir. Traore ve arkadaşları mikroskobik görüntülerin sınıflandırılması için yaptıkları çalışmada evrişimsel sinir ağlarını kullanmıştır. Böylelikle sıtma ve kolera hastalıklarının etken virüslerinin tespit edilmesi için akıllı mikroskoplara karar verme sürecinde destek olması sağlanmıştır (Traore vd., 2018).

ESA modeli iki basamaklı çalışır. İlkinde birbiriyle bağlantılı öznitelikler çıkarılır. İkincisinde ise çok katmanlı sinir ağlarından faydalanılarak sınıflandırma yapılması sağlanır. Çıkarılan özniteliklerin haritaları, küçük boyutlu filtrelerin sinir ağının giriş verisi üzerinde konvolüsyon uygulanmasıyla elde edilir. Eğitim boyunca, hatanın geriye yayılımı algoritması kullanılarak filtrelerin parametreleri güncellenir. Bu şekilde filtrelerin içerikleri öğrenilmiş olur. Daha sonra öznitelik haritaları havuzlama işlemine tabi tutulur. Bu sayede çıkarılan özelliklerin çeşitliliği ve boyutu en aza indirgenmiş olur. Daha sonra haritalar vektör şekline getirilerek sinir ağının girdisi olarak verilir (Işık ve Artuner, 2018). Şekil 3.15’de genel ESA yapısı görülmektedir.



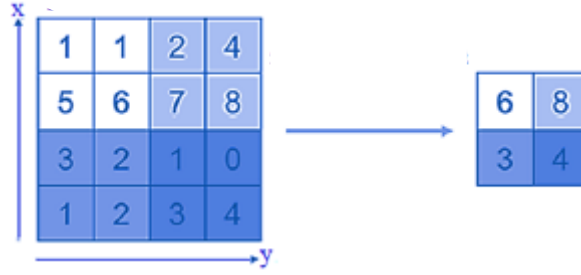
Şekil 3.15. Genel ESA yapısı (MathWorks,2018a)

Bir boyutlu ESA, ÇKYSA ve birbiriyle tamamen bağlı olan veya havuzlama işlemi gören bir veya daha fazla konvolüsyonel tabakaları kullanır. Konvolüsyon işlemi konvolüsyon tabakaları arasındaki her basamakta uygulanır. Bundan dolayı daha derin tabakalar her bir giriş bilgisinin özelliğine bağlı olarak çıkışı etkiler. Konvolüsyon tabakası matrisleri kullanarak özellik çıkarımı yapar. Bu tabakanın çıkışı bir özellik haritası çıkarır. Havuzlama katmanı önemli bilgileri elde etmek için özelliklerin sayısını düşürür (Choubey vd., 2019).

Verilerin giriş katmanına uygulanmasında boyutu ve özellikleri çok önemlidir. Boyut ve özellikler arttıkça hafıza gereksinimi, eğitim süresi ve test süresi artarken, ağın başarımlı yükselecektir (İnik ve Ülker, 2017).

Sisteme girdi olarak verilen verinin belirli özelliklerinin açığa çıkarılmasını sağlamak için iç içe filtrelerden geçirilir (Mathworks, 2018a). Konvolüsyon katmanlarının kullanılmasının amacı girdilerin özelliklerini tespit etmektir. Konvolüsyon katmanı, ağı temel birimini ifade eder. Ağı hesaplama yapan bölümünün esas kısmını taşır (Mishra, 2019).

Havuzlama katmanı sinir ağına öğrenmesi gereken özelliklerin sayısını azaltmak için, doğrusal olmayan verilerin çıktılarını basite indirir. Ayrıca ağıdaki hesaplama karmaşasını ve mekânsal boyutu azaltmayı sağlar (Ghoury, 2019). Şekil 3.16’da iki adımlı maksimum havuzlama işlemi görülmektedir.



Şekil 3.16. Havuzlama işlemi (Mishra,2019)

Tam bağlı katman, sinir ağının sonundaki nöron katmanıdır. Tam bağlı bir katmandaki nöronlar, diğer sinir ağlarına benzer şekilde, kendinden önceki katmanlardaki her bir aktivasyonla tam bağlantıya sahiptir ve çalışma prensibi aynıdır. Bundan dolayı, bir matris çarpımı ve sonra eğim etkisi ile hesaplanabilir. Bu katman giriş ve çıkış arasındaki eşlemeye yardım eder (Mishra, 2019). Son katman, K boyutlu bir vektörü veren tam bağlı bir katmandır. K sinir ağının tahmin etmesi gereken sınıfların sayısıdır (Mathworks, 2018a).

3.6 K-En Yakın Komşu Algoritması

K-En Yakın Komşu (K-EK) algoritması, veriler arasındaki mesafenin ölçülmesi aracılığıyla sınıflandırma yapar. Örnek uzayda birbirine yakın olan verilerin, büyük ihtimalle aynı sınıfa ait olduğu çıkarımı yapılır. Algoritmanın hedefi, verilere ait özelliklerden faydalananak, verileri önceden belirlenen sınıf veya gruplara göre yerleştirmektir (Elesan, 2019). Algoritmanın performansı k en yakın komşu sayısı, mesafe ölçütü ve öznitelik sayısı gibi değişkenlere bağlı olarak değişiklik gösterir. K en yakın komşu sayısının nasıl seçileceğine dair belli bir standart bulunmamaktadır. Deneme yanılma yöntemiyle veya sezgisel olarak belirlenebilir (Bhatia ve Vandana, 2010). Parametrik olmadığı için veriyle ilgili varsayımları çok azdır. Fakat k en yakın komşu sayısının veya mesafe hesabı için hangi ölçüm metodunun kullanılması gerektiğine dair netliğin olmaması da algoritmanın dezavantajlarından (Cunningham ve Delany, 2007).

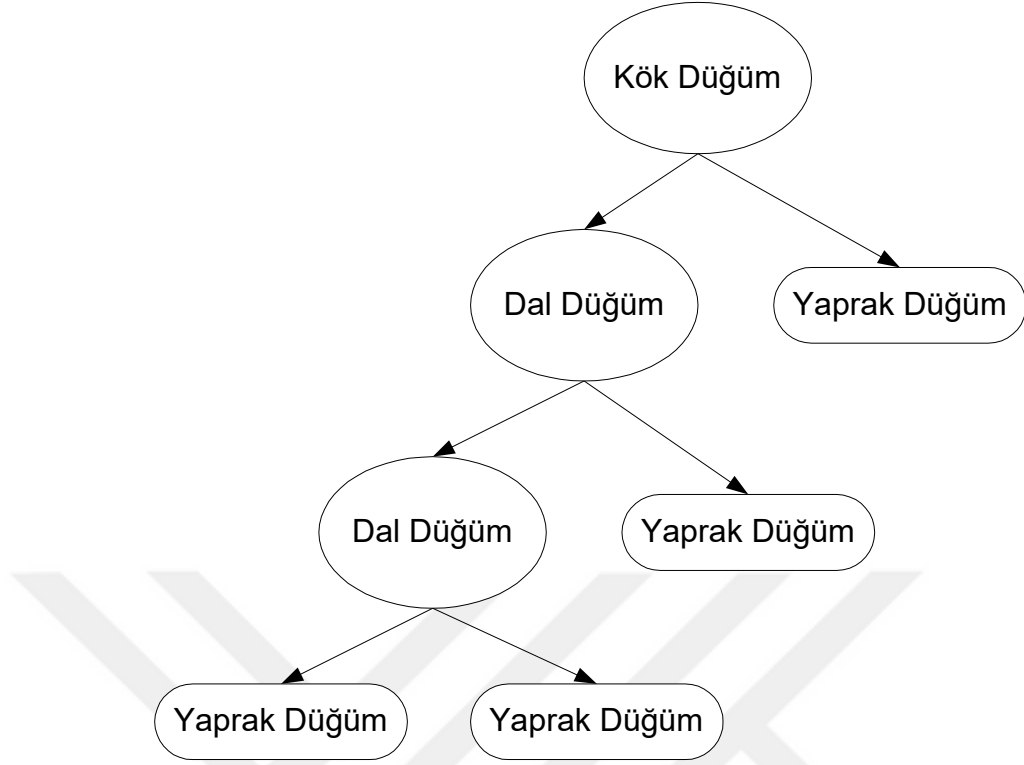
K-EKA parametrelerinden öklid uzaklığı algoritma için en çok tercih edilen mesafe ölçütüdür. Öklid bağıntısında iki nokta arasındaki en kısa mesafe olan doğrusal uzunluk bulunur. P ve Q arasındaki Öklid mesafesi $P = (x_1, x_2, \dots, x_n)$ ve $Q = (y_1, y_2, \dots, y_n)$ olmak üzere, Eşitlik (3.4) kullanılarak bulunur (Laribi, 2018).

$$d(x, y) = (\sqrt{\sum_{i=1}^n (x_i - y_i)^2}) \quad (3.4)$$

K-en yakın komşu sayısı ise yeni sınıflandırılacak olan verilerin hangi sınıfa ait olduğunu belirlemede önemlidir. k sayısı, yeni bir verinin sınıflandırılmasında, en yakın komşu sayısının değerlendirilmesinde kullanılır (Elesan, 2019).

3.7 Karar Ağaçları

Karar ağaçları, değişik veri tabanı sistemleriyle birlikte çalışabilmesi, anlaşılmasının kolay olması ve güvenilirlik açısından iyi sonuçlar vermesi nedeniyle makine öğrenmesi yöntemleri içinde en yaygın kullanılanlardan biridir. Karar ağaçları dal, yaprak ve düğüm noktalarından meydana gelen ve yorumlanması oldukça kolay olan bir yöntemdir. Düğüm noktaları değişken elemanları temsil ederken dal ve yapraklar da diğer elemanları temsil eder. Ağaç yapısı giriş verilerini içeren kök yapısıyla başlar ve bu yapıda verilerin sınıflandırılması ve karar ağacının oluşturulması için bazı sorular oluşturulur ve bunlara verilen cevaplara göre yapraklara kadar olan yapı oluşturulur. (Köktürk, 2012). İlk sorunun sorulduğu kök düğümünden ağaç yapısı oluşturulmaya başlanır. Daha sonra alınan cevaplara göre yeni düğüm noktaları oluşur. Bunlar da yeni dallara ayrılır ve eğer yeni soru sorulmadığı yere gelinmişse, herhangi bir sınıfı belirten yaprağa ulaşılmıştır. (Akman, 2010). Karar ağaçlarının genel yapısı Şekil 3.17’de verilmiştir.



 ekil 3.17. Karar a acının genel yapısı (Elesan, 2019)

Karar a a  modelinde her dal d ğ m  bir kuralı temsil eden fonksiyondur. Sınıflandırma a amasında nihai sonuca ulaşılamamışsa, yeni dallar karar vermek i in tekrar olu turulur. Bu dallanmaların sonunda meydana gelen karar d ğ mlerinin sayısı modelin derinli ini ifade eder. Karar a a larındaki derinlik, sınıflandırmak i in se ilen veri setinin  zelli ine ve boyutuna g re de i kenlik g sterir (Carvalho ve Freitas, 2004).

Karar A a larının avantajları olarak ; yorumlama ve anla ılma kolaylı ı, farklı ve devamlı  zelliklerle  alı abilir olması, olu turdu u hiyerar ik yapı  zelliklerin ba lantı y n n  ve  nem de erini ortaya koyması ve b y k veri setleri i in uygulanmasının kolay olması sayılabilir (Murthy, 1998). Dezavantajları olarak ise; sınıflandırılmak istenen veri setinin  zelliklerine ba lı olarak ihtiya  duyulan zaman ve a a  karma ıklı ının de i mesi, verilerden yeterli sayıda  zellik  ıkartılamaz ise karar a acının yeterince bilgiye ula amayabilmesi, uygun sınıflandırma ba arımını yakalamak i in derinli in artırılması ve bunun sonucunda karma ık a a ların olu ması gibi durumlar sayılabilir (Ramazanov, 2020).

Çalışmada karar ağaç çeşitlerinden olan XGBoost (Extreme Gradient Boosting) Algoritması ve Rassal Orman Algoritması kullanılmıştır.

3.7.1 XGBoost Algoritması

XGBoost (eXtreme Gradient Boosting) algoritması Chen ve He tarafından 2015’de geliştirilen kollektif bir ağaç algoritmasıdır. XGBoost algoritması farklı derinlikte karar ağaçları oluşturmada ve aynı anda farklı algoritmaları optimize edebilmesindeki performansı sayesinde regresyon ve sınıflandırma problemlerinde etkili ve verimli çözümler sunar (Zhou vd., 2019).

XGBoost Algoritması (XGBA) birçok mühendislik simülasyonları için hızlı ve güvenilir bir çözüm sunar (Zhang vd., 2020). Algoritmanın genel çalışma prensibi, sürekli yeni ağaçlar oluşturmak ve yeni özellikler çıkararak ağacın büyümesini sağlamaktır. Her zaman yeni bir ağaç ekleyerek son yapılan tahmini iyileştirecek şekilde çalışır. Eğitim tamamlandığında, veri setinin tahminleme aşamasına geçilir (Peng vd. 2019).

XGBA’yı aşağıdaki şekilde ifade edebiliriz. $D=\{(x_i,y_i)\}$ şeklinde bir data seti olarak düşünülebilir. Bu veri setinde n adet örnek ve m adet özellik bulunsun ($|D| = n$, $x_i \in R^m$, $y_i \in R$). Önerilen ağaç modelinin yaklaşık cevabı eşitlik (3.5)’deki gibi olur (Zhang vd., 2020).

$$\hat{y}_i = \phi(x_i) = \sum_{z=1}^Z f_z(x_i), \quad f_z \in F \quad (3.5)$$

F regresyon ağacının uzayını temsil etmektedir. F , eşitlik (3.6)’daki gibi ifade edilebilir (Zhang vd., 2020).

$$F = \{f(x) = w_{q(x)}\} (q: R^m \rightarrow T, w \in R^T) \quad (3.6)$$

Eşitlik (3.6)’da bulunan q ağaçların yapısını belirtir. w yaprak düğümlerinin ağırlığını ve T ise ağaçtaki yaprak sayılarını ifade eder. XGBA için eşitlik (3.7)’deki ifade kullanılarak hatalar azaltılmaya çalışılır ve en iyileme yapılır. Eşitlik (3.7)’de

bulunan l konveks kayıp fonksiyonunu temsil eder. $\hat{y}$ tahmin edilen değeri gösterirken, y_i gerçekte ölçülen değerleri belirtir. t ise hataları minimize etmek için kullanılan iterasyon sayısını gösterir (Zhang vd., 2020).

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_t^{t-1} + f_t(x_i)) + (f_i) \quad (3.7)$$

XGBA'nın avantajları aşağıda maddeler halinde verilmiştir (Morde, 2020).

- Karar ağaçlarını oluştururken iç ve dış döngüler arası geçiş yaparak paralel çalışabilir olması,
- Ağacın maksimum derinliğini belirleyerek karmaşıklığı önlemesi,
- Farklı regülerizasyon tekniklerini beraber kullanarak overfitting (ezberleme) önlenabiliyor olması,
- Ağacı doğru noktalardan ayırarak ağırlık değerlerini ataması

3.7.2 Rassal Orman Algoritması

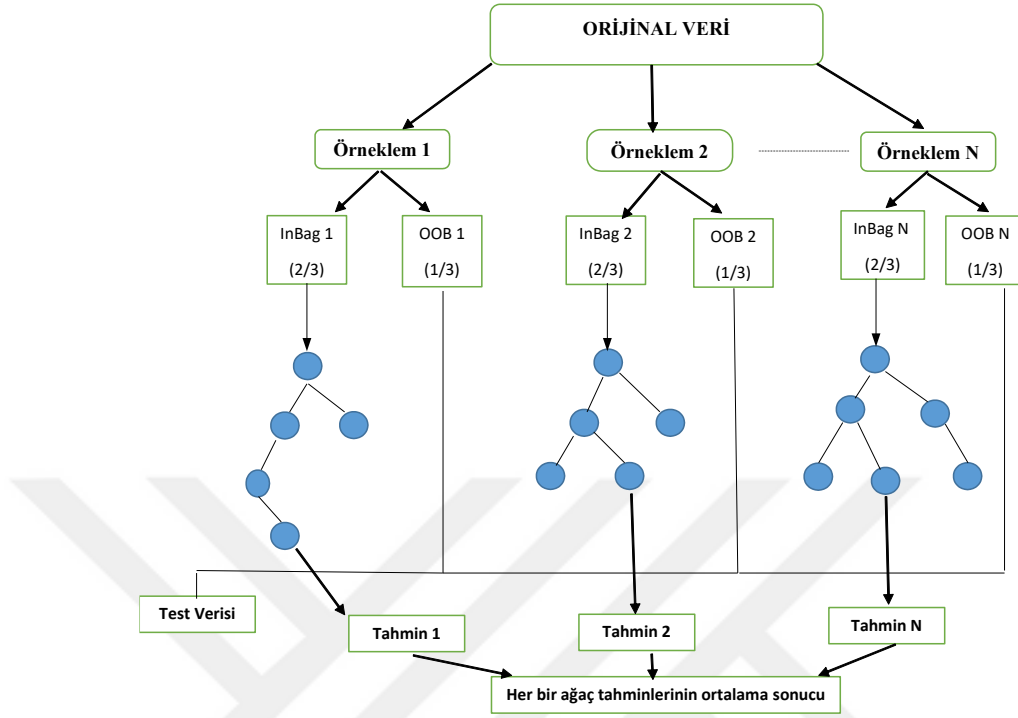
Karar ağaçları sınıflandırma işleminde kullanıldığı zaman, yaşanan en büyük problemlerden biri doğru ağaç yapısını her zaman için oluşturamamaktır. Bu yüzden sınıflandırma sonucu istenilen başarı elde edilemeyebilir. Bu olumsuz etkiyi ortadan kaldırmak veya minimize etmek için rassal orman algoritması geliştirilmiştir. Rassal orman algoritması ilk 1996 yılında Breiman tarafından ortaya atılmıştır. Rassal ormanı oluştururken, rasgele vektörler kullanılarak ağaçlar geliştirilmiş ve veri setinden rastgele seçim için ise bagging (torbalama) tekniği kullanılmıştır (Breiman, 1996). Karar ağaçlarının büyüklüğünü ayarlamak ve karar düğümlerinin nasıl ayrılacağı konusunu çözüme kavuşturmak için 1998 yılında rastgele alt uzay tekniği sunulmuştur. Karar ağaçlarında ezberleme problemini engellemek için veri seti özelliklerinin içinden rastgele bir alt uzay seçilir. Karar ağaçları oluşturulurken, karar düğümlerinde kullanılacak özellik buradan seçilir (Ho, 1998). Breiman 2001 yılında bagging ve alt uzay yöntemlerini birleştirerek yeni bir rassal orman algoritması geliştirmiştir. Torbalama tekniğiyle yeni veri setleri oluşturmuş ve alt uzay yöntemiyle de yeni karar ağaçları ve karar düğümleri elde edilmiştir (Breiman, 1996).

Rassal orman algoritmasında kullanılan InBag (in bag) ve OOB (out of bag) terimleri de verilerin oluşturulması aşamasında önemlidir. Algoritmanın işleyişi öğrenme veri setiyle başlar. Torbalama yöntemiyle veri setinden örneklem seçilir ve bu öğrenme verisinin $2/3$ 'ünü oluşturur. Bu oluşturulan veri seti InBag diye isimlendirilir. Bu sebeple oluşturulan her bir ağacın InBag verisi değişiklik gösterecektir. Veri setinden geriye kalan $1/3$ 'lük kısım algoritmanın tahminleme aşamasındaki başarısını görmek için OOB (out of bag) verisi olarak değerlendirilir (Sabancı, 2019).

Karar ağaçları oluşturulma aşamasında, gerçek veri setindeki n adet gözlem ile birebir aynı sayıda olacak şekilde torbalama tekniğiyle N adet alt veri seti elde edilir. Rassal Orman Algoritması aşamaları aşağıda verilmiştir (Akman, 2010; Yao vd., 2013; Yılmaz, 2014; Abellan vd., 2017; Liaw ve Wiener, 2018).

- Gerçek veri setindeki verilerin $2/3$ 'ü InBag, $1/3$ 'ü ise OOB verisi olarak seçilir.
- Sınıflandırma ve regresyon karar ağaçları InBag verileri kullanılarak yapılır. Karar ağacı oluşturulurken, karar düğümlerinde en önemli değişkenleri kullanmak yerine bunlardan toplam m adet olan bağımsız değişkenlerden p adet rastgele seçilir. p değeri, m değerinden küçük seçilmelidir. Çünkü ezberleme engellenmelidir. Rastgele seçilen bu değişkenlerden bilgi kazancı en büyük olan değer ile ağaç dalları oluşturulur.
- Rassal Orman oluşuncaya kadar önceki aşamalar tekrar edilmektedir. N adet oluşturulan her bir ağacın yaptığı tahminler birleştirilerek yeni bir tahmin yapılır. Regresyon ağaçlarında sınıflandırma için yapılan oylamaların aritmetik ortalaması alınarak son tahmin yapılırken, sınıflandırma karar ağaçları için ise oy çokluğu alan sınıf en son tahmin olarak değerlendirilir.
- Out of bag (OOB) verileri ile yapılan sınıflandırma tahminleri ise, Rassal Orman algoritmasının hata oranını hesaplamak için kullanılır. Algoritmanın performansı ölçülmüş olur.

Rassal Orman Algoritmasının yapısı Şekil 3.18’de gösterilmiştir.



Şekil 3.18. Rassal Orman Algoritmasının yapısı (Sabancı, 2019)

Rassal Orman algoritmasında karar ağaçları oluşturulurken, verileri sınıflandırmak için verilere ait özelliklerden en etkili olanlarını seçmek çok önemlidir. Böylece bir sonraki sınıflandırmanın yapılması için hangi özelliklerin önemli hangilerinin ise daha az etkili veya önemsiz olduğu tespit edilmelidir. Böylece daha etkili bir model oluşturulabilir (Cutler vd., 2011). Özelliklerin önem derecesini tespit etmek için iki farklı metot kullanılır. Bunlardan biri, algoritma yapısı oluşurken kullanılacak olan özelliklerin önem derecesini bulmaktır. Bunun için “gini temelli önem” ölçütü kullanılır. Bu ölçüte göre en önemli özellikler tespit edilir ve algoritmanın yapısının oluşturulmasında kullanılır. Diğer metot ise “doğruluk temelli önem” olarak isimlendirilir. Bu metotta rastgele orman modeli inşa edilir ve modelin performansı hesaplanır. Daha sonrasında eğitim setindeki bir parametrenin tüm sütun değerleri rasgele dağıtılır. Yeni oluşturulan veri setiyle modelin performansı tekrar ölçülür. Gerçek veri setindeki sınıflama sayısından, değiştirilmiş veri setindeki sınıflama sayısı çıkarılır ve bu her karar ağacı için yinelenir. Aradaki oluşan farkların

ortalama değeri hesaplanarak özelliğin önem derecesi bulunur. Veri setindeki tüm değişkenler için bu önem değeri hesaplanır (Akman, 2010).

Rastgele orman algoritmasının avantajları aşağıda maddeler halinde verilmiştir (Brieman, 1998; Akman, 2011; Yao vd., 2013; Yılmaz, 2014; Abellan vd., 2017; Liaw ve Wiener, 2018).

- En iyi performansı gösterecek şekilde sınıflandırmayı seçmektedir.
- Kayıp değerlere ve aykırı değerlere karşı dayanıklıdır.
- Farklı özelliklerdeki veri setleriyle uyumlu çalışabilmektedir. Özellikle küçük veri setlerinde ve değişken sayısının fazla olduğu koşullarda üstün performans gösterir.
- ROA tek bir karar ağacına göre daha iyi sonuçlar verir.
- Out of bag verileriyle yüzlerce kez çapraz doğrulama yaparak güvenilirliği artırmaktadır.
- Algoritma çok hızlı çalışmaktadır.

Dezavantajları ise şöyle sıralanabilir;

- Yapay sinir ağları ve regresyon gibi algoritmalarda ortaya çıkan doğruluk aralıkları, rastgele orman algoritmasında oluşmamaktadır.
- Bilgisayarlarda fazla bellek kullanılmasını gerektirmekte ve işlem zorluğuna neden olmaktadır.
- Çok fazla sayıda karar ağacı oluştuğu için, yapılan işlemler görsel olarak değil, renk sıcaklık haritalarıyla görülebilmektedir.

3.8 Sınıflandırma Başarım Ölçütleri

Bu bölümde farklı makine öğrenmesi tekniklerinin sınıflandırma başarımlarını analiz edebilmek için kullanılan ölçütler anlatılmaktadır.

3.8.1 Karışıklık Matrisi

Verilerin sınıflandırılması sonucu çıktıların en az iki veya daha fazla olduğu durumlarda performans ölçütü olarak kullanılabilir. Çizelge 3.3 karışık matrisini ifade etmektedir.

Çizelge 3. 3. Karışıklık matrisi

TAHMİN EDİLEN DEĞERLER	GERÇEK DEĞERLER	
	POZİTİF	NEGATİF
	P	N
P	Gerçek Pozitif (GP)	Yanlış Pozitif (YP)
N	Yanlış Negatif (YN)	Gerçek Negatif (GN)

Gerçek Pozitif (GP), Verideki alkolik birine ait bilgiyi algoritma doğru tahmin ederse. Yani “bu veri alkolizme yatkın birine aittir” diye tahmin yaparsa, buna gerçek pozitif denir. Gerçek Negatif (GN), Verilerdeki kontrol grubuna ait birini, algoritma doğru tahmin ederse, yani “bu veri alkolik olmayan birine aittir” diye tahmin yaparsa, buna gerçek negatif denir. Yanlış Pozitif (YP), veri setindeki alkolik olmayan birini algoritma sınıflandırırken, “bu kişi alkoliktir” diye tahminde bulunursa, buna yanlış pozitif denir. Yanlış Negatif (YN), verilerdeki alkolik olan birisini algoritma sınıflandırırken, “bu kişi alkolik değildir” diye tahmin yaparsa, bu durum da yanlış negatif olarak değerlendirilir.

3.8.2 Doğruluk

Doğruluk değeri, modelde doğru tahmin ettiğimiz alanların toplam veri kümesine oranı ile hesaplanmaktadır. Eşitlik (3.8) ‘de doğruluk denklemi gösterilmiştir (Powers ve Ailab, 2011).

$$\text{Doğruluk} = \frac{GP+GN}{GP+YP+GN+YN} \quad (3.8)$$

3.8.3 Kesinlik

Kesinlik, pozitif olarak tahmin edilen değerlerin gerçekten kaç adedinin pozitif olduğunu ifade eder. Eşitlik (3.9)'da kesinlik denklemi gösterilmiştir (Powers ve Ailab, 2011).

$$\text{Kesinlik} = \frac{GP}{GP+YP} \quad (3.9)$$

3.8.4 Duyarlılık

Duyarlılık ise pozitif olarak tahmin edilenlerin, işlemlerin ne kadarını pozitif olarak tahmin ettiğimizi gösteren bir metriktir (Powers ve Ailab, 2011). Eşitlik (3.10)'da duyarlılık denklemi verilmiştir.

$$\text{Duyarlılık} = \frac{GP}{GP+YN} \quad (3.10)$$

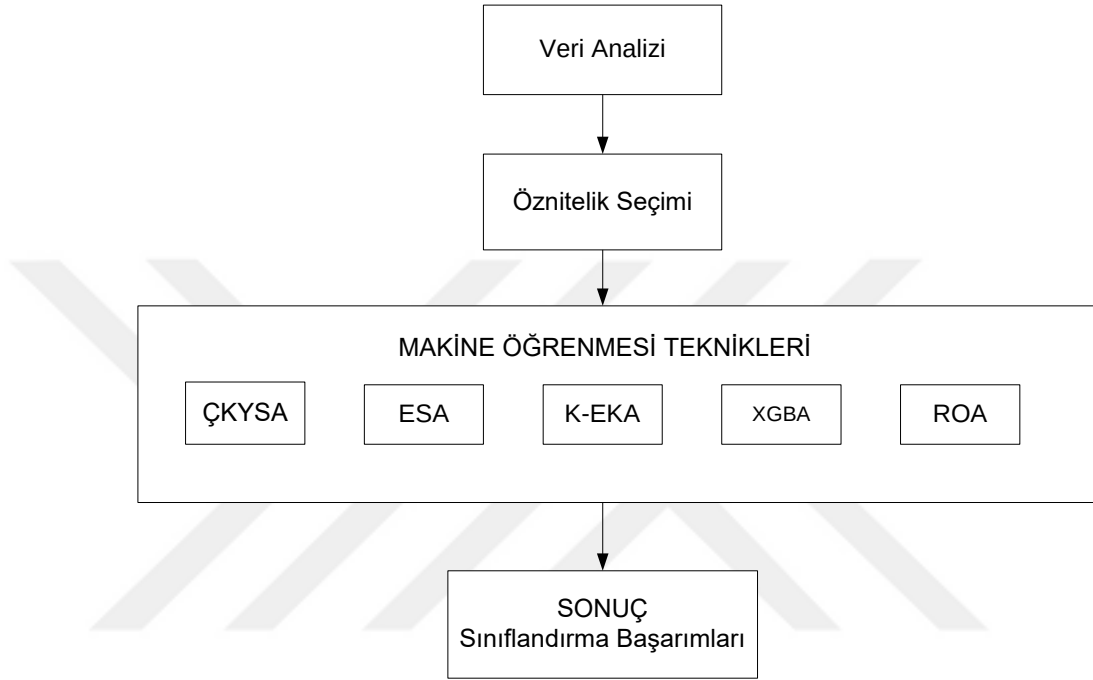
3.8.5 F1-Score

Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını göstermektedir. Harmonik ortalama kullanılmasının sebebi çok uç durumları göz önünde bulundurma zorunluluğudur. Basit bir aritmetik ortalama bizi yanıltabilir (Gülcan, 2020, Powers ve Ailab, 2011). Eşitlik (3.11)'de ifade verilmiştir.

$$\text{F1 - Skor} = 2 * \frac{\text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.11)$$

4. BULGULAR

Çalışma temel olarak EEG işaretleri kullanılarak veri analizi, öz nitelik seçimi, makine öğrenmesi tekniklerinin sınıflandırma başarımları ve çalışma sürelerinin değerlendirilmesini içermektedir. Çalışmanın blok diyagramı Şekil 4,1’de verilmiştir.



Şekil 4.1. Çalışmanın blok diyagramı

Çalışmalar, işlemcisi Intel core-i7, 6500U CPU 2.50 GHz, 8GB Ram’i olan dizüstü bilgisayarda gerçekleştirilmiştir. Kullanılan kütüphaneler ve versiyonları Çizelge 4.1’de sunulmuştur.

Çizelge 4.1. Kullanılan kütüphane ve versiyonları

Paket	Versiyon
pandas	0.24.2
numpy	1.16.5
Matplotlib	3.1.1
Keras	2.3.1
graphviz	0.14
imagesize	1.1.0
ipykernel	5.1.2
Keras-Applications	1.0.8
Keras-Preprocessing	1.1.0
Mlxtend	0.17.2
Plotly	4.5.4
Scikit-learn	0.21.3
Seaborn	0.9.0
Spyder	3.3.6
Tensorflow	2.0.0
Xgboost	1.1.0

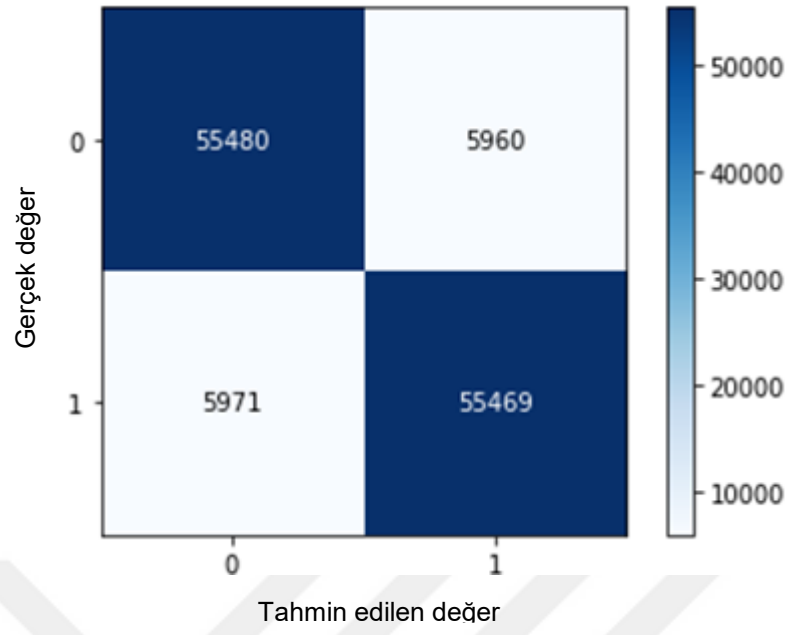
4.1 Çok Katmanlı Yapay Sinir Ağı Sınıflandırma Başarım Değerlendirmesi

Çok Katmanlı Yapay Sinir Ağı (ÇKYSA) 4 katmandan oluşmaktadır. İlk 3 katmanda ReLu, çıkış katmanında ise sigmoid fonksiyonu kullanılmıştır. Ağda kayıp fonksiyonu olarak ortalama karesel hata değeri kullanılmıştır. Eğitim süresi 17dk 16sn, test süresi 1,32sn sürmüştür. Ortalama karesel hata değeri ise 0.0870 olarak hesaplanmıştır. Çizelge 4.2’de başarımlar sonuçları verilmiştir.

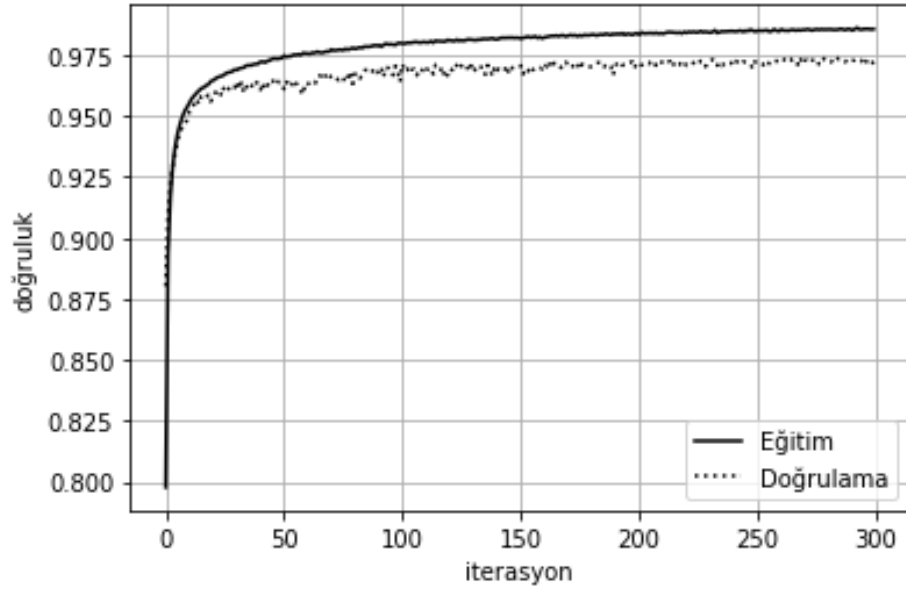
Çizelge 4. 2. ÇKYSA için başarımlar sonuçları

Sınıf	Kesinlik	Duyarlılık	F1-Skor	Doğruluk
a-0	%90	%90	%90	%90
c-1	%90	%90	%90	

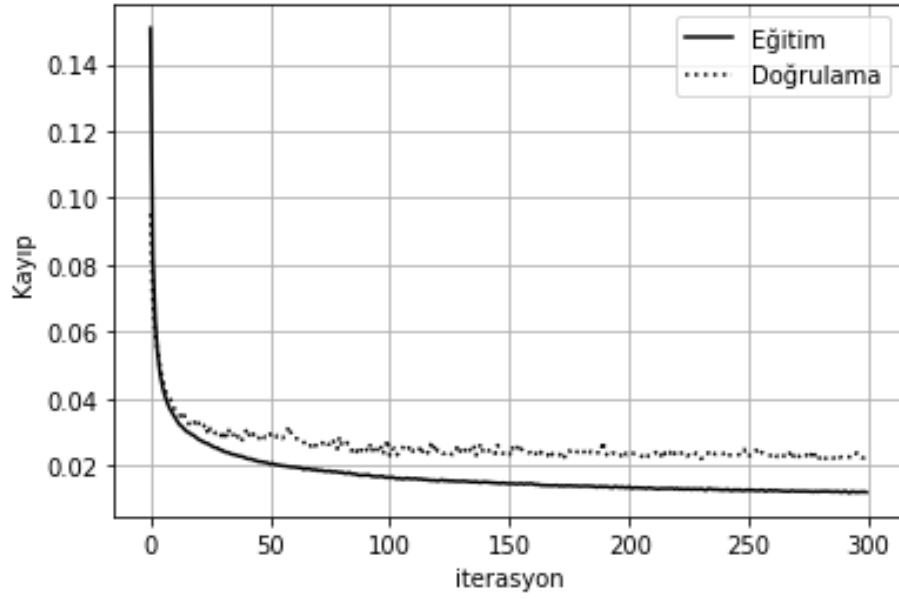
Şekil 4.2’de ÇKYSA için Karışıklık matrisi verilmiştir. Şekil 4.3 ve Şekil 4.4’te ise modelin doğruluk ve kayıp eğrileri görülmektedir.



Şekil 4.2. ÇKYSA için karışıklık matrisi



Şekil 4.3. ÇKYSA için modelin doğruluk eğrisi



Şekil 4.4. ÇKYSA için modelin kayıp eğrisi

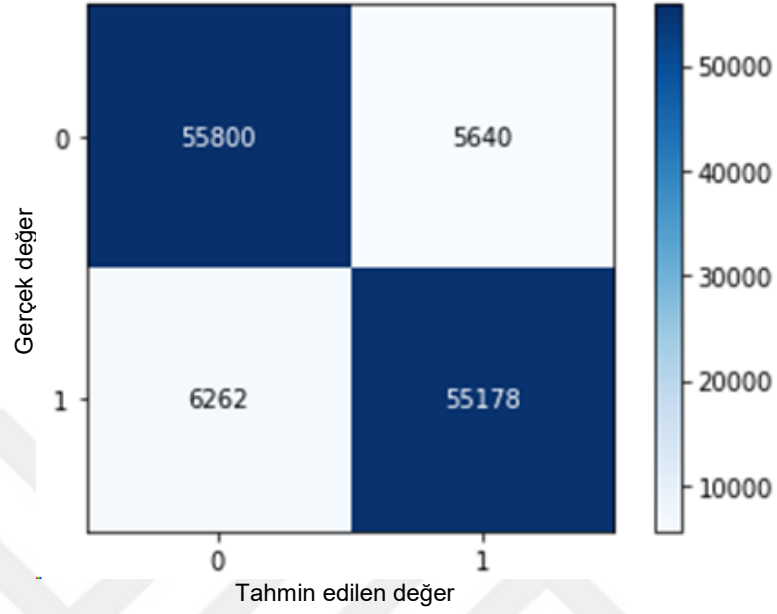
4.2 Evrişimsel Sinir Ağı Sınıflandırma Başarım Değerlendirmesi

1D-Evrişimsel Sinir Ağı (ESA) yapısı kullanılmıştır. Giriş ve ara katmanlarda Relu, çıkışta sigmoid fonsiyonu tercih edilmiştir. Verilerin özelliklerini saptamak için evrişim katmanı, uygunluk ve ağırlık sayısı için ise maksimum havuzlama katmanı kullanılmıştır. Ağda kayıp fonskiyonu olarak ortalama karesel hata değerinden faydalanılmıştır. Ortalama karesel hata değeri ise 0.0864 olarak hesaplanmıştır. Eğitim süresi 21 dk 53sn, test süresi 1.68sn sürmüştür. Çizelge 4.3’de başarımları verilmiştir.

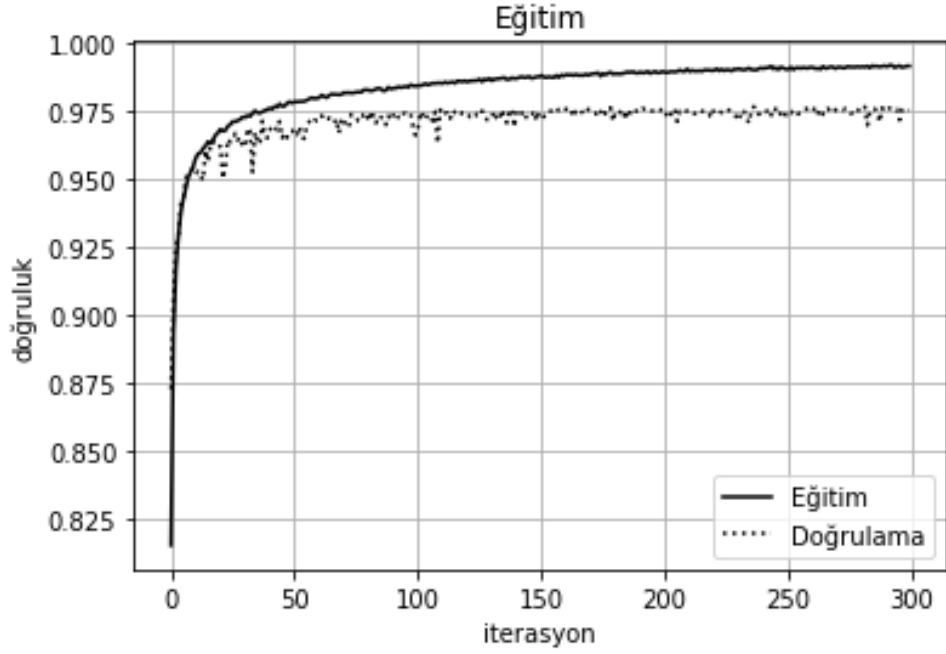
Çizelge 4. 3. ESA için başarımları sonuçları

Sınıf	Kesinlik	Duyarlılık	F1-Skor	Doğruluk
a -0	%90	%91	%90	%90
c -1	%91	%90	%90	

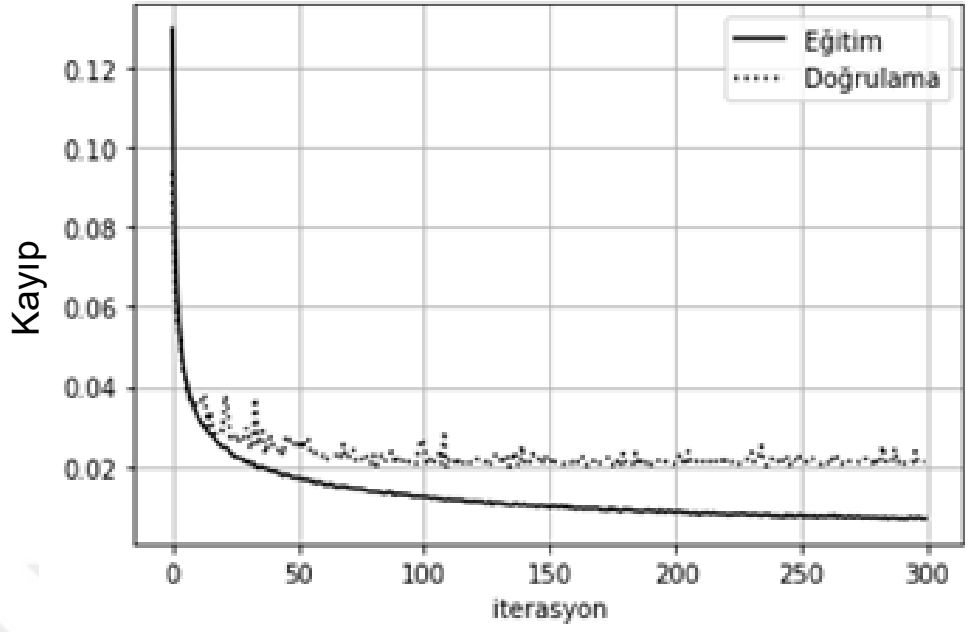
Şekil 4.5’de karışıklık matrisi verilmiştir. Şekil 4.6 ve Şekil 4.7’de ESA için modelin doğruluk ve kayıp eğrisi gösterilmiştir.



Şekil 4.5. ESA için karışıklık matrisi



Şekil 4.6. ESA için modelin doğruluk eğrisi



Şekil 4.7. ESA için modelin kayıp eğrisi

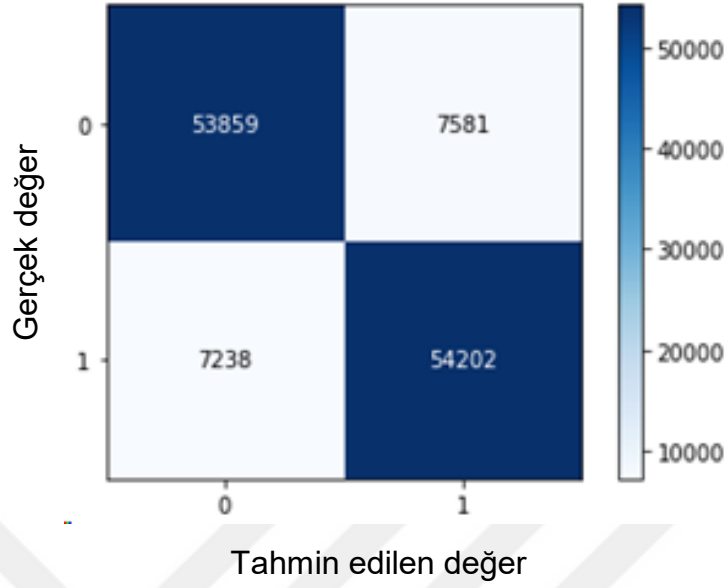
4.3 XGBoost Sınıflandırma Başarım Değerlendirmesi

Karar Ağacı temelli XGBoost algoritma (XGBA) yapısı kullanılmıştır. Derinlik 50 olarak alınmıştır. Ortalama karesel hata değeri ise 0.120 olarak hesaplanmıştır. Eğitim süresi 3 dk 65sn, test süresi 1.44 sn sürmüştür. Çizelge 4.4'te XGBoost için başarımlar verilmiştir.

Çizelge 4. 4. XGBoost için başarımlar

Sınıf	Kesinlik	Duyarlılık	F1-Skor	Doğruluk
a - 0	%88	%88	%88	%88
c -1	%88	%88	%88	

Şekil 4.8’de XGBoost için karışıklık matrisi verilmiştir.



Şekil 4.8. XGBoost için karışıklık matrisi

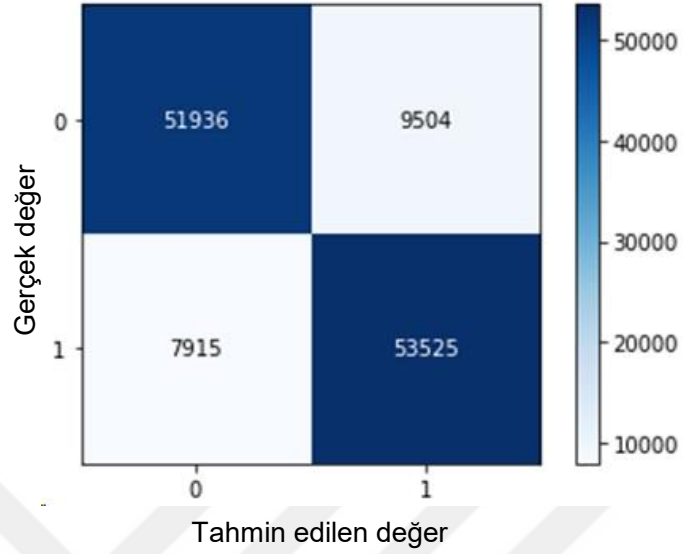
4.4 Rassal Orman Algoritması Sınıflandırma Başarım Değerlendirmesi

Rassal Orman Algoritması (ROA) için ağaç sayısı 400 olarak atanmıştır. Yapılan benzetimler sonucu ortalama karesel hata değeri ise 0.141 olarak hesaplanmıştır. Eğitim süresi 9 dk 34sn, test süresi 20.59 sn sürmüştür. Çizelge 4.5’te Rassal Orman Algoritması için başarımları verilmiştir.

Çizelge 4. 5. Rassal Orman algoritması için başarımları

Sınıf	Kesinlik	Duyarlılık	F1-Skor	Doğruluk
a -0	%87	%85	%86	%86
c -1	%85	%87	%86	

Şekil 4.9’da karışıklık matrisi gösterilmiştir.



Şekil 4.9. ROA için karışıklık matrisi

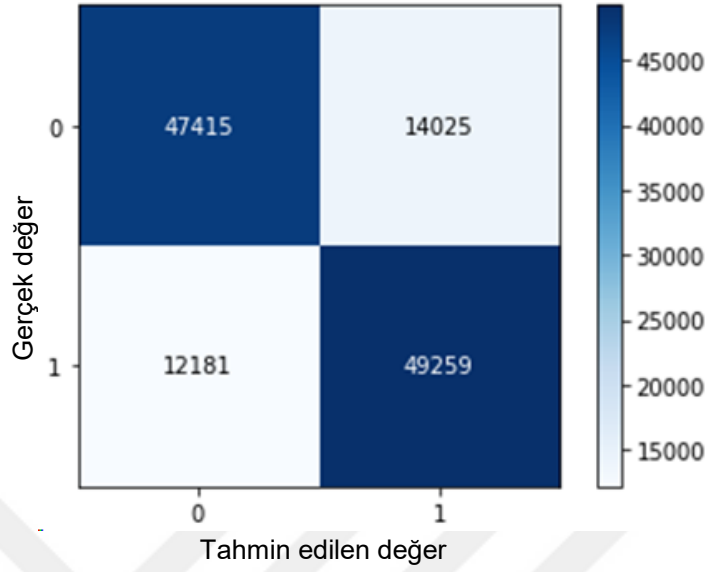
4.5 K-En Yakın Komşuluk Sınıflandırma Başarım Değerlendirmesi

K En Yakın Komşuluk (K-EKA) algoritması için komşu sayısı 5 olarak atanmıştır. Ortalama karesel hata değeri 0.2132 olarak hesaplanmıştır. Eğitim süresi 0.554 sn, test süresi 13dk 69 sn sürmüştür. Çizelge 4.6’da başarımları sunulmuştur.

Çizelge 4. 6. K-en yakın komşu algoritması için başarımları

	Kesinlik	Duyarlılık	F1-Skor	Doğruluk
a-0	%80	%77	%78	%79
c-1	%78	%80	%79	

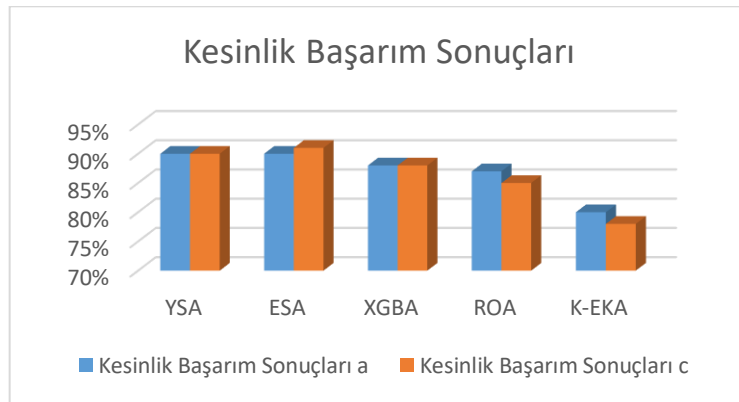
Şekil 4.10’da karışıklık matrisi verilmiştir.



Şekil 4.10. K-EKA için karışıklık matrisi

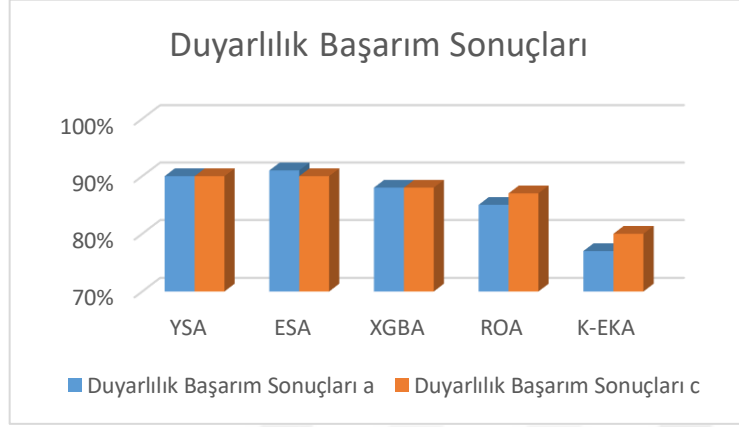
4.6 Sınıflandırma Başarımlarının Karşılaştırılması

Tez çalışmasında YSA, ESA, XGBA, ROA ve K-EKA olarak beş farklı makine öğrenmesi teknikleri kullanılmıştır. Şekil 4.11’de kullanılan algoritmaların kesinlik başarımları sonuçları verilmiştir. Kesinlik başarımları en yüksek ESA ve YSA olduğu görülmektedir.



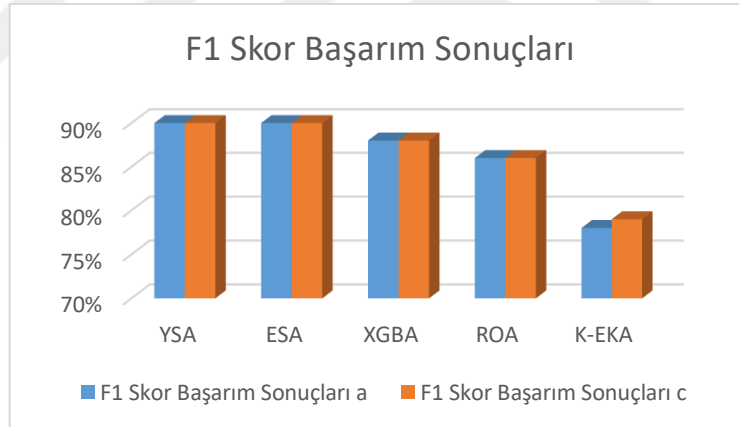
Şekil 4.11. Kesinlik başarımları sonuçları

Şekil 4.12’de kullanılan algoritmaların duyarlılık başarımları gösterilmiştir. YSA ve ESA duyarlılık başarımları en yüksektir.



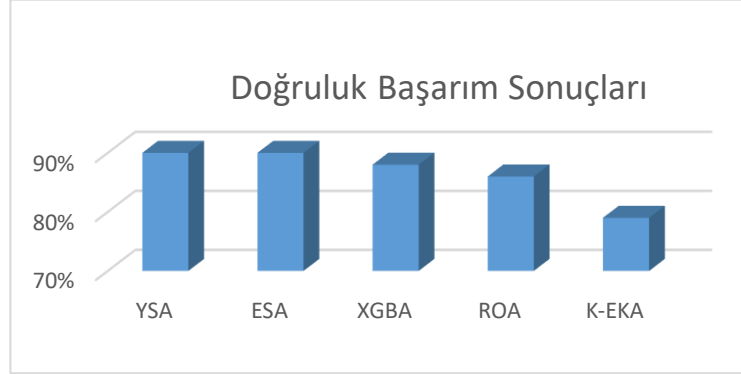
Şekil 4.12. Duyarlılık başarımları

Şekil 4.13’de kullanılan algoritmaların F1 Skor başarımları gösterilmiştir. YSA ve ESA başarımları en yüksektir.



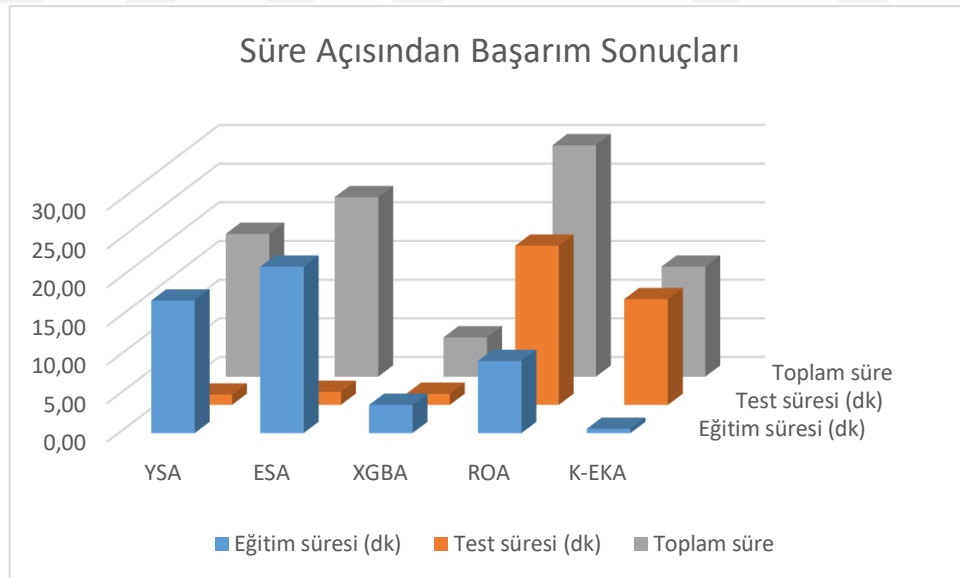
Şekil 4.13. F1 Skor başarımları

Şekil 4.14’de doğruluk başarımları verilmiştir. Şekil 4.14’e bakıldığında YSA ve ESA doğruluk açısından diğerlerine göre %90 oranında başarımları göstermiştir. En düşük başarımları ise K-EKA göstermiştir.



Şekil 4.14. Doğruluk başarımları

Şekil 4.15’de algoritmaların eğitim ve test süreleri açısından başarımları verilmiştir. Eğitim süreleri açısından analiz edildiğinde, ESA eğitim süresi en yüksek, K-EKA eğitim süresi en düşük çıkmıştır. Test süreleri açısından ise ROA en fazla süreyi içermektedir.



Şekil 4.15. Süre açısından başarımları

5.SONUÇ VE ÖNERİLER

Tez çalışmasında EEG işaretleri kullanılarak bağımlılığa yatkınlığın farklı makine öğrenmesi teknikleri ile sınıflandırma başarımları ve çalışma süreleri değerlendirilmiştir. EEG verileri alkolizme olan yatkınlığın EEG ile ilişkisini içermektedir. Alkolik ve kontrol olmak üzere iki sınıftan oluşmaktadır. Çalışmada ilk olarak veri analiz edilmiştir. Dağınık ve çok sayıda olan csv dosyaları tek bir csv dosyasına dönüştürülerek daha düzenli bir veriseti elde edilmiştir. Veriler arasındaki korelasyon incelenmiş, öznitelik seçimi yapılmıştır. L2 normalizasyon kullanılmıştır. Öznitelik seçiminde özyinelemeli öznitelik seçimi kullanılmıştır. Kullanıma hazır olan veri için beş farklı makine öğrenme tekniği kullanılarak sınıflandırma yapılmıştır. Kullanılan teknikler; Çok Katmanlı Yapay Sinir Ağları (ÇKYSA), Evrişimsel Sinir Ağı (ESA), K-En Yakın Komşu Algoritması (K-EKA), XGBoost Algoritması (XGBA), Rassal Orman Algoritması (ROA)'dır. ÇKYSA, eğitim süresi 17dk 16sn, test süresi 1,32sn sürmüştür. Ortalama karesel hata değeri ise 0.0870 olarak hesaplanmıştır. Doğruluk oranı %90'dır. ESA için ortalama karesel hata değeri ise 0.0864 olarak hesaplanmıştır. Eğitim süresi 21 dk 53sn, test süresi 1.68sn sürmüştür. Doğruluk oranı %90 elde edilmiştir. XGBA için ortalama karesel hata değeri ise 0.120 olarak hesaplanmıştır. Eğitim süresi 3 dk 65sn, test süresi 1.44 sn sürmüştür. Doğruluk oranı %88 olarak elde edilmiştir. ROA, ortalama karesel hata değeri ise 0.141 olarak hesaplanmıştır. Eğitim süresi 9 dk 34sn, test süresi 20.59 sn sürmüştür. Doğruluk oranı %86 çıkmıştır. K-EKA, ortalama karesel hata değeri 0.2132 olarak hesaplanmıştır. Eğitim süresi 0.554 sn, test süresi 13dk 69 sn sürmüştür. Doğruluk oranı %79 elde edilmiştir. Sonuçlara bakıldığında ortalama karesel hata en düşük ESA'da elde edilmiştir. Eğitim süresi en uzun ESA'da, en kısa ise K-EKA'da elde edilmiştir. Test süresi en fazla süren ROA olurken, en az süren ise ÇKYSA'dır. Toplam süre açısından bakıldığında ise en hızlı XGBA çıkmıştır. Doğruluk oranı açısından en iyi sonuçları ÇKYSA ve ESA vermiştir.

Sonraki alıřmalar olarak farklı znitelik ıkarım teknikleri uygulanabilir, tasarlanan modellerdeki hiper parametreler deęiřtirilerek bařarımları daha da artırmaya ynelik alıřmalar yapılabilir.

EEG sinyalleri zerinde yapılan alıřmalar incelendięinde sıklıkla hastalıęı daha doęru tespit etmek iin alıřmalar yrtlmřtr. Bu alıřmayı dięer alıřmalardan ayıran fark ise baęımlılık zerine bir arařtırma gerekleřtirmemizdir. Literatre bakıldıęında zellikle lkemizde bununla ilgili yapılan alıřma sayısı yok denecek kadar azdır. Ama, bu alıřmanın dięer baęımlılıklarla ilgili yapılacak olan alıřmalara rnek teřkil edebilmesidir. İnsan beyni grdę ve duyduęu her řeyden etkilenmektedir. Baęımlılık tedavisinde genetik yatkınlıkların EEG ile iliřkisi belirlenebilirse hangi sinyallerden beynin olumlu etkilendięi tespit edilebilir ve baęımlı olan ya da baęımlı olmaya aday bireyleri baęımlılık yaratacak kt alıřkanlıklardan koruyabilir veya iyileřtirebiliriz.

6.KAYNAKLAR

- Abellan J, Mantas CJ, Castellano JG (2017) “A Random Forest Approach Using Imprecise Probabilities” *Knowledge-Based Systems*, 134, 72-84.
- Akman M (2010) *Veri Madenciliğine Genel Bakış ve Random Forest Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*, Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara.
- Akman M, Genç Y, Ankaralı H (2011) “Random Forests Yöntemi ve Sağlık alanında bir uygulama”. *Türkiye Klinikleri*, 30:1,36-48.
- Bavkar S, Iyer B, Deosarkar S (2019) “Detection of Alcoholism: An EEG Hybrid Features and Ensemble Subspace K-NN Based Approach” *International Conference on Distributed Computing and Internet Technology, Lecture Notes in Computer Science*, vol 11319, 161-168.
- Bhatia N, Vandana A (2010) “Survey of nearest neighbor techniques”, *International Journal of Computer Science and Information Security*, 8(2):302-305.
- Breiman L (1996) “Bagging Predictors” *Machine Learning* 24, 123–140.
- Brownlee J (2020) Recursive Feature Elimination (RFE) for Feature Selection in Python, <https://machinelearningmastery.com/rfe-feature-selection-in-python/>, (28 Haziran 2020).
- Calp MH (2019) “İşletmeler için personel yemek talep miktarının yapay sinir ağları kullanılarak tahmin edilmesi” *Politeknik dergisi*, 22(3).
- Carvalho DR, Freitas AA (2004) “A hybrid decision tree/genetic algorithm method for Data mining” *J Inf Sci.*, 163(1-3):13-35.
- Chambon S, Galtier MN, Arnal PJ, Wainrib G, Gramfort A (2018) “A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series.”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26:4, 758-769.
- Choubey RN, Amar L, Khare S (2019) “Internet traffic classifier using artificial neural network and 1D-CNN”, *International Conference on Information Technology (ICIT)*, 291-296, Bhubaneswar, India.
- Cutler A, Cutler DR and Stevens JR (2011) “Random Forests” *Ensemble Machine Learning*, Publisher :Springer, 157-176.
- Cunningham P and Delany SJ (2007) “K-neighbor classifiers”, *J Multiple Classifier Systems*, 34(8):1-17.

- Dandıl E, Çevik K (2011) “Yapay sinir ağları için net platformunda görsel bir eğitim yazılımının geliştirilmesi.” Bilişim Teknolojileri Dergisi, 5(1), 19-28.
- Edenberg HJ, Gelernter J, Agrawal A (2019) “Genetics of Alcoholism”, Current Psychiatry Reports , 21:26.
- Elesan S (2019) Veri madenciliğinde farklı karar ağaçları ve k-en yakın komşuluk yöntemlerinin incelenmesi: Kadın hastalıkları ve doğum verisinde bir uygulama, Yüzüncü Yıl Üniversitesi, Sağlık Bilimleri Enstitüsü, Doktora Tezi, Van.
- Elmas Ç (2016) Yapay Zeka Uygulamaları, 3.Baskı, Ankara, Seçkin Yayıncılık.
- Ersöz A, Özşen S (2011) Uyku EEG Sinyalinin Yapay Sinir Ağ Modeli İle Sınıflandırılması, Elektrik-Elektronik ve Bilgisayar Sempozyumu, 298-301, Elazığ.
- Fedjaev J (2017) Decoding EEG Brain Signals Using Recurrent Neural Networks, Master Thesis, Neuroscientific System Theory Technische Universität München,, Münih.
- Gadade YA (2019) limitlessdatascience.wordpress.com/2019/04/16/normalisation-l1-l2-norms/. (19 Haziran 2020).
- Gajic D, Djurovic Z, Gennaro SD, Gustafsson F (2014) ”Classification of eeg signals for detection of epileptic seizure based on wavelets and statistical pattern recognition” Biomedical Engineering: Applications, Basis and Communications, 26(02):145021.
- Ge S, Yang Q, Wang R, Lin P, Gao J, Leng Y, Yang Y, Wang H (2017) “A Brain-Computer Interface Based on a Few-Channel EEG-fNIRS Bimodal System” IEEE Access, vol 5, pp 208-218.
- Ghoury S (2019) Konvolüsyonel sinir ağlarını kullanarak üzüm ve üzüm yapraklarının hastalıklarının tespit edilmesi, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Gül E (2018) EEG sinyallerinin wavelet yöntemiyle dönüştürülerek yapay sinir ağları ile sınıflandırılması, Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, Sağlık Bilimleri Enstitüsü, Eskişehir.
- Gülcan O (2020) <https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1l%C4%B1k-recall-ya-da-f1-score-300c925feb38>, (24 Haziran 2020).
- Güler İ, Gökçil Z and Gülbandılar E (2009) “Evaluating of traumatic brain injuries using artificial neural networks”, Expert Systems with Applications, 36(7), 10424-10427.

- Gürsoy UTŞ (2009). Veri madenciliği ve Bilgi Keşfi: Pegem Akademi.
- Ho TK (1998) “The Random Subspace Method for Constructing Decision Forests” IEEE Transactions on Pattern Analysis and Machine Intelligence, 832-844.
- Iasemidis LD (2003). “Epileptic seizure prediction and control”, IEEE Transactions on Biomedical Engineering, 50(5),549-558.
- Işık G and Artuner H (2018) “A dataset for Turkish dialect recognition and classification with deep learning” 26th Signal Processing and Communications Applications Conference, 1-4, İzmir.
- İmak A (2019) Epileptik EEG sinyallerinin makine öğrenmesi yöntemleri ile sınıflandırılması, Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
- İnik Ö, Ülker E (2017) “Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri”, Gaziosmanpaşa bilimsel araştırma dergisi, 6(3), 85-104.
- İpek B, (2018) Eeg sinyallerinin epileptik rahatsızlıkların teşhisi için konvolüsyonel sinir ağları ve destek vektör makineleri ile tasnif edilmesi, Yüksek Lisans Tezi, Karatay Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- Jarrett K, Kavukcuoglu K, Ranzato M, LeChun Y (2009) “What is the best multi-stage architecture for object recognition? ” IEEE 12th International Conference on Computer Vision, pp. 2146-2153, Kyoto.
- Kalaivani M, Kalaivani V and Devi VA (2014) “Analysis of EEG Signal for the Detection of Brain Abnormalities.”, International Journal of Computer applications (IJCA), 2, 1-6.
- Kantardzic M (2011) Data mining: concepts, models, methods, and algorithms: John Wiley & Sons.
- Köktürk F (2012) K en yakın komşuluk, yapay sinir ağları ve karar ağaçları yöntemlerinin sınıflandırma başarılarının karşılaştırılması, Doktora Tezi, Bülent Ecevit üniversitesi Sağlık Bilimleri Enstitüsü, Zonguldak.
- Kruizinga H (2016) Your 5 brainwaves: Delta, theta, alpha, beta and gamma., <https://lucid.me/blog/5-brainwaves-delta-theta-alpha-beta-gamma/>(25 Haziran 2020).
- Lakshmanan S (2019) How, When, and Why Should You Normalize / Standardize / Rescale Your Data? <https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff>, (29 Haziran 2020)

- Laribi P (2018) Genetik algoritma ve k-en yakın komşu kullanarak metin belgelerinin sınıflandırılması, Yüksek Lisans Tezi, Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van.
- Liaw A, Wiener M (2018) The R Project for Statistical Computing. <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>, (23 Haziran 2020).
- Mathworks, (2018a) Convolutional Neural Network. <https://uk.mathworks.com/solutions/deep-learning/convolutional-neural-network.html> (20 Haziran 2020)
- Mishra M (2019) Convolutional Neural Networks, Explained [online], <https://www.datascience.com/blog/convolutional-neural-network>, (10 Haziran 2020).
- Morde V (2020) <https://towardsdatascience.com/https-medium-com-vishal-morde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>, (22 Haziran 2020).
- Murthy KS (1998) “Automatic Construction of Decision Tree from Data: A Multi-Disciplinary Survey”, Data Mining and Knowledge Discovery, 2: (345-389).
- Murugesan M, Sukanesh R (2009) “Automated detection of brain tumor in EEG signals using artificial neural networks”, International Conference on Advantages in Computing, Control and Telecommunication Technologies, December 2009, 284-288.
- Næs T, Knut K, Tomas I, Charles M (1993). “Artificial neural networks in multivariate calibration”. Journal of Near Infrared Spectroscopy, 1, 1-11.
- Schomer DL and Silva FH (2017) Electroencephalography: Basic principles, clinical applications and related fields. Publisher: Oxford University Press, Print ISBN-13: 9780190228484.
- O’shea A, (2017) “Neonatal seizure detection using convolutional neural networks.”, arXiv preprint arXiv: 1709.05849,2017.
- Öztemel E (2006) Yapay Sinir Ağlatı. Papatya Yayınları, İstanbul.
- Peng Z, Huang Q, Han Y (2019) "Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm," IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT), Jinan, China, 2019, pp. 168-172, doi: 10.1109/ICAIT.2019.8935894.
- Powers W, Ailab (2008) Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. 2. 2229-3981. 10.9735/2229-3981.

- Ramazanov S (2020) Türkiye’de fertlerin elektronik ticaret kullanım sürelerinin destek vektör makineleri ve karar ağaçları yöntemleri ile karşılaştırmalı analizi, Kafkas Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Kars.
- Rieg T, Frick J, Hitzler M, Buettner R (2019), “High-performance detection of alcoholism by unfolding the amalgamated EEG spectra using the Random Forests method”, Proceedings of the 52nd Hawaii International Conference on Systems Sciences, 3769-3777.
- Sabancı D (2019) Rastgele Orman Yaklaşımı Kullanılarak Çok Değişkenli Uyumlu Regresyon Şeritlerinde Model Seçimi, Doktora Tezi, Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü, Samsun.
- Sharanreddy M, Kulkarni P (2013) “Automated EEG signal analysis for identification of epilepsy seizures and brain tumour.”, Journal of Medical Engineering & Technology, 37(8), 511-519.
- Şen Z (2004) Yapay sinir ağları ilkeleri. İstanbul: Su Vakfı Yayınları.
- Tosun M (2004). İnholasyon anestezisinde sevofluran oranının neuro-fuzzy sistemi ile kontrolü. Sakarya Üniversitesi, Fen bilimleri enstitüsü, Sakarya.
- Traore B, Kamsu F, Tangarab F (2018) “Deep convolution neural network for image recognition”, Ecological informatics 48, 257-268.
- Tuncer E (2015) Uyku Evrelemesinde Çeşitli Dalgacık ve Sınıflandırıcıların Performans Analizi, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.
- Tüfekçi M, Karpat F (2019), “Derin öğrenme mimarilerinden konvolüsyonel sinir ağları üzerinde görüntü işleme- sınıflandırma kabiliyetinin artırılmasına yönelik yapılan çalışmaların incelenmesi” International congress on human- computer interaction, optimization and robotic applications, 5-7 temmuz, Nevşehir.
- Tülay EE (2009) Beyin elektriksel aktivitesinin ölçümü ve sinyal analizi, Yüksek lisans tezi, İstanbul Kültür Üniversitesi Fen bilimleri enstitüsü, İstanbul.
- UCI, veritabanı, <https://archive.ics.uci.edu/ml/datasets/eeg+database>, erişim tarihi: 10 haziran 2020.
- Vieira S, Pinaya WHL, Mechelli A (2017) “Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications.”, Neuroscience& Biobehavioral Reviews, 74:58-75.
- Yao D, Yang J, Zhan X (2013) “A Novel Method for Disease Prediction: Hybrid of Random Forest and Multivariate Adaptive Regression Splines” Journal of Computers, 8:1, 170-177.

- Yılmaz H (2014) Random Forest Yönteminde Kayıp Veri Probleminin İncelenmesi ve Sağlık Alanında Bir Uygulama. Yüksek Lisans Tezi, Osmangazi Üniversitesi Sağlık Bilimleri Enstitüsü, Eskişehir.
- Wang SH, Muhammad K, Hong J, Sangalah AK, Zhang YD (2020) “Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization”, Neural Computing and Applications, 32:665-680.
- Zhang X, Nguyen H, Bui XN (2020) “Novel Soft Computing Model for Predicting Blast-Induced Ground Vibration in Open-Pit Mines Based on Particle Swarm Optimization and XGBoost” Natural Resources Research, Vol.29, No.2.
- Zhang A, Lipton Z, Li M, Smola AJ (2018) Dive into Deep Learning., <http://en.diveintodeeplearning.org>, (12 Haziran 2020).
- Zhao M, Tang B, Deng L, Pecht M (2020) “Multiple wavelet regularized deep residual networks for fault diagnosis” Measurement, Vol 152, ISSN 0263-2241.
- Zhou J, Li E, Wang M, Chen X, Shi X, Jiang L (2019) “Feasibility of Stochastic Gradient Boosting Approach for Evaluating Seismic Liquefaction Potential Based on SPT and CPT Case Histories”, Journal of Performance of Constructed Facilities. 33. 04019024. 10.1061/(ASCE)CF.1943-5509.0001292.

7. ÖZGEÇMİŞ

Adı Soyadı : Veysel YARĞI

Doğum Yeri ve Tarihi : Diyarbakır 27/04/1988

Lisans Üniversite : Dicle Üniversitesi

Elektronik posta : veyselyargi@yandex.com

İletişim Adresi : TCDD Genel Müdürlüğü Hipodrom Cad.
Anafartalar Mah. No:3 Altındağ/ ANKARA

Yayın Listesi : -