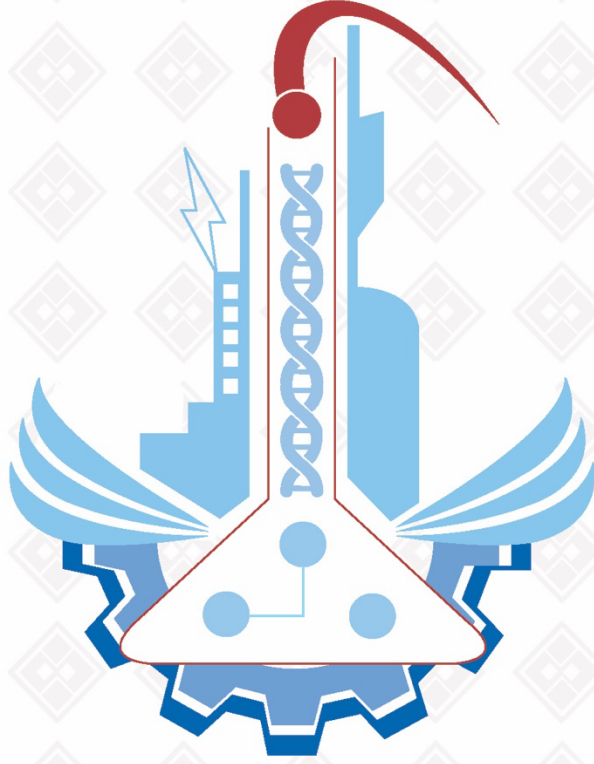


ISSN: 1308-9064

Cilt: 34 / Sayı: 1 / Yıl: 2022

# FIRAT ÜNİVERSİTESİ FEN BİLİMLERİ DERGİSİ





**Fırat Üniversitesi**  
**Fen Bilimleri Dergisi (FBD)**  
**Cilt:34, Sayı:1, Yıl:2022**  
**ISSN: 1308 - 9064**

### **İletişim**

Fırat Üniversitesi  
Fen Bilimleri Enstitüsü  
23119, ELAZIĞ

Tel: 0 424 212 27 07  
Faks: 0 424 236 99 55  
e-posta: [fenbilim@firat.edu.tr](mailto:fenbilim@firat.edu.tr)

**Bu dergi yılda 2 kez ve hakemli olarak yayımlanır.**

***Derginin Eski Adı: Fırat Üniversitesi Fen ve Mühendislik Bilimleri Dergisi***  
***Derginin Eski ISSN Numarası: 1300-2708***

*<http://dergipark.gov.tr/fufbd>*





**Fırat Üniversitesi**  
**Fen Bilimleri Dergisi (FBD)**

**Sahibi**

**Prof. Dr. Fahrettin GÖKTAŞ**  
Fırat Üniversitesi Rektörü

**Genel Yayın Yönetmeni**

**Prof. Dr. Kürşat Esat ALYAMAÇ**  
F.Ü. Fen Bilimleri Enstitüsü Müdürü

**Baş Editör**

**Doç. Dr. Fatih ÖZKAYNAK**  
Fırat Üniversitesi Enstitüsü Müdür Yrd.

**Editör**

**Doç. Dr. Emrah YILMAZ**  
Fırat Üniversitesi Enstitüsü Müdür Yrd.

**DANIŞMA KURULU**

<b>Leyla AÇIK</b>	Gazi Ü., Biyoteknoloji	<b>Turgay SEÇKİN</b>	İnönü Ü., Kimya
<b>Haluk KORALAY</b>	Gazi Ü., Fizik	<b>İbrahim KARATAŞ</b>	Selçuk Ü., Kimya
<b>Mahmut ELP</b>	Kastamonu Ü., Su Ürünleri	<b>Zihni DEMİRBAĞ</b>	K.T.Ü., Biyoloji
<b>Yusuf AYVAZ</b>	S.Demirel Ü., Biyoloji.	<b>Seyfettin ÇAKMAK</b>	S.Demirel Ü., Fizik
<b>Mikail ET</b>	Fırat Ü., Matematik	<b>Orhan ERMAN</b>	Fırat Ü., Biyoloji
<b>Hikmet GEÇKİL</b>	İnönü Ü., Biyoloji	<b>Ömer AKIN</b>	T.O.B.B. Matematik
<b>Kenan GÜLLÜ</b>	Muğla S.K.Ü. Su Ürünleri	<b>Sinan AKPINAR</b>	Fırat Ü., Fizik
<b>Mehmet KAYA</b>	Fırat Ü., Kimya	<b>Ahmet ŞAHİN</b>	Erciyes Ü., Biyoloji
<b>İlhami KIZIROĞLU</b>	Hacettepe Ü., Biyoloji	<b>Kadir DEMİRELLİ</b>	Fırat Ü., Kimya
<b>S.Güneş KÜRKÇÜOĞLU</b>	Osmangazi Ü., Fizik	<b>Kenan GÜLLÜ</b>	Muğla Ü., Su Ürünleri
<b>Dursun ŞEN</b>	Fırat Ü., Su Ürünleri	<b>Vedat ASİL</b>	Fırat Ü., Matematik
<b>Serdar SALMAN</b>	Marmara Ü. Metal. Müh.	<b>Ertan GOKALP</b>	K.T.Ü. Harita. Müh.
<b>İzzet KARA</b>	Pamukkale Ü., Fizik	<b>Ekrem ARTUNÇ</b>	S.Demirel Ü., Fizik



**Fırat Üniversitesi Fen Bilimleri Dergisi (FBD)**  
**34-1, 2022**

**İÇİNDEKİLER / CONTENTS**

1. **Bazı Karbonil Bileşikleri ile Kitosan Aşı Kopolimerlerinin Sentezi ve Karakterizasyonu**  
*Synthesis and Characterization of Chitosan Graft Copolymers with Some Carbonyl Compounds*  
Rüveyda Şener Gürsoy, Uğursoy Olgun ..... 1-10
2. **Bisfenol A (BFA) ve Halk Sağlığı Açısından Önemi**  
*Bisfenol A (BFA) and Its Importance in Terms of Public Health*  
Sait TAN, Pelin DEMİR, Ali ARSLAN ..... 11-17
3. **Türkçe Ses Kayıt Verilerinin CountVectorizer ve TF-IDFVectorizer Yöntemleri ile BERT Modelleri Olarak Google Colab Platformunda ve RapidMiner'da Makine Öğrenmesi Algoritmalarıyla Analizi**  
*Analysis of Turkish Voice Recordings Data with CountVectorizer and TF-IDF Vectorization Methods as BERT Models on Google Colab Platform and RapidMiner with Machine Learning Algorithms*  
Abdülkadir TEPECİK, Engin DEMİR ..... 19-29
4. **Siber Saldırıları için Rastgele Orman Algoritması Kullanılarak Öznitelik Seçimi**  
*Feature Selection Using Random Forest Algorithm to Cyber Attack*  
Abdulkadir BİLEN, Ahmet Bedri ÖZER ..... 31-37
5. **Yüz Tanıma Sistemlerinde Kullanılan ESA, YGH-DVM ve DSA Algoritmalarının Performans Testleri**  
*Performance Tests of ESA, YGH-DVM and DSA Algorithms Used in Face Recognition Systems*  
Faruk AYATA, Hayati ÇAVUŞ ..... 39-48
6. **Kentsel İçme Suyu Şebekelerinde Suyun Kalite Seviyesinin Yapay Zekâ Yaklaşımları Kullanılarak Belirlenmesi**  
*Determining Water Quality in Urban Drinking Water Networks Using Artificial Intelligence Approaches*  
Abdullah ŞENER, Burhan ERGEN, Hamit MIZRAK ..... 49-57



## Bazı Karbonil Bileşikleri ile Kitosan Aşı Kopolimerlerinin Sentezi ve Karakterizasyonu

Rüveyda Şener Gürsoy<sup>1\*</sup>, Uğursoy Olgun<sup>2</sup>

<sup>1</sup> Sakarya Üniversitesi, Fen-Edebiyat Fakültesi, Kimya Bölümü, Serdivan Sakarya, Türkiye

<sup>2</sup> Polimer Tekmer Polimer Malzeme ve Teknolojileri Araştırma Uygulama Merkezi, Polimer, Kaplama ve Kauçuk Teknolojileri Lab. Sargem Araştırma-Geliştirme Uygulama ve Araştırma Merkezi, Sakarya Üniversitesi, Serdivan, Sakarya, Türkiye

\*<sup>1</sup> ruveyda.sener1@ogr.sakarya.edu.tr, <sup>2</sup> uolgun@sakarya.edu.tr

(Geliş/Received: 19/07/2021;

Kabul/Accepted: 01/11/2021)

**Öz:** Bu çalışmada, kitosan polimeri sodyum kloroasetat, tereftalaldehit ve sinneroil klorür ile modifiye edilmiştir. Kitosanın sodyum kloroasetat ile modifikasyon reaksiyonları; 1) öğütülerek, çözücüsüz, 120 °C’de ve 2) etilen glikol çözücüsü kullanılarak 120 °C’de olmak üzere iki farklı yöntem ile gerçekleştirilmiştir. Kitosanın tereftalaldehit ile modifikasyon reaksiyonları; etilen glikol içerisinde 1) sodyum hidroksit baz katalizörülüğünde ve 2) sülfanilik asit katalizörülüğünde olmak üzere iki farklı yöntem ile yapılmıştır. Sentezlenen kitosan aşı kopolimerlerinin yapıları FT-IR kızıl ötesi spektroskopisi, UV-vis. absorpsiyon, floresans emisyon ve taramalı elektron mikroskopu (SEM) teknikleri kullanılarak karakterize edilmiştir. Bu çalışmada farklı kitosan aşı kopolimerlerinin sentezi ile kitosanın çözünürlüğünün artırılması, kitosana farklı fonksiyonel gruplar eklenmesi, kitosanın UV görünür bölgedeki ışık soğurma ve floresans emisyon özelliklerinin kazandırılması hedeflenmiştir. Tereftalaldehit ile sentezlenen aşı kopolimerinin DMSO’da 275 nm ve 328 nm’de absorpsiyon yaptığı ölçülmüştür. Ayrıca; 447 nm’de floresans emisyon gösterdiği belirlenmiştir. Sentezlenen biyouyumlu kitosan aşı kopolimerlerinin tıbbi görüntüleme, tedavi ve ilaç salımı amaçlı kullanım potansiyalleri yüksektir.

**Anahtar kelimeler:** Kitosan, Modifikasyon, Biyopolimer, Spektroskopi, Aşı Kopolimeri

### Synthesis and Characterization of Chitosan Graft Copolymers with Some Carbonyl Compounds

**Abstract:** In this study, chitosan polymer was modified with sodium chloroacetate, terephthalaldehyde and cinnamoylchloride. Modification reaction of chitosan with sodium chloroacetate; It was carried out by two different methods, 1) by grinding solvent-free at 120 °C and 2) using ethylene glycol solvent at 120 °C. The modification reactions of chitosan were made with terephthalaldehyde in ethylene glycol with two different methods; 1) sodium hydroxide base catalyzed and 2) sulfanilic acid catalyzed. The structures of the synthesized chitosan graft copolymers were characterized using FT-IR spectroscopy, UV-vis. absorption, fluorescence emission and scanning electron microscopy (SEM) techniques. In this study, it was aimed to increase the solubility of chitosan with the synthesis of different chitosan graft copolymers. In order to provide in proved light absorption and fluorescent emission properties of chitosan in the UV visible region, the different functional groups was attached to the chitosan. It was noted that the graft copolymer synthesized with terephthalaldehyde gives absorption bands at 275 nm and 328 nm in DMSO. Also, it showed fluorescence emission at 447 nm. Synthesized biocompatible chitosan graft copolymers have high potential for use in medical imaging, treatment and drug delivery purposes.

**Key words:** Chitosan, Modification, Biopolymer, Spectroscopy, Graft Copolymer

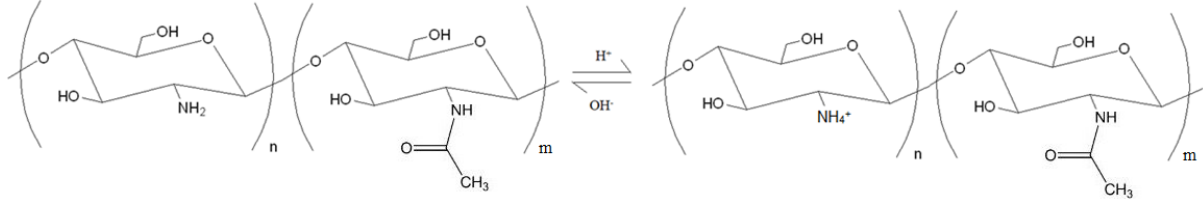
### 1. Giriş

Kitosan, kitinin kısmi N-deasetilasyon reaksiyonu ile hazırlanan N-deasetile türevidir [1] ve poli- $[\beta-(1,4)-2\text{-amino-2-deoksi-}\beta\text{-D-glukopiranoz}]$  yapısındadır [2]. İlk olarak 1859 yılında Rouget tarafından kitinin derişik potasyum hidroksit ile kaynatılması sonucunda elde edilmiştir [3]. Bazı mantar hücre duvarlarında doğal olarak bulunmakla birlikte çoğunlukla yengeç, karides, kril kabukları ve mantarlardan elde edilir [4,5]. Rengi açık sarıdan beyaza kadar çeşitli tonlarda deęişebilir ve toz hali oldukça yumuşaktır [6]. Saf kitosan sulu ortamda çözünmeyen bir biyopolimerdir [7]. Serbest amino gruplarına sahip olması sebebiyle sulu asidik ortamlarda çözünür [8]. Kitosan biyopolimerinin ortam pH’ına göre yapısı Şekil 1’de verilmiştir.

Kitosan kimyasal olarak uygun fonksiyonel gruplara sahip olması sebebiyle polimerizasyon derecesini bozmadan modifiye edilebilir [4]. Pozitif yüke sahiptir ve bu durum kitosanın negatif yüklü yağlar, lipitler, metal iyonları, kolesterol, protein ve makromoleküllerle bağlanmasını sağlar. Kitosan ve kitosan türevleri ticari olarak gittikçe önem kazanan polimerler haline gelmiştir [2]. Son yıllarda yapılan çalışmalar patojenik bakterilerin

\* Sorumlu yazar: [ruveyda.sener1@ogr.sakarya.edu.tr](mailto:ruveyda.sener1@ogr.sakarya.edu.tr). Yazarların ORCID Numarası: <sup>1</sup> 0000-0003-3967-5331, <sup>2</sup> 0000-0001-7104-9926

antibakteriyel ürünlere karşı direncinin giderek arttığını göstermiştir [9]. Bu nedenle, mikroorganizma ve virüs kaynaklı enfeksiyonların önlenmesi için yeni antimikrobiyal ve antiviral maddelerin geliştirilmesi günümüzde büyük önem taşımaktadır [10]. Kitosan modifikasyonları kitosana çeşitli fonksiyonel gruplar bağlayarak antimikrobiyal aktivitesi ve çözünürlüğünü arttırmak, diğer maddeler ile etkileşim özelliklerini değiştirmek, mekanik dayanım ve kimyasal kararlılığı iyileştirmek, hidrofobik, katyonik ve anyonik özellikleri kontrol etmek amacıyla gerçekleştirilmektedir [11,12].



Şekil 1. Doğal kitosan biyopolimerinin ortam pH'sına göre molekül yapısı.

Kitosan ve kitosan türevleri; ilaç, gıda ve beslenme, kozmetik, ambalaj, tıp, ziraat, tekstil gibi çeşitli endüstrilerde sınırsız kullanım alanına sahiptir [13,14]. Medikal alanda antitümör, antiülser ve antikoagülant olarak, kanamanın durdurulması, diş uygulamaları, kolesterol kontrolü ve ilaçlarda, dikiş ipi, yara bandı, sargı bezi, medikal yapay deri, kontakt lens yapımı vb. şekillerde kullanılmaktadır. Bunun yanı sıra yara tedavisini %30 oranında hızlandırmaktadır [14,15,16]. Gıda alanında şarap, meyve suyu, bira gibi içeceklerde etkili bir ajan olarak bulanıklığı giderir. Gıda işletmelerindeki atık suların proteinlerin, yağların ve metal iyonlarının arıtılmasında kullanılır [13,17]. Gıda ambalajlamada raf ömrünü ve gıda kalitesini artırmadaki etkililiği sayesinde akıllı ambalaj geliştirmek için yenilebilir filmler veya kaplamalar olarak, bununla birlikte doğal bir gıda koruyucusu olarak görev yapar [18,19,20]. Tarımda ve biyoteknolojide kompostasyon hızlandırıcı olarak, kozmetikte nemi korumak, akneyi tedavi etmek, saç esnekliğini artırmak, cildi tonlamak gibi amaçlarla diş macunu, diş beyazlatıcı ve sakızlar gibi diş bakım ürünlerinde, saç şampuanı, saç spreyi ve saç boya gibi saç bakımında kullanılan ürünlerde, yüz, el ve vücut kremlerinde, koku giderme ürünlerinde, ruj, göz farı ve oje gibi renkli kozmetik ürünlerin üretiminde kullanılmaktadır [21,22]. Bu çalışmada kitosan bazı karbonil bileşikleriyle modifiye edilmiştir. Kitosanın karboksilasyon ve schiff bazı aşı kopolimerlerinin sentez işlemlerinde literatürden farklı çözücüler, asit-baz katalizörleri ve sıcaklık koşulları kullanılmıştır. Kıyaslamalı olarak reaksiyon şartlarının kitosan ürün profiline etkileri ele alınmıştır. Hazırlanan kitosan aşı kopolimerlerinin karakterizasyonları FT-IR, UV-vis. absorpsiyon, floresans emisyon ve taramalı elektron mikroskobu (SEM) analizleri ile yapılmıştır [23].

## 2. Deneysel

### 2.1 Metaryal ve metod

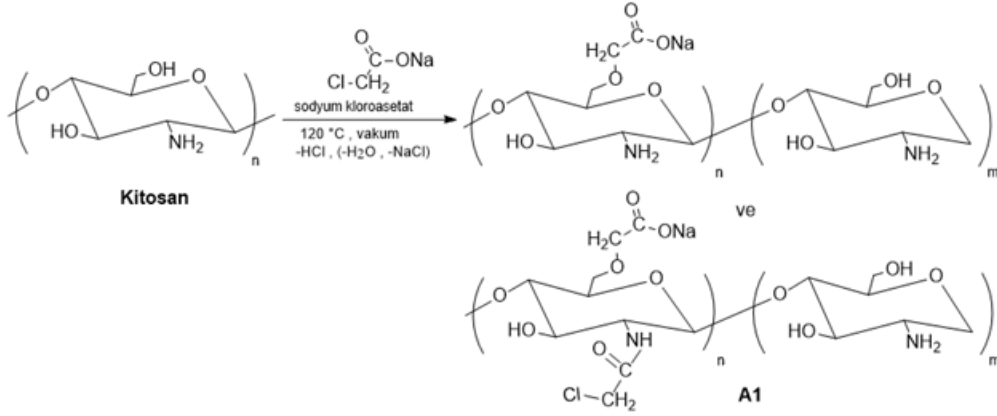
Kitosan, Sodyum kloroasetat, sinamoill klorür, sülfanilik asit, ve sodyum hidroksit Sigma Aldrich'den; etilen glikol, etil asetat ve tetrahidrofur (THF), Merck'den; teraftalaldehit Alfa Easer'dan alınmıştır.

### 2.2 Sodyum kloroasetat ile kitosan aşı kopolimerlerinin sentezi

Kitosan bu çalışmada farklı sıcaklıklarda ve farklı çözücülerde iki farklı sentez yöntemi ile modifiye edilmiştir. Kitosanın ClCH<sub>2</sub>COOH ile karboksilasyon reaksiyonları literatürde çalışılmıştır [24]. Çalışmamızda kullanılan sodyum kloroasetat toksik olmaması ve biyolojik olarak kolayca bozunabilir olması açısından tercih edilmiştir [25]. Sentezlenen aşı kopolimerleri sırasıyla A1 ve A2 olarak adlandırılmıştır.

#### 2.2.1 Katı fazda sodyum asetat-kitosan-ko-kitosan aşı kopolimeri (A1) sentezi

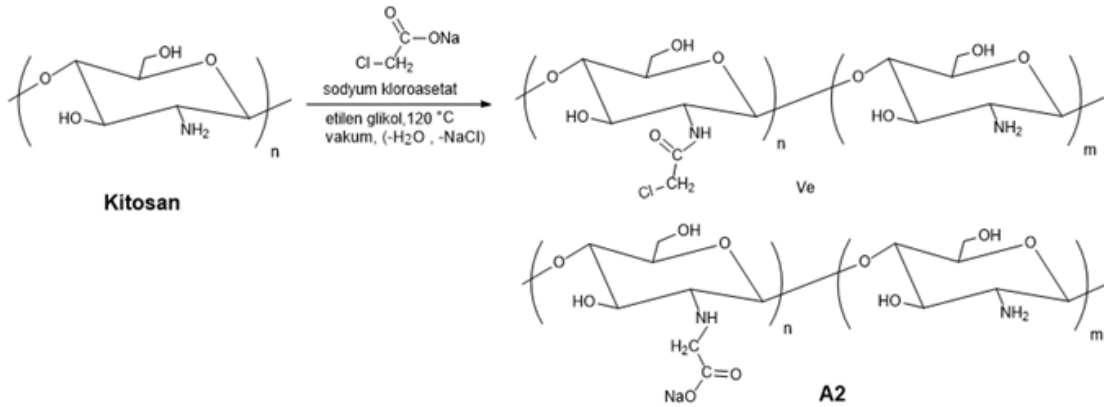
Sodyum asetat-kitosan aşı kopolimerinin (A1) sentezi için tartımı alınan 0,20 g kitosan üzerine yaklaşık 4 mmol 0,376 g sodyum kloroasetat katılıp öğütülerek 120 °C de vakum ortamında 24 saat manyetik karıştırıcıda karıştırılmıştır. Reaksiyon sonunda numunede sararma görüldü. Vakum ortamında gaz fazda sıvı damlacıklar oluşmuştur ve reaksiyonda HCl çıkışı olduğu düşünülerek oluşan damlacıklarda pH ölçülmüş ve pH'ın 2 olduğu belirlenmiştir. Düşük pH 2 değeri kitosan ile sodyum kloroasetat arasındaki reaksiyonun sonucunda HCl oluştuğunu göstermektedir. Sentezlenen polimer (A1) tartım sonucu 0,4334 g olarak bulunmuştur. Sodyum asetat-kitosan-ko-kitosan aşı kopolimeri sentez (A1) reaksiyonu Şekil 2'de verilmiştir.



Şekil 2. Sodyum asetat-kitosan-ko-kitosan aşısı kopolimeri (A1)'nin sentez reaksiyonu.

### 2.2.2 Etilen glikolde kloro asetat-kitosan-ko-kitosan aşısı kopolimeri (A2) sentezi

Kloro asetat-kitosan-ko-kitosan aşısı kopolimeri (A2) sentezi farklı bir yöntemle etilen glikol içinde sentezlenmiştir. 0,20 g kitosan üzerine 10 mL etilen glikol ve yaklaşık 4 mmol 0,37 g sodyum kloroasetat katıldı.  $120\text{ }^\circ\text{C}$  de vakum ortamında 24 saat manyetik karıştırıcıda karıştırıldı. 24 Saat sonunda oluşan koyu kahverengi çökelti su ile yıkanıp, dekante edildi. Santrifüjlenerek  $105\text{ }^\circ\text{C}$  de etüvde 2 saat kurutulmuştur. Sentezlenen kloro asetat-kitosan-ko-kitosan aşısı kopolimeri (A2) hassas terazide tartılmıştır ve tartım sonucu 0,2725 g olarak bulunmuştur. Kurutulan A2 polimeri sentez reaksiyonu Şekil 3'te verilmiştir.



Şekil 3. Kloro asetat-kitosan-ko-kitosan aşısı kopolimeri (A2)'nin sentez reaksiyonu.

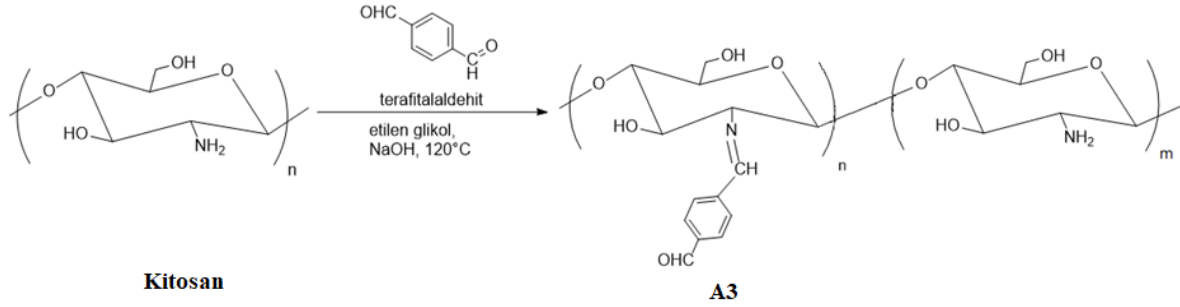
### 2.3 Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşısı kopolimerlerinin Sentezi

Kitosan baz katalizörü olarak NaOH ve asit katalizörü olarak da sülfanilik asit kullanılarak iki farklı sentez yöntemi ile modifiye edilerek teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşısı kopolimerleri sentezlenmiş ve elde edilen kopolimer ürünleri sırasıyla A3 ve A4 olarak isimlendirilmiştir.

#### 2.3.1 Baz katalizörlü teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşısı kopolimeri (A3) Sentezi

Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşısı kopolimeri (A3) sentez reaksiyonu için 0,20 g kitosan üzerine yaklaşık 0,55 g (4 mmol) teraftalaldehit ve 5 mL etilen glikol ilave edilerek  $120\text{ }^\circ\text{C}$  de vakum ortamında 4 saat manyetik karıştırıcıda karıştırıldı. Yaklaşık 4 saat sonrasında 2 mL 0,2 M NaOH ilave edilerek karıştırılmaya devam edilmiştir. Oluşan koyu sarı renkli karışım 30 dk daha  $120\text{ }^\circ\text{C}$  sıcaklıkta kırmızı renk oluşuncaya kadar ısıtmaya devam edilmiştir. Sonrasında kırmızı renkli karışım, reaksiyon süresi toplam 24 saat oluncaya kadar ısıtmaya devam edilmiştir. Meydana gelen sarı kahverengi çökelti reaksiyona girmemiş başlangıç maddeleri ile diğer yan ürünleri uzaklaştırmak için 10 mL etilasetat ve sonrasında THF fazlası ile birkaç defa yıkanmıştır. Santrifüjlenerek ayrılan kitosan aşısı kopolimeri  $105\text{ }^\circ\text{C}$  de etüvde 2 saat kurutulmuştur. Elde edilen ürün tartımı

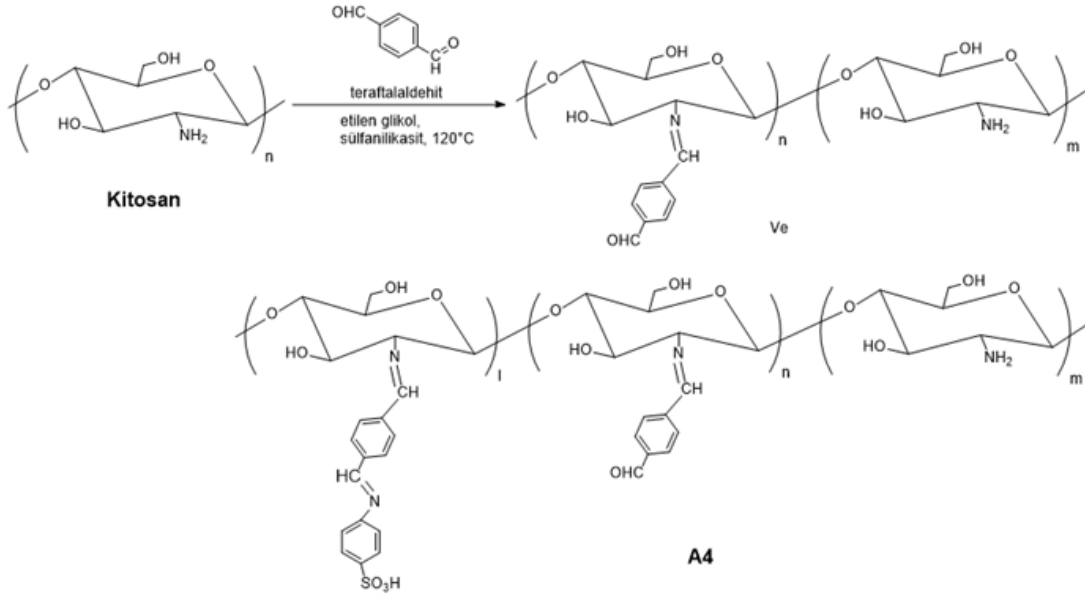
0,4545 g olarak ölçülmüştür. Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A3) sentez reaksiyonu Şekil 4'te verilmiştir.



Şekil 4. Baz katalizörlü teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A3)'nin sentez reaksiyonu.

### 2.3.2 Asit katalizörlü teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4) sentezi

Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4) sentezi için 0,20 g kitosan ve 0,55 g (4 mmol) teraftalaldehit üzerine 5 mL etilen glikol eklenerek yaklaşık 120 °C de vakum ortamında 1 saat 30 dk manyetik karıştırıcıda karıştırılmıştır. Oluşan hafif sarı renkli karışım üzerine 0,33 g sülfanilik asit eklenerek reaksiyon süresi 24 saate oluncaya kadar ısıtmaya devam edilmiştir. Elde edilen reaksiyon ürünü katı çökelti santrifüjlenerek ayrılmıştır. Sentezlenen teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4) 105 °C de etüv'de kurutulmuştur. Katı ürün terazide tartılarak, tartım sonucu 0,6372 g bulunmuştur. Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4) sentez reaksiyonu Şekil 5'te verilmiştir.



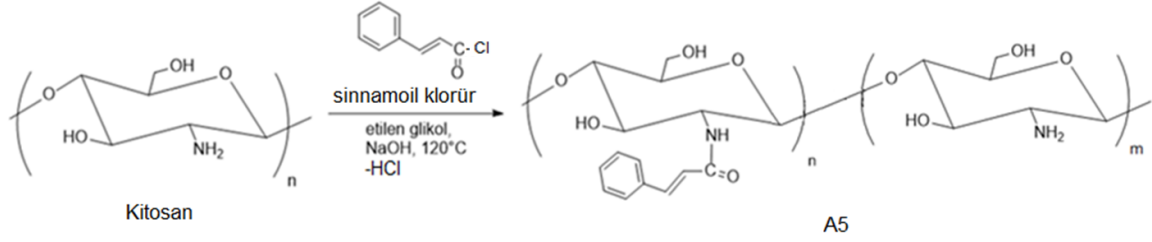
Şekil 5. Asit katalizörlü teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4)'nin sentez reaksiyonu.

### 2.4. Baz katalizörlü sinamoil klorür-kitosan-ko-kitosan aşı kopolimeri (A5) sentezi

Baz katalizörlü sinamoil klorür-kitosan-ko-kitosan aşı kopolimeri (A5) sentez reaksiyonu için 0,20 g kitosan ve yaklaşık 0,68 g (4 mmol) sinamoil klorür üzerine 5 mL etilen glikol ilave edilerek 120 °C de vakum ortamında 4 saat manyetik karıştırıcıda karıştırılmıştır. Yaklaşık 4 saat sonrasında, baz katalizörü olarak karışım üzerine 2 mL 0,2 M NaOH ilave edilmiştir. Bu sırada beyaz renkli çökelti meydana gelmiştir. Reaksiyon süresi 24 saat oluncaya kadar sıcakta karıştırma işlemine devam edilmiştir. Reaksiyon sonunda koyu kahve renkli çökelti oluşmuştur. Sıvı kısım dekantasyonla ayrıldıktan sonra kalan ürün saf su ile birkaç kez yıkanmıştır. Santrifüj işlemi ile ayrılan katı numune etüvde 105 °C'de 3 saat kurutulmuştur. Sentezlenen sinamoil klorür-kitosan-ko-kitosan



aşı kopolimeri (A5) terazide tartılarak kütlesi 0,33 g olarak bulunmuştur. Baz katalizörlü olarak Sınnaoil klorür-kitosan-ko-kitosan aşı kopolimeri (A5) sentez reaksiyonu Şekil 6'da verilmiştir.



Şekil 6. Baz katalizörlü sınaoil klorür-kitosan-ko-kitosan aşı kopolimeri (A5) sentez reaksiyonu.

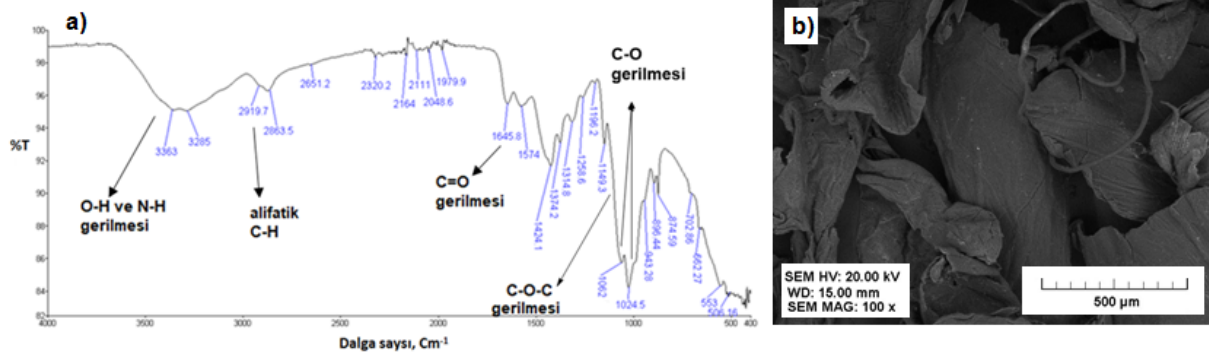
## 2.5. Karakterizasyon yöntemleri

Sentezlenen kitosan aşı kopolimerlerinin FT-IR analizleri Perkin-Elmer FT-IR spektrofotometresi ile yapılmıştır. Sentezlenen kitosan aşı kopolimerleri 400-4000  $\text{cm}^{-1}$  kızıl ötesi bölgesinde ölçülmüştür. DMSO'da çözünen Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimerinin (A4) absorpsiyon spektroskopisi analizi UV-2600 Shimadzu marka UV-Vis spektrofotometre cihazı ile ve floresans emisyon spektroskopisi analizi ise F-7000 Hitachi marka floresans spektrofotometre cihazı ile yapılmıştır. Sentezlenen kitosan aşı kopolimerleri'nin tanecik boyutları ve yüzey morfolojileri Vega Tescan marka, taramalı elektron mikroskobu (SEM) kullanılarak incelenmiştir.

## 3. Deneysel Bulgular ve Tartışma

Bu çalışmada farklı kitosan aşı kopolimerleri sentezlenerek yapısal özellikleri incelenmiştir. Kopolimerlerin FT-IR ve SEM elektron mikroskobu ile yapılan analizleri başlangıç maddesi olarak kullanılan kitosan kopolimerinin Şekil 7'de verilen analiz sonuçları ile kıyaslanarak yorumlanmıştır.

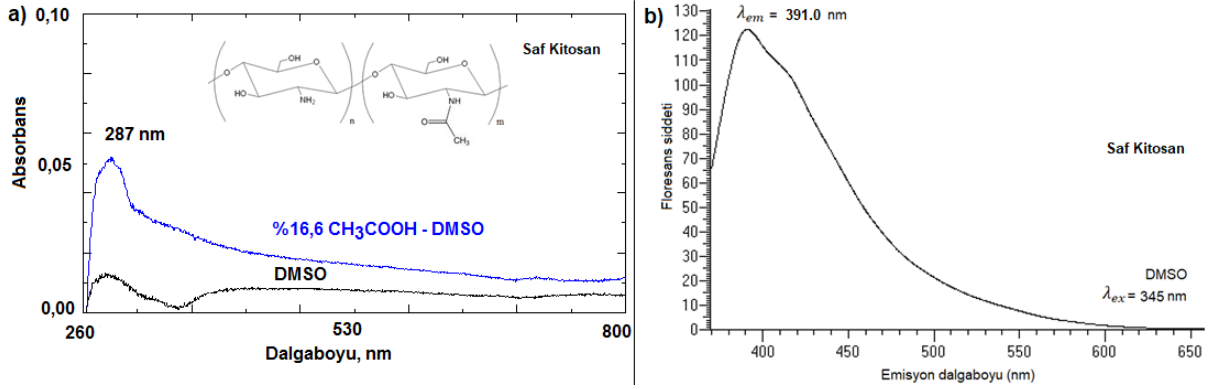
Kitosan'ın kızılötesi spektrumunu Şekil 7a'da verilmiştir. Benzer şekilde aşı kopolimerlerinin sentezinde kullanılan kitosanın alınan taramalı elektron mikroskobu (SEM) görüntüsü Şekil 7b'de verilmiştir.



Şekil 7. Kopolimer sentezlerinde kullanılan kitosanın a) FT-IR spektrumu ve b) SEM görüntüsü.

Spektrumun 3285-3363  $\text{cm}^{-1}$  bölgesindeki güçlü bant, molekül içi hidrojen bağlarının yanı sıra N-H ve O-H gerilme titreşimlerine karşılık gelmektedir. Yaklaşık 2919  $\text{cm}^{-1}$  ve 2863  $\text{cm}^{-1}$  civarındaki absorpsiyon bantları alifatik  $\text{CH}_2$  gruplarına ait sırasıyla C-H simetrik ve asimetric gerilme titreşimleridir. Kitosan yapısında deasetillenmeden kalan N-asetil gruplarının varlığı, sırasıyla yaklaşık 1645  $\text{cm}^{-1}$  (amid C=O gerilmesi) ve 1314  $\text{cm}^{-1}$ 'deki (amid C-N gerilmesi) bantları ile doğrulanmaktadır. Yaklaşık 1574  $\text{cm}^{-1}$ 'deki titreşim bandı, birincil amine ait N-H bükülmesine karşılık gelmektedir. Yapıda yer alan CH-OH bükülme titreşimi ve  $\text{CH}_2$ -OH simetrik deformasyon titreşimleri sırasıyla 1424  $\text{cm}^{-1}$  ve 1374  $\text{cm}^{-1}$  civarındaki bantların varlığı ile doğrulanmıştır. Yaklaşık 1149  $\text{cm}^{-1}$ 'deki absorpsiyon bandı, C-O-C köprüsünün asimetric gerilme titreşimidir. Spektrumda görülen 1062  $\text{cm}^{-1}$  ve 1024  $\text{cm}^{-1}$ 'deki bantlar C-O gerilme titreşimlerine karşılık gelmektedir.

Kitosanın UV görünür bölge spektrumu DMSO'da ve %16,6  $\text{CH}_3\text{COOH}$ -DMSO'da ölçülerek Şekil 8a'da verilmiştir. Kitosanın floresans emisyon spektrumu ise DMSO çözeltisi 345 nm dalga boyunda uyarılarak ölçülmüştür ve elde edilen spektrum Şekil 8b'de verilmiştir.



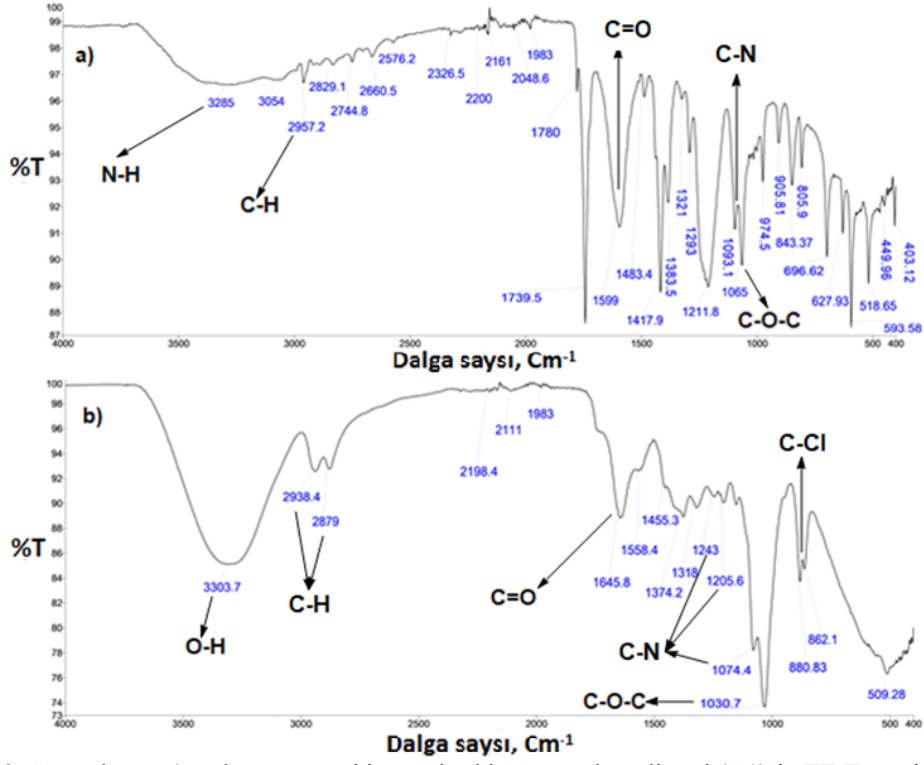
Şekil 8. Kopolimer sentezlerinde kullanılan kitosanın a) DMSO'da ve %16,6 CH<sub>3</sub>COOH-DMSO'da ölçülen UV spektrumu ve b) DMSO'da 345 nm'de uyarılarak ölçülen floresans spektrumu.

Kitosanın DMSO'da çözünürlüğünün çok düşük olduğu görülmüştür. Ancak CH<sub>3</sub>COOH ilavesi ile kitosanın çözünürlüğünün arttığı görülmektedir. Kitosan %16,6 CH<sub>3</sub>COOH-DMSO'da 287 nm'de absorpsiyon sinyali vermiştir. Kitosan DMSO çözeltisinde 391.0 nm'de emisyon yaptığı belirlenmiştir.

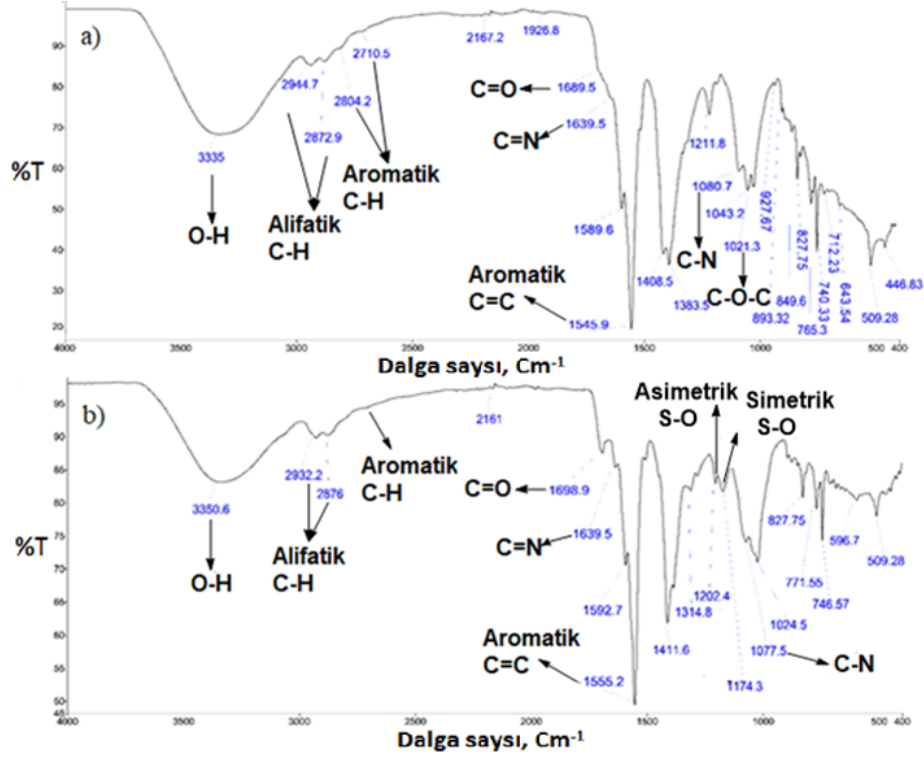
Kitosanın teraftalaldehit ile reaksiyonu literatürde çalışılmıştır. Li Z. ve arkadaşları hazırladıkları NiFe<sub>2</sub>O<sub>4</sub>-COF-Kitosan-teraftalaldehit nanokompozit filmin antibiyotik kirleticilerin giderilmesi için su arıtımında büyük potansiyele sahip olduğunu görmüşlerdir [26]. Xu B. ve arkadaşları çalışmalarında eğreltiotu yaprağı benzeri  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> mikropartikülleri ve kitosan (CS) hibrit filmine dayanan yeni bir elektrokimyasal DNA biyosensörü geliştirmiş ve teraftalaldehiti iki fonksiyonlu bir bağlayıcı olarak kullanılmıştır [27]. Çapraz bağlı kitosan hidrojelleri, nötr ortamda kitosan amino grupları ve teraftalaldehitin aldehit grupları arasında bir Schiff bazı bağının oluşturulması yoluyla hazırlanabilir. Kitosan ve teraftalaldehit arasındaki Schiff baz bağları da ilginç bir pH'a duyarlı stabilizeye sahiptir [28].

### 3.1. FT-IR kızıl ötesi spektroskopisi analizleri

Şekil 9'da kitosan A1 ve A2 polimerlerinin FT-IR spektrumları verilmiştir. Kitosan A1 polimerinin kızıl ötesi spektrumunda; spektrumun 3000-3500 cm<sup>-1</sup> bölgesindeki O-H bandı, kitosanın -OH grubuna karbonil grubu bağlı olması sebebiyle daha az şiddetlidir. Yaklaşık 3285 cm<sup>-1</sup>'deki pik NH<sub>2</sub> grubundan gelen N-H gerilme titreşimidir. Alifatik CH<sub>2</sub> gruplarına ait C-H simetrik ve asimetric gerilme titreşimleri 2957 cm<sup>-1</sup> civarında görülmektedir. Yapıda yer alan 1599 cm<sup>-1</sup>'deki absorpsiyon bandı amid C=O gerilmesi, 1093 cm<sup>-1</sup>'deki absorpsiyon bandı C-N gerilme titreşimidir. Yaklaşık 1065 cm<sup>-1</sup> de C-O-C köprüsünün asimetric gerilmesi görülmektedir. Kitosan A2 polimerinin kızılötesi spektrumunda; karboksilden gelen -OH grubu fazla olduğu için 3303 cm<sup>-1</sup>'de görülen O-H bandı daha belirgin şekilde görülmüştür. Yaklaşık 2938 cm<sup>-1</sup> ve 2879 cm<sup>-1</sup> civarındaki absorpsiyon bantları alifatik C-H gerilme titreşimleridir. Kitosan yapısında deasetillenmeden kalan N-asetil grubu amid C=O gerilmesi yaklaşık 1645 cm<sup>-1</sup>'de görülmüştür. Spektrumda yer alan 1455 cm<sup>-1</sup>, 1205 cm<sup>-1</sup> ve 1074 cm<sup>-1</sup> pikleri C-N gerilme titreşimini göstermektedir. Yaklaşık 1030 cm<sup>-1</sup>'de bulunan titreşim bandı yeni C-O-C gerilmesidir. Parmak izi bölgesindeki 860 cm<sup>-1</sup> ve 765 cm<sup>-1</sup>'de gözlenen pikler C-Cl gerilme titreşimidir. Şekil 10'da kitosan A3 ve A4 polimerlerinin kızılötesi spektrumları verilmiştir.

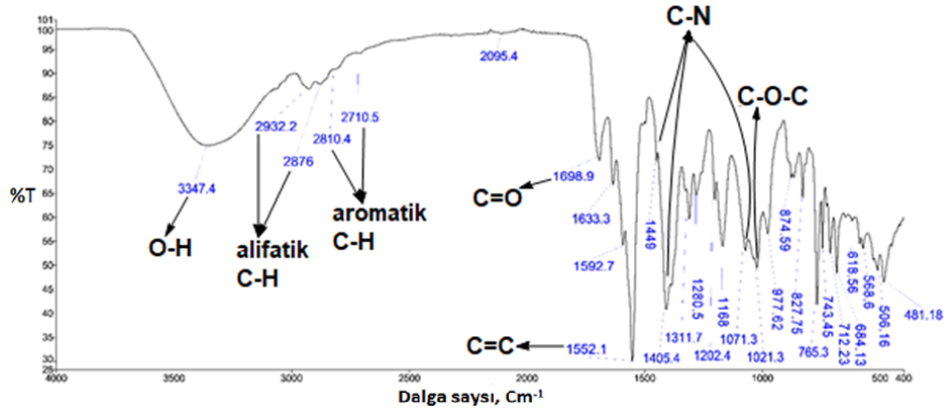


Şekil 9. Sentezlenen a) sodyum asetat-kitosan-ko-kitosan aşı kopolimeri (A1)'in FT-IR spektrumu, b) kloro asetat-kitosan-ko-kitosan aşı kopolimeri (A2)'nin FT-IR spektrumu.



Şekil 10. Sentezlenen a) teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri-baz katalizörlü ürün (A3)'ün FT-IR spektrumu, b) teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri-asit katalizörlü ürün (A4)'ün FT-IR spektrumu.

Kitosan A3 polimerin kızılötesi spektrumunda; yaklaşık 2944-2872  $\text{cm}^{-1}$ 'deki bantlar alifatik C-H, 2804-2710  $\text{cm}^{-1}$ 'deki bantlar aromatik C-H titreşimlerini gösterir. Yapıda yer alan C=N schiff bazı spesifik bağı 1639  $\text{cm}^{-1}$ 'de görülmüştür. Yaklaşık 1545  $\text{cm}^{-1}$ 'deki absorpsiyon bandı aromatik C=C gerilmesini, 1080  $\text{cm}^{-1}$ 'deki absorpsiyon bandı C-N gerilmesini göstermektedir. Kitosan A4 polimerin kızılötesi spektrumunda farklı olarak Yaklaşık 1202  $\text{cm}^{-1}$  ve 1174  $\text{cm}^{-1}$ 'de  $-\text{SO}_3\text{H}$  grubuna ait sırasıyla asimetric S-O ve simetric S-O titreşimleridir. Şekil 11'de Kitosan A5 polimerinin kızılötesi spektrumu verilmiştir. Kitosan A5 polimerin kızılötesi spektrumunda; C-N bağı titreşimi 1449  $\text{cm}^{-1}$ , 1405  $\text{cm}^{-1}$  ve 1071  $\text{cm}^{-1}$  civarındaki bantların varlığı ile doğrulanmıştır.



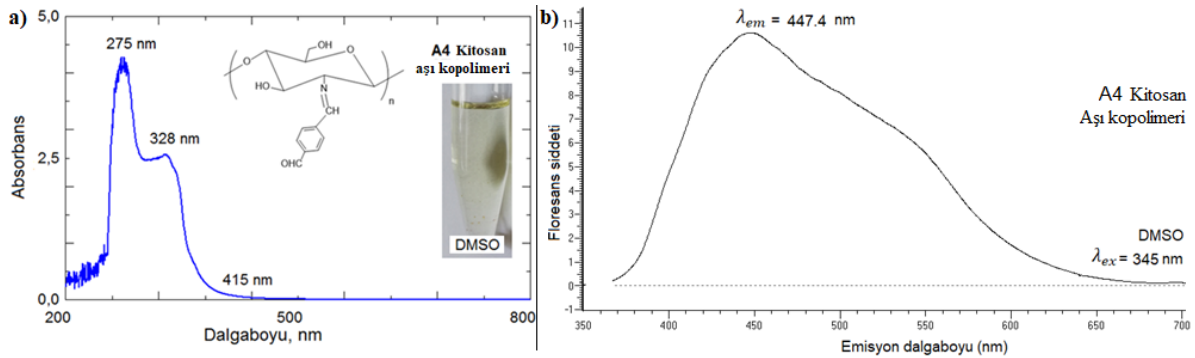
Şekil 11. Sinamoill klorür-kitosan-ko-kitosan aşı kopolimeri-baz katalizörlü ürün (A5)'ün FT-IR spektrumu.

### 3.2 UV-vis absorpsiyon spektroskopisi analizleri

Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4) DMSO'da çözülmüştür. DMSO'da çözülerek hazırlanan A4 polimer numunesinin UV-vis. Spektrumu Şekil 12'de verilmiştir. Spektrumda 275 nm'de gözlenen maksimum absorpsiyon bandı benzen halkasındaki  $\pi-\pi^*$  ve  $\sigma-\sigma^*$  bağlarından kaynaklanmaktadır. Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimerinde, teraftalaldehit, kitosan ana zinciri boyunca amin grupları ile reaksiyona girerek aromatik Schiff bazı oluşturacak şekilde bağlanmaktadır. Elde edilen aşı kopolimeri Schiff bazlarına özgü karakteristik sarı renkli olarak elde edilmiştir. Bu renk oluşumu kromofor grup olan  $-\text{CH}=\text{N}-$  fonksiyonel bağ yapısının varlığına ve bu gruptan kaynaklı olarak 328 nm'de gözlenen  $n-\pi^*$  geçişlerine işaret etmektedir. Sonuç olarak, ürünün UV-vis. spektrumunda 328 nm'de gözlenen maksimum absorpsiyon sinyali Schiff bazı oluşumunu teyit etmektedir.

### 3.3 Floresans emisyon spektroskopisi analizleri

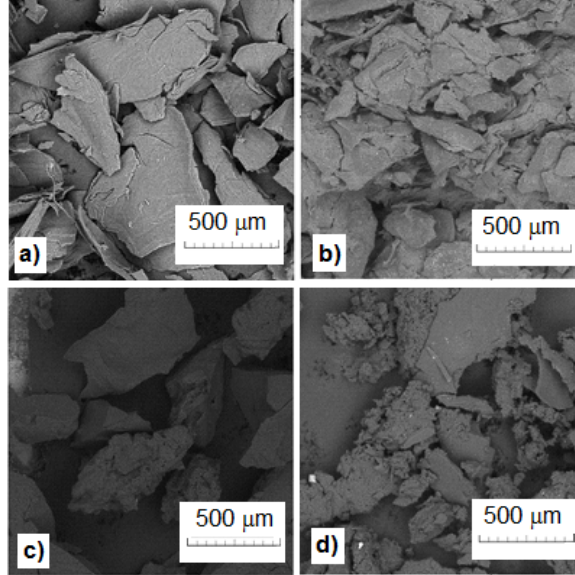
Sentezlenen teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimeri (A4)'nin 345 nm dalga boyunda uyarılarak DMSO'da ölçülen floresans spektrumu Şekil 12'de verilmiştir. Kitosan A4 aşı kopolimerinin DMSO çözeltisi 345 nm'de uyarıldığında 447,4 nm'de floresans emisyon sinyali vermiştir.



Şekil 12. Teraftalaldehit-kitosan Schiff bazı-ko-kitosan aşı kopolimerinin (A4) DMSO'da ölçülen a) UV spektrumu ve b) 345 nm'de uyarılarak ölçülen floresans spektrumu.

### 3.4. Taramalı elektron mikroskobu (SEM) analizleri

A2, A3, A4 ve A5 polimerlerinin 500  $\mu\text{m}$ ' de alınan Taramalı Elektron Mikroskobu görüntüleri Şekil 13'te verilmiştir.



Şekil 13. a) A2 polimerinin 500  $\mu\text{m}$ ' de sem görüntüleri b) A3 polimerinin 500  $\mu\text{m}$ ' de sem görüntüleri c) A4 polimerinin 500  $\mu\text{m}$ ' de sem görüntüleri d) A5 polimerinin 500  $\mu\text{m}$ 'de SEM görüntüleri.

### 4. Sonuçlar

Farklı karbonil bileşikleri kullanılarak sentezlenen kitosan aşılı kopolimerlerinin FT-IR spektrumları ölçülmüş ve kitosana kıyasla farklı frekanslarda karakteristik yeni pikler verdikleri belirlenmiştir. Belirlenen yeni pikler sentezlenen kitosan aşılı kopolimerlerindeki kimyasal bağlarla uyumludur. Sentez reaksiyonundan sonra kitosan aşılı kopolimerlerinde yer alan bu yeni bağlar; kitosan A1 polimerinde C-N, kitosan A2 polimerinde C-N ve C-Cl, kitosan A3 polimerinde C=N, C-N ve C=C, kitosan A4 polimerinde S-O, C=N, C-N, C=C ve kitosan A5 polimerinde C-N bağlarıdır. Teraftaldehit-kitosan Schiff bazı-ko-kitosan aşılı kopolimeri (A4)'ün DMSO'daki çözeltisi 345 nm'de uyarılarak alınan floresans spektrumu 447,4 nm'de emisyon vermiştir. Etilen glikol içerisinde baz katalizöründe gerçekleştirilen aşılı kopolimeri A3 sentez reaksiyonu, kitosan ortamda yeterince çözünmediği için oldukça yavaş ve katı partikül yüzeylerinde düşük verimle gerçekleşmiştir. Buna karşılık, asidik ortamda kitosan çok daha iyi çözüldüğü için A4 aşılı kopolimeri sentezi çok daha yüksek bir verimle yapılabildiği görülmüştür. Sentezlenen A2, A3, A4 ve A5 kitosan aşılı kopolimerlerinin taramalı elektron mikroskobu SEM görüntüleri incelendiğinde aşılı modifikasyon reaksiyonları ile sentezlenen kopolimerlerin mikro yapılarının ve yüzey morfolojilerinin başlangıç maddesi kitosan'dan oldukça farklı olduğu görülmüştür. Kitosan ile etilen glikolde sentezlenen kitosan A2 polimerlerinin SEM görüntüleri kıyaslandığında tanecik boyutu ve morfolojik yapılarının farklı olduğu belirlenmiştir. Kitosan taneciklerin daha yassı ve küçük olduğu yüzeylerinin ise daha düzgün olduğu görülmektedir. Yapılan elektron mikroskobu analizlerinde baz katalizörü ile sentezlenen teraftaldehit-kitosan Schiff bazı-ko-kitosan aşılı kopolimeri (A3) ile asit katalizörü ile sentezlenen teraftaldehit-kitosan Schiff bazı-ko-kitosan aşılı kopolimeri (A4) numunelerinin mikro yapılarının önemli ölçüde farklılık gösterdiği görülmüştür. Bu çalışma sonucunda sentezlenen kitosan aşılı kopolimerleri kitosana kıyasla daha yüksek çözünürlük, ışık emilimi ve floresans emisyon özellikleri göstermiştir. Sentezlenen kitosan aşılı kopolimerlerinin antibakteriyel, kozmetik ve tıbbi amaçlı kullanım potansiyeli yüksektir. Bunun yanı sıra kitosan kopolimerlerinin ilaç salımı, kanser tedavisi, tıbbi görüntüleme, işaretleyici ajanlar, florometrik sensörler ve boya duyarlı güneş pilleri gibi teknolojik alanlarda kullanımına yönelik çalışmalarımız devam etmektedir.

## Kaynaklar

- [1] Elçi SD. Kitosan türevlerini içeren hidrojellerin ve çözeltilerin hemostatik özelliklerinin değerlendirilmesi. Doktora Tezi, Hacettepe Üniversitesi, Biyoloji Bölümü, 2019.
- [2] Tokatlı K, Demirdöven A. Kitosan ve kitosan bazlı yenilebilir film uygulamaları. *Akad Gıda* 2015; 13(4): 348–353.
- [3] Tokatlı K. Karides atıklarından kitosan üretim koşullarının optimizasyonu ve kitosandan elde edilen yenilebilir film kaplamanın kirazların raf ömrüne etkisi. Doktora Tezi, Gazi Osmanpaşa Üniversitesi, Fen Bilimleri Enstitüsü, Gıda Mühendisliği, 2016.
- [4] Negm NA, Hefni HHH, Abd-Elaal AAA, Badr EA, Abou Kana MTH. Advancement on modification of chitosan biopolymer and its potential applications. *Int J Biol Macromol*. 2020; 152: 681–702.
- [5] Rinaudo M. Chitin and chitosan: Properties and applications. *Prog Polym Sci* 2006; 31: 603–632.
- [6] Demir C. Yara örtüsü olarak papain yüklü kitosan nanopartiküllerinin üretilmesi ve antibakteriyel özelliklerinin incelenmesi. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Bölümü, 2016.
- [7] Salama A, Hasanin M, Hesemann P. Synthesis and antimicrobial properties of new chitosan derivatives containing guanidinium groups. *Carbohydr Polym* 2020; 241: 116363.
- [8] Çankaya N, Sökmen Ö. Kitosan-kil biyomimetik kompozitleri. *Journal of Polytechnic* 2016; 19(3): 283–295.
- [9] Olgun U, Tunc K, Hos A. Preparation and antibacterial properties of nano biocomposite Poly( $\epsilon$ -caprolactone)-SiO<sub>2</sub> films with nanosilver. *Journal of Polymer Research* 2019; 26: 24 .
- [10] Hoş A, Tunç K, Olgun U. Antibacterial nano biocomposite poly( $\epsilon$ -caprolactone) films with nano Ag-hydroxyapatite filler particles. *Composite Interfaces*. 2019.
- [11] Acet Ö. Cu<sup>2+</sup> Takılı O-karboksümetil kitosan schiff bazı kompleksleri gömülü kompozit membranların hazırlanması ve protein adsorpsiyon davranışlarının incelenmesi. Yüksek Lisans Tezi, Aksaray Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Bölümü, 2015.
- [12] Baran T. Suda çözünebilen O-karboksümetil kitosan-imine destekli bazı komplekslerin sentezi ve uygulamaları. Doktora Tezi, Aksaray Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Bölümü, 2015.
- [13] Bostan K, Aldemir T, Aydın A. Kitosan ve antimikrobiyal aktivitesi. *Türk Mikrobiyol Cem Derg*. 2007; 37(2):118–127.
- [14] Hayta EN. Polikaprolakton-Polietilen Glikol-kitosan bazlı mikrokürecik üretimi ve kontrollü ilaç salınımında kullanılması için ön çalışma. Yüksek lisans tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Mühendisliği bölümü, 2014.
- [15] Salama A. Recent progress in preparation and applications of chitosan/calcium phosphate composite materials. *Int J Biol Macromol* 2021; 178: 240–252.
- [16] Altındağ Gİ. Modifiye kitosan türevlerinde immobilize edilen lakkaz ile boya gideriminin incelenmesi. Yüksek Lisans Tezi, Celal Bayar Üniversitesi, Fen Bilimleri Enstitüsü, Kimya bölümü, 2013.
- [17] Kuzgun NK, İnanlı AG. Kitosan üretimi ve özellikleri ile kitosanın kullanım alanları. *Türk Bilim. Derlemeler Derg* 2013; 6(2): 16–21.
- [18] Irastorza A, Zarandona I, Andonegi M, Guerrero P, Caba K. de la, The versatility of collagen and chitosan: from food to biomedical applications. *Food Hydrocoll*. 2021; 116: 106633.
- [19] Wang H, Ding F, Ma L, Zhang Y. Edible films from chitosan-gelatin: Physical properties and food packaging application. *Food Biosci* 2021; 40: 100871.
- [20] Hafsa J, Smach MA, Mrid R, Ben Sobeh M, Majdoub H, Yasri A. Functional properties of chitosan derivatives obtained through Maillard reaction: A novel promising food preservative. *Food Chem* 2021; 349: 129072.
- [21] Pal P, Pal A, Nakashima K, Yadav BK. Applications of chitosan in environmental remediation: A review. *Chemosphere* 2021; 266, 128934.
- [22] Abd El-Hack ME, El-Saadony MT, Shafi ME, Zabermawi NM, Arif M. Batiha GE, Khafaga AF, Abd El-Hakim YM, Al-Sagheer AA. Antimicrobial and antioxidant properties of chitosan and its derivatives and their applications: A review. *Int. J Biol Macromol* 2020; 164: 2726–2744.
- [23] Şener Gürsoy R. Kitosan aşırı kopolimerlerinin sentezi ve karakterizasyonu. Yüksek lisans tezi, Sakarya üniversitesi, Fen Bilimleri Enstitüsü, Kimya Bölümü, 2021.
- [24] Wang W. Xue C. Mao X. Chitosan: Structural modification, biological activity and application. *Int J Biol Macromol* 2020; 164, 4532–4546.
- [25] Elizabe L. Kariuki BM. Harris KDM. Tremayne M. Epple M. Thomas JM. Topochemical rationalization of the solid-state polymerization reaction of sodium chloroacetate: Structure determination from powder diffraction data by the monte carlo method. *J. Phys. Chem*. 1997; 101: 44.
- [26] Lia Z. Liua Y. Zoub S. Lua C. Baia H. Mua H. Duana J. Removal and adsorption mechanism of tetracycline and cefotaxime contaminants in water by NiFe<sub>2</sub>O<sub>4</sub>-COF-chitosan-terephthalaldehyde nanocomposites film. *Chemical Engineering Journal* 2020; 382: 123008.
- [27] Xu B. Zheng D. Qiu W. Gao F. Jiang S. Wang Q. An ultrasensitive DNA biosensor based on covalent immobilization of probe DNA on fern leaf-like  $\alpha$ -Fe<sub>2</sub>O<sub>3</sub> and chitosan Hybrid film using terephthalaldehyde as arm-linker. *Biosensors and Bioelectronics* 2015; 72: 175-181.
- [28] Chu LY. Xie R. Ju XJ. Wang W. Preparation and Properties of Monodisperse pH-Responsive Microgels. *Smart Hydrogel Functional Materials*. Springer. 2031: 155-170.



## Bisfenol A (BFA) ve Halk Sağlığı Açısından Önemi

Sait TAN<sup>1</sup>, Pelin DEMİR<sup>2\*</sup>, Ali ARSLAN<sup>3</sup>

<sup>1,2,3</sup> Firat Üniversitesi, Veteriner Fakültesi, Besin Hijyeni ve Teknolojisi Bölümü, Elazığ, Türkiye  
<sup>1</sup> saittan0221@gmail.com, <sup>2</sup> p.demir@firat.edu.tr, <sup>3</sup> aarslan2@firat.edu.tr

(Geliş/Received: 06/09/2021;

Kabul/Accepted: 30/03/2022)

**Öz:** Bisfenol A, polikarbonat plastiklerin ve epoksi reçinelerin üretiminde kullanılan insan sağlığı açısından en tehlikeli kimyasallardandır. Gıda sanayinde kullanılan bazı plastik ve konserve gibi gıda ambalaj materyalleri de yüksek oranda Bisfenol A içermektedir. Bisfenol A, endokrin sistem bozucu, toksik, teratojenik, karsinojenik, mutajenik ve östrojenik etkilere sahip olup insan ve hayvan sağlığını önemli ölçüde etkilemektedir. Gerek ülkemizde gerekse dünya otoritelerince günlük alımına sınırlamalar getirilen Bisfenol A'nın bugün üretimi 8 milyon tonu bulmuştur. Bu derlemede Bisfenol A'nın gıda sektöründe ve halk sağlığı açısından önemi hakkında bilgi verilecektir.

**Anahtar kelimeler:** Bisfenol A, Gıda, Halk Sağlığı, Kanser, Polikarbonat (PC) Plastikleri.

### Bisfenol A (BFA) and Its Importance in Terms of Public Health

**Abstract:** Bisphenol A is one of the most dangerous chemicals in terms of human health used in the production of polycarbonate plastics and epoxy resins. Food packaging materials such as some plastic and canned food used in the food industry also contain high levels of Bisphenol A. Bisphenol A has endocrine disruptor, toxic, teratogenic, carcinogenic, mutagenic and estrogenic effects and significantly affects human and animal health. The production of Bisphenol A, whose daily intake is limited by both our country and the world authorities, has reached 8 million tons today. In this review, information will be given about the importance of Bisphenol A in the food industry and in terms of public health.

**Key words:** Bisphenol A, Food, Public Health, Cancer, Polycarbonate (PC) Plastics.

#### 1. Giriş

Bisfenol A (BFA) ilk olarak 1891'de Rus kimyager Dianin tarafından sentezlenen ve dünya çapında en fazla kullanılan kimyasallardandır [1, 2]. BFA başka kimyasallarla kombine edilerek polikarbonat ve epoksi reçinelerin yapımında kullanılan bir kimyasaldır [1, 3]. Kimyager Charles Edward Dodds BFA'yı 1930'larda yapay bir östrojen olarak tanımlamıştır. BFA 1950'lerden beri sert polikarbonat plastik ve epoksi reçine yapımında kullanılmaktadır [4]. BFA 1960'lardan beri gıda paketlerinde kullanılmakta olup endokrin bozucu olarak 1996 yılında dikkat çekmeye başlamıştır [3, 5]. BFA'nın 1980'lerde yıllık üretimi 1 milyon tona ulaşırken günümüzde 8 milyon ton'a ulaşmıştır [4, 6-9]. Kanada Çevre Departmanı 2008 yılında Bisfenol A'yı toksik madde olarak ilan etmiştir [4].

Gıda endüstrisinin gelişmeye başlamasıyla gıda ambalajlarında kullanılan maddeler değişmiş olup, özellikle plastik ambalaj ve konserveler kullanılmaya başlanmıştır. Bazı plastik ambalajlar ve konserveler yüksek miktarda BFA denilen kimyasal bir madde içermektedirler [10].

Dünya genelinde, polikarbonat plastiklerin %70'inde ve epoksi reçinelerin %25'inin üretiminde BFA kullanılmaktadır [6]. Bunun yanı sıra diğer alanlarda yaygın olarak kullanılan ve üretimi fazla olan endüstriyel bir kimyasaldır [8-10]. Polikarbonatlar (PC), plastik yemek takımları, plastik poşetler, mikrodalga fırın kapları, gözlük camları, oyuncaklar, emzikler, darbeye dayanıklı güvenlik ekipmanları, kompakt diskler ve otomobil parçaları gibi ürünlerde bulunmaktadır. Epoksi reçineler elektronik cihazlar, yapışkanlar, epoksi-reçine esaslı boyalar, konserve yiyecek ve içecek kaplarının koruyucu kaplamalarında, içme suyu depolama tanklarında, şarap fıçısı kaplamalarında, zemin kaplamalarında ve bazı dental kompozitlerde kullanılmaktadır. BFA ayrıca sigara filtresi, medikal cihazlar, diş dolguları, mürekkepler, mikrodalga fırın kapları, kompakt diskler, kaplanmış teneke kutular, termal kağıtlar (kasa fişleri vs.) ve alev geciktiriciler gibi yiyeceklerle ilgili olmayan bazı alanlarda da kullanılır. Epoksi esaslı kaplama, yapıştırıcı, boya, elektronik cihaz ve mürekkeplerin aşınması sonucunda BFA, ev tozu içinde de yoğunlaşabilir [11-17].

BFA'nın dünyadaki üretimi 2006 yılında 3.9 milyon tondan fazladır. Dünyada özellikle ABD, Almanya, Hollanda ve Japonya gibi gıda endüstrisinin geliştiği ülkelerde BFA yüksek oranda üretilmektedir. 2003 yılında

\* Sorumlu yazar: [p.demir@firat.edu.tr](mailto:p.demir@firat.edu.tr). Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-8596-9261, <sup>2</sup> 0000-0002-0824-1672, <sup>3</sup> 0000-0002-3011-5592.

Amerika'da Bisfenol A kullanımı yaklaşık 862 bin ton olmuştur ve bu kullanımın 3/4'nün gıda ve içeceklerin muhafaza edildiği polikarbonat kapların yapımında kullanıldığı belirlenmiştir [6]. 2011 yılında küresel BFA üretimi yaklaşık 4,4 milyon tona ulaşmıştır. 2012 yılında 4,7 milyon ton sınırını aşarak 372.000 tonun biraz üzerinde büyümüştür. Asya, toplam üretim hacminin yaklaşık %53'lük payıyla baskın BFA üreticisidir. Avrupa %25, Kuzey Amerika ise %18 pazar payına sahiptir [18]. Küresel BFA Pazarı 2015-2019 raporuna göre, analistler küresel BFA pazarının yıllık ortalama büyüme oranının (Bileşik Yıllık Büyüme Oranı) 5,1 oranında büyüyeceğini tahmin etmiştir [19].

EFSA, 2014 yılında BFA'nın günlük tolere edilebilir alım miktarını 5 µg/kg vücut ağırlığı/gün olarak önermiştir. Ayrıca 2011 yılından itibaren AB, bebek biberonlarında BFA kullanımını yasaklayan kanunu yürürlüğe koymuştur. Yetişkinlerde ve çocuklarda BFA'ya maruziyetin en büyük kaynağı gıdadır [20]. Hayvanlar üzerinde yapılan çalışmalar, BFA'ya maruz kalma üreme, diyabet ve kardiyovasküler hastalıklar içeren çoklu sağlık problemlerine neden olabildiğini göstermiştir [21].

### 1.1. Bisfenol A'nın fiziksel ve kimyasal özellikleri

Bisfenol A (2, 2-bis (4-hidroksifenil) propan) aseton ve fenolün kombinasyonu ile oluşmaktadır [22]. BFA organik bir bileşiktir ve iki fenol halkasının metil köprüsü ile birbirine bağlanmasıyla oluşmaktadır [6, 23]. BFA, oda sıcaklığında katı beyaz kristal halde bulunur ve hafif fenolik kokuya sahiptir [24].

**Tablo 1.** Bisfenol A'nın fiziksel ve kimyasal özellikleri [6, 25-26].

Parametre	Değer
Kimyasal formülü	C <sub>15</sub> H <sub>16</sub> O <sub>2</sub> ; (CH <sub>3</sub> ) <sub>2</sub> C(C <sub>6</sub> H <sub>4</sub> OH) <sub>2</sub>
Moleküler ağırlığı	228.291 g/mol
Erime noktası	152-153°C
Kaynama noktası	360.5°C (760 mmHg)
Yanma noktası	212°C (açık kaplarda)
Yoğunluk	1.195 g/cm <sup>3</sup> (25°C)
Suda çözünürlük	300 mg/l (25°C)
Buhar basıncı	5.3x10 <sup>-6</sup> Pa

### 1.2. Bisfenol A'ya maruz kalınma yolları?

Genel nüfus, diyet (gıda, içme suyu) ve diyet dışı (toz, hava, termal kağıt, kozmetik, CD oyuncak vb.) kaynaklardan BFA'ya maruz kalabilir [19, 27-28]. Bisfenol A, en fazla gıda ve su ile vücuda alınır [4, 29]. Bisfenol A'nın konserve yiyeceklerin plastik iç kaplamasından gıdalara geçtiği bilinmektedir [30]. İnsan maruziyeti mesleki, çevresel ve gıda yoluyla olmaktadır [31].

Avrupa Gıda Güvenliği Ajansı (EFSA) tarafından 2015 yılında yayınlanan bilimsel görüşte, BFA'ya diyetle maruz kalma değerlendirilmesi yapılmıştır. Paketlenmemiş gıdalarla karşılaştırıldığında en yüksek BFA konsantrasyonları paketlenmiş ürünlerde (18.68 µg/kg) bulunmuştur. (1.5 µg/kg). BFA'nın özellikle paketlenmiş gıda kategorileri arasında anne sütü, kahvaltılık gevrekler, et, balık, baharatlar, yemek, atıştırmalık ve dondurma hazır gıdalar yer almaktadır. Bu kategoriler 30 µg/kg BFA'nın üzerindeki değerlere ulaşmaktadır. Paketlenmemiş gıdalar arasında en yüksek BFA konsantrasyonları sırasıyla 9.4 µg/kg ve 7.4 µg/kg BFA değerleriyle et ve balıkta saptanmıştır [32]. Özellikle anne sütü ile beslenmeyen, toz ve sıvı şeklinde muhafaza edilen sütler ile beslenen bebekler BFA'ya maruz kalan en duyarlı gruptur [33]. FAO/WHO [34], 0-6 aylık bebeklerin BFA migrasyonuna %81'i PC şişelerden, %19'u PC kapları veya epoksi ile kaplı paketli sıvı mamalardan, %1'i ise süt tozu formülleriyle temas halinde olan epoksi reçinesinden maruz kalınmaktadır. Kanada'da yapılan bir çalışmada konserve alkolsüz içeceklerin büyük çoğunluğunda BFA değerinin düşük olduğu fakat ölçülebilir düzeyde olduğu test edilmiştir [4, 35].

Yapılan bir çalışmada, taze ve konserve gıdalar, plastik ambalajlı gıdalar ile plastik ambalajda ve konserve olarak satılan kedi ve köpek mamalarından oluşan 105 örnekten 63'ünde BFA bulunmuştur [4]. Serbest BFA, termal kağıtlar, diş dolguları ve karbonsuz kopya kağıtlarında yüksek oranda bulunmuştur [4, 36-37]. Termal kağıtların en yaygın olarak kullanıldığı alanlar alışveriş fişleri, sinema biletleri, uçak biletleri ve etiketlerdir. İsveç'te yapılan bir çalışmada 13 termal kağıdın 11'inde 8-17 g/kg oranında BFA içerdiği tespit edilmiştir [4].



### 1.3. Bisfenol A Migrasyonunu Etkileyen Faktörler

Belirli koşullar altında ambalaj materyalinin üretiminde kullanılan kimyasalların temas ettiği gıda ile etkileşerek gıdaya göç etmesine migrasyon denir [38]. Migrasyon üzerinde, ambalaj malzemesinin dolaylı ya da direk olarak gıdaya teması, temas süresi ve sıcaklığı, gıda ile temas eden materyalin özelliği (örneğin; kâğıtlar için kâğıdın kalınlığı ve geçirgenliği), migrantın kimyasal özelliği (Buhar basıncı, polaritesi, moleküler büyüklüğü ve yapısı vb.), ambalaj materyalindeki migrantın başlangıçtaki konsantrasyonu, materyal ile temas eden gıdanın yapısı (pH, yağ içeriği), materyal ile temas eden bileşenler (gıda ya da uyarıcılar), temas eden yüzey alanının büyüklüğü etkili olmaktadır [38-40].

Kanada’da yapılan bir çalışmada [41], bebek biberonlarının mikrodalga fırında bekletilmesi ve birden fazla kullanılan bebek biberonlarının sıcak su ile doldurulması veya mikrodalga fırında ısıtılmasının BFA geçişini arttırdığı belirtilmiştir. Türkiye’de Hatay ilinde 2013 yılında konserve gıda ve içeceklerden BFA geçişi üzerine yapılan bir çalışmada gıdadaki glikoz ve NaCl konsantrasyonu ile BFA geçişi arasında güçlü bir korelasyonun olduğu gözlemlenmiş olup ayrıca bu ürünlerin son kullanım tarihi ile de pozitif yönde bir ilişki olduğu görülmüştür [42]. Kore Cumhuriyeti Çevre Bilimleri departmanı tarafından yapılan bir çalışmada (2009), bebek biberonlarının 10. kez tekrar kullanımı ile BFA migrasyon düzeyinin 1.1 ppb’ye kadar yükseldiği ve 60. kez tekrar kullanıma kadar bu seviyeyi koruduğu gözlemlenmiştir. Bu çalışmada 60. kez kullanımdan sonra, 100. kez tekrar kullanımda BFA migrasyon seviyesi hızlı bir şekilde 3.08 ppb seviyesine yükselmiştir. Migrasyon düzeyinin 80°C’nin üzerinde hızlı bir şekilde arttığı gözlemlenmiştir [43].

### 1.4. Bisfenol A’nın farmakokinetiği

İnsan çalışmalarında oluşan etik ve bireysel duyarlılığı belirleme ile ilgili sıkıntılardan ötürü BFA ile ilgili in vivo çalışmalar daha çok laboratuvar hayvanları üzerinde yapılmıştır [6, 10]. Yapılan bu deneylerde karaciğerin önemli rol aldığı gözlemlenmiştir (6). İnsanlarda oral yolla alınan BFA, gastrointestinal kanalda emildikten sonra daha çok CYP2C18 ve daha az olarak da CYP2C19 ve CYP2C9 enzimleri ile metabolize uğrar [1]. Daha sonra glukuronik asit ve sülfat ile konjuge olarak metabolitleri olan BFA glukuronit (BFAG) ve BFA sülfat’a dönüşür [6, 1]. Konjuge olmuş BFAG serbest BFA miktarını düşürür. Ama plasenta ve diğer dokularda yüksek konsantrasyonda bulunan beta-glukuronidaz enzimi ile dekonjugate olabilir [4].

BFA’nın yarılanma ömrü 6 saat olup [44] ve tamamına yakınının 42 saat içinde idrarla atıldığı belirtilmiştir [1]. Az sayıda gönüllüler üzerinde yapılan çalışmada BFA’nın %9.5’inin hiç değişmeden, %69.5 BFAG ve %21’inin BFA sülfat konjugatları şeklinde idrarla atıldığı gösterilmiştir [11]. İdrardan başka BFA’nın kan, fetüs, yağ dokusu, semen, plasenta, kolostrum, göbek kordonu kanı ve tükürükte de bulunduğu tespit edilmiştir [1]. BFA lipofilik karaktere sahiptir. Bundan dolayı yağ dokularında birikebilir [45].

Bazı hayvan testlerinde BFA’nın oral yolla alındıktan sonra kolay bir şekilde plasenta bariyerini geçerek fetüsü etkilediği görülmüştür. Ayrıca Japonya’da gebe fareler ve gebe maymunlar üzerinde yapılan bir çalışmada farelere 100 mg/kg maymunlara ise 50 mg/kg dozunda BFA deri altı yolla uygulanmış. Farelerde 30 dakika sonra yavrunun beyin, serum, karaciğer, uterus ve testis gibi organlarında BFA saptanmıştır. Maymunlarda ise 1 saat sonra beyin, karaciğer ve böbrek gibi iç organlarında BFA tespit edilmiştir [6]. Yapılan son araştırmalarda, insanların ter analizlerinde BFA biyoakümülyasyonun olduğu saptanmıştır. Yirmi kişi üzerinde gerçekleştirilen bir çalışmada ter, kan ve idrar örnekleri incelenmiş, 16 kişide serum ve idrar örneklerinde BFA’ya rastlanılmamasına rağmen ter örneklerinde BFA’ya rastlanılmıştır [45].

### 1.5. Bisfenol A’nın insan sağlığı üzerine etkisi

FAO/WHO’nun bildirdiği bir çalışmada; dişi üreme dokusu, göğüs kanseri dokusu, erkek üreme dokusu, hipofiz, sinir hücreleri ve pankreatik/yağ dokusu, bağışıklık hücreleri ve embriyonik kültürlerde BFA’ya çok düşük miktarda rastlanılmıştır [10]. Bisfenol A’nın dişi cinsiyet hormonuna çok benzer etki gösteren (ksenöstrojen) sentetik bir yapıya sahip olduğu bilinmektedir [37]. Hayvanlar üzerindeki çalışmalara dayanılarak, yüksek dozlarda (TDI değerinden 100 kat daha fazla dozlarda) BFA’nın karaciğer ve böbrekler üzerine yan etkilere sebep olabileceği ihtimali bulunmaktadır. Aynı zamanda kemirgenlerin meme bezleri üzerinde etki yapma ihtimali bulunmaktadır [46]. BFA’nın dişi ve erkek üreme, sinir, immun, metabolik ve kardiyovasküler sistem üzerine toksikolojik etkisi mevcuttur. Aynı zamanda kanser oluşumuna da etkisi söz konusudur [37, 46]. Ayrıca BFA gebelik esnasında maruz kalındığında bu kimyasallar; hamileliğin hangi döneminde maruz kalındığına bağlı olarak hem anneyi hem de fetal gelişimi etkileyebilmektedir [47]. Son zamanlarda yoğun çalışmalar BFA’nın toksik, teratojenik, karsinojenik ve özellikle östrojenik mekanizmayı etkilediği saptanmıştır [45].

### 1.6. Bisfenol A hangi tür plastiklerde bulunur?

Paketlemede kullanılan plastikler polietilen Tereftalat (PET, PETE), düşük ve yüksek dansiteli polietilen (LDPE, HDPE), polivinil klorürler (V, PVC), polipropilen (PP), stirenin polimerizasyonu ile elde edilen polistiren ve stirenbutadien kauçukları (PS), polikarbonatlar (PC) ve diğerleri olup, bunlardan geri dönüşüm kodları 1, 2, 4, 5, 6 olanlar polimerizasyon ya da paketleme formunda BFA içermedikleri için gıdalara ve içeceklere BFA geçişi olmaz. Ancak geri dönüşüm kodu 3 ve 7 olanlar BFA içermektedirler [37].

### 1.7. Bisfenol A ve metabolitlerinin analizinde kullanılan analitik yöntemler

BFA ve metabolitlerinin ölçülmesinde çok sayıda yöntem kullanılmaktadır, bunlardan bazıları Enzyme-Linked Immunosorbent Assay (ELISA), Yüksek Performans Sıvı Kromatografisi (HPLC), Ultra Performance Liquid Chromatography (UHPLC) ve Gaz Kromatografisi Kütle Spektrometresi (GC/MS) olarak adlandırılabilir. BFA Analizi için hem single hem de triple quadrupole kütle spektrometresine dayanan (LC/MS-MS) daha duyarlı yöntemler de bulunmaktadır [1]. HPLC ve farklı belirleme yöntemleri (UV, MS, fluorescence vs.), Gas kromatografi (GC) MS ile birlikte, elektrokimyasal metodlar ve Misel elektrokinetik kromatografi (MEKC) yöntemleri ile tespit edildikleri belirtilmektedir [48]. Solid Phase Extraction (SPE), Liquid-Liquid Extraction (LLE), Solid Phase Microextraction (SPME), Stir Bar Sorptive Extraction (SBSE), Ultrasound-Assisted Dispersive Liquid-Liquid Microextraction (UA-DLLME), Pressurised Liquid Extraction (PLE) and Quick Easy Cheap Effective Rugged Safe (QUECHERS) en yaygın kullanılan ekstraksiyon yöntemleridir [49].

### 1.8. Bisfenol A ile ilgili Türkiye ve Dünyada yapılan yasal düzenlemeler

Bisfenol A, dünyada ilk kez Kanada'da Nisan 2008 tarihinde insan ve çevre için toksik madde olarak kabul görmüştür. Avrupa'da ve Türkiye'de 2011 yılında, Amerika Birleşik Devletlerinde ise 2012 yılında bebek biberonlarında kullanımı tamamen yasaklanmıştır [1].

EFSA'nın 2016 Ekim ayında yayımladığı yeni verilere göre yapılan hayvan deneylerinde BFA'nın immun sistem üzerine etkili olabileceği kanıtlanmış, ancak bu kanıtları insan sağlığı açısından yorumlamak için bu çalışmalar çok sınırlıdır. Ocak 2015'te EFSA BFA maruziyeti ve toksisitesini ele alan kapsamlı yeni bir değerlendirme yayınlanmıştır. Bu değerlendirmede günlük tolere edilebilir (TDI) BFA miktarı'nı 50 µg/kg'dan 4 µg/kg vücut ağırlığı/gün olarak güncellenmiştir [45]. EFSA bu TDI miktarı geçici olarak hükme bağlamış olup ABD Ulusal Programı tarafından 2 yıl içerisinde bitmesi planlanan bir çalışmaya bağlı olarak 2017 yılında tekrar değerlendirileceği belirtilmiştir [14].

Fransa Hükümeti gıda ile temas eden madde yasalarına ilişkin 2 rehber belge yayınlamıştır. Birincisi 24 Aralık 2012 tarih ve 2012-1442 sayılı Kanunla değiştirilen 2010-729 sayılı Kanun. İkincisi ise 21 Temmuz 2014 tarih ve 2014-108 sayılı kanundur. Fransa'da BFA yasağı 2 aşamada uygulanmıştır. 1. aşama 1 Ocak 2013 tarihinden itibaren geçerli olmak üzere yürürlüğe giren; 3 yaşına kadar olan çocukların kullanımına yönelik tüm mutfak gereçleri, gıda ambalajları ve tüm kaplarda BFA kullanımının yasaklanması, 2. aşama ise 1 Ocak 2015 tarihinden itibaren yürürlüğe girmiş olan bu aşamada tüm gıda ambalajları, kaplar ve mutfak gereçlerinde BFA'nın yasaklandığı karardır [50].

EFSA'nın 2015 yılında BFA risk değerlendirme açısından ana sonuçları şunlardır [45]:

- 1) BFA tüketiciler için sağlık riski oluşturmaz. Çünkü bu kimyasala maruz kalma miktarı zarar veremeyecek kadar düşüktür.
- 2) BFA'ya maruz kalma miktarı 4 µg/kg vücut ağırlığı/gün'e düşürülmüştür. Bu oran daha önceki orandan 12.5 kat daha düşüktür.
- 3) Hem yiyecek hem de yiyecek dışı kaynaklardan alınan maksimum BFA miktarı, bütün yaş gruplarında TDI oranından 3-5 kat daha düşüktür.
- 4) Bütün yaş gruplarında gıda ile alınan BFA oranı 15 kat daha düşüktür. Oysa daha önce 4 kat olarak tahmin edilmekteydi.
- 5) Hayvanlar üzerinde yapılan çalışmalara dayanarak yüksek dozda BFA (yaklaşık TDI'nın 100 katı) böbrek ve karaciğerde olumsuz etkilere neden olabilir. Kemirgenlerin meme bezleri üzerinde de etkileri muhtemel olarak değerlendirilmektedir.
- 6) Ratlarda doğum öncesi ve sonrası yapılan uzun süreli bir çalışma ile BFA'ya maruz kalmanın sağlık üzerindeki potansiyel etkilerinin saptanmasında yardımcı olabilir.
- 7) BFA'nın meme bezi, üreme, metabolik, nörodavranışsal ve bağışıklık sistemi üzerindeki potansiyel sağlık etkilerini çevreleyen belirsizlikler, TDI'ya göre hesaplanmış ve ölçülmüştür.

Haziran 2010'da Avustralya hükümeti BFA içeren polikarbonat bebek şişelerinin bayiler tarafından gönüllü olarak satışının iptal edildiğini duyurmuştur. Ve yine (FSANZ), 2016 yılında 24. Avustralya Toplam Diyet Araştırmasının 2. bölümünü yayınladığı bildirilmiştir. Bu bölümde kimyasallar içeren ambalaj materyalleri taranmış ve sonuç olarak BFA'nın hiç bulunmadığı ya da güvenlik seviyesinin oldukça altında kaldığı belirtilmiştir [51].

BFA'nın gıda ile temas etmesi amaçlanan verniklerde ve kaplamalarda kullanılması ve 10/2011 Sayılı (AB) Tüzüğü'nün revize edilmesi sonucu bu maddenin plastik food contacts material FCM'lerde kullanımına ilişkin olarak, malzeme ve eşyalara uygulanan verniklerden veya kaplamalardan BFA geçişi, 0.05 mg BFA/kg gıda specific migration limitini (SML) geçmeyecektir. Bebekler ve küçük çocuklar için gıda veya özel beslenme ürünleri ile temas etmesi amaçlandığında bu malzemelerden göçe izin verilmeyecektir [51]. Amerika da ise BFA ile ilgili bütün yasal düzenlemeler FDA'ya verilmiştir. BFA 1960'lı yıllardan beri FDA'nın çalışma kapsamına alınmıştır [4, 21]. EFSA, yalnızca BFA için 4 µg/kg/vücut ağırlığı/günlük geçici bir tolere edilebilir günlük alım (tTDI) belirlemiştir ve Avrupa Birliği, BPA'ya spesifik migrasyon limitini (SML) gıda ile temas eden malzemelere uygulanan vernik veya kaplamalardan gıda içine/üzerine 0,05 mg/kg olarak sabitleyerek kullanımına ilişkin kısıtlamalar getirmiştir [45, 51].

Türkiye'de ise BFA ile ilgili yapılan düzenlemeler Avrupa Birliği direktiflerine uyumlu olup [1, 53], son olarak 2013/34 sayılı tebliğ 10/2011/EU sayılı "Gıda ile Temas Eden Plastik Madde ve Malzemeler Üzerine Komisyon Tüzüğü dikkate alınarak Avrupa Birliği'ne uyum çerçevesinde hazırlanmıştır. Bu Tebliğ'de "TGK-Bebek Formülleri ve TGK-Devam Formülleri Tebliğlerinde bebek olarak tanımlanan tüketici grubu için kullanılan, polikarbonat madde ve malzemelerin üretiminde kullanılamaz" tanımlaması yer almıştır [53]. Türk Gıda Kodeksinin 30989 sayılı resmi gazetede yer alan 2019/44 numaralı Gıda Maddeleri ile Temasta Bulunan Plastik Madde ve Malzemeler Tebliği'nde Bisfenol A'nın spesifik migrasyon limiti (SML) 0,05 mg/kg olarak belirtilmiştir [54].

### 1.9. Bisphenol-A (BFA)'dan etkilenmemek için neler yapılabilir?

Hayvanlar üzerinde yapılan bazı çalışmalar bebekler ve çocukların BFA'ya karşı en duyarlı grup olduklarını göstermiştir. Sağlık otoriteleri maruziyet miktarının minimize edilmesi için aşağıdaki hususlara dikkat edilmesini önermektedirler [37, 54].

- Polikarbon plastik gıda kaplarını mikrodalga fırınlarda kullanmamalı. Çünkü yüksek sıcaklık gıdalara BFA geçişini artırır.
- Plastik kapların geri dönüşüm kodlarına dikkat edilmelidir. 3 ve 7 no'lu kodları barındıran plastikleri kullanmaktan kaçınılmalıdır.
- Konserve gıda tüketimi azaltılmalı,
- Mümkün olduğunca özellikle sıcak yemekler ve sıvılar için cam, porselen ve paslanmaz çelik kaplar tercih edilmeli,
- BFA içeren bebek biberonları kullanılmamalı,
- Polikarbonatlı plastikler bulaşık makinasında yıkanmamalıdır,
- Rezin esaslı diş dolgu maddeleri mümkünse hamilelik döneminde kullanılmamalı.

## 2. Sonuç

Sanayiinin gelişmeye başlaması ve artan dünya nüfusunun sonucu olarak plastikler ve konserveler gıda endüstrisinin vazgeçilmezleri arasına girmiştir. Bu plastik ve konserveler gıdaların ambalajlamasında ve raf ömrünün uzatılması amacıyla yaygın olarak kullanılmaktadır. Ama bu plastik ve konserve ambalajların yapısında BFA (Bisfenol A) gibi östrojenik, kanserojenik, mutajenik, kalp sağlığı, bağışıklık sistemi, diyabet ve diğer birçok olumsuz etkiye sahip kimyasallar içermektedirler. Özellikle bebek biberonları ve çocuk oyuncakları bu kimyasal bakımından tehlike oluşturmaktadırlar. Ayrıca BFA bu ürünler hariç CD, kasa fişleri gibi birçok üründe de bulunmaktadır. BFA'nın ambalajdan gıdaya geçişi birçok faktöre bağlıdır. Özellikle gıdanın türüne, temas süresi, yiyecek veya içeceğin sıcaklığına, depolama koşullarına, pH, tuz oranı gibi faktörlere bağlıdır. Başta EFSA ve FDA gibi kurumların öncülük ettiği birçok çalışma sonucunda günlük alınan ortalama BFA miktarı TDI değerinin altında kalmaktadır. Ancak uzun süre maruz kalımı dokularda birikime neden olabileceğinden halk sağlığı bakımından ciddi sağlık sorunlarına neden olabileceği vurgulanmaktadır.

## Kaynaklar

- [1] Battal D. Mersin ilinde yaşayan bireylerdeki Bisfenol A düzeyinin belirlenmesi. Yüksek Lisans tezi, Mersin Üniversitesi Sağlık Bilimleri Enstitüsü, Mersin, Türkiye, 2012.
- [2] Denizli A, Yavuz H. İnsan Sağlığı için olası bir tehdit: Bisfenol A Bilim Teknik Dergisi Eylül 2015; 40-42.
- [3] Anonymous, 2017. National Institute of Advanced Industrial Science and Technology Bisphenol A (BFA) Risk Assessment Document. <https://unit.aist.go.jp/riss/crm/mainmenu.20.04.2017>.
- [4] Vogel SA. "The politics of plastics: the making and unmaking of bisphenol a "safety". American Journal of Public Health 2009; 99: 559-566.
- [5] Food Drug and Administration FDA. Bisphenol A. <https://www.fda.gov/food/ingredientspackaginglabeling/foodadditivesingredients> 20.04.2017.
- [6] Er B. Ton balığı konservelerinde katı faz ekstraksiyon ve hplc metodu ile Bisfenol A varlığının incelenmesi. Doktora Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara, Türkiye, 2010.
- [7] Tzatzarakis MN, Karzi V, Vakonaki E, et al. Bisphenol A in soft drinks and canned foods and data evaluation, Food Additives & Contaminants 2016; 1-6.
- [8] Gorini F, Bustaffa E, Coi A, Iervasi G, Bianchi F, Bisphenols as environmental triggers of thyroid dysfunction: clues and evidence. Int. J. Environ. Res. Public Health 2020; 17(8): 2654. <https://doi.org/10.3390/ijerph17082654>.
- [9] Kim JJ, Kumar S, Kumar V, Lee YM, Kim YS, Kumar V. Bisphenols as a legacy pollutant, and their effects on organ vulnerability. Int. J. Environ. Res. Public Health 2020; 17(1): 112. <https://doi.org/10.3390/ijerph17010112>.
- [10] Anonymous, 2010. Report of Joint FAO/WHO Expert Meeting 2-5 November 2010 and Report of Stakeholder Meeting on Bisphenol A 1 November 2010 Ottawa, Canada Toxicological and Health Aspects of Bisphenol A 2010.
- [11] EFSA (European Food Safety Authority), 2013. Public Consultation on the Draft Opinion on Bisphenol A (BFA) - Exposure Assessment [314 Pages]. European Food Safety Authority [Accessed 23 August 2013]. <http://www.efsa.europa.eu/en/consultations/call/130725.Htm>.
- [12] FAO/WHO (Food and Agriculture Organization and World Health Organization), 2011.
- [13] NTP (US National Toxicology Program), 2008. In: CERHR (Ed.), NTP-CERHR Monograph on the Potential Human Reproductive and Developmental Effects of Bisphenol.
- [14] EFSA, Bisphenol A <http://www.efsa.europa.eu/en/topics/topic/bisphenol>. Accessed date: 27/10/2021.
- [15] National Pivonello C, Muscogiuri G, Nardone A, Garifalos F, Provisiero DP, Verde N, de Angelis C, Conforti A, Pisco M, Auriemma RS, Colao A, Pivonello R. Bisphenol A: an emerging threat to female fertility. Reproductive Biology and Endocrinology 2020 Mar 14;18(1):22. <https://doi.org/10.1186/s12958-019-0558-8>.
- [16] Wu W, Li M, Liu A, Wu C, Li D, Deng Q, Zhang B, Du J, Gao X, Hong Y, 2020. Bisphenol A and the risk of obesity a systematic review with meta-analysis of the epidemiological evidence. Dose Response 2020; 18 (2): 1-10. <https://doi.org/10.1177/1559325820916949>.
- [17] Zwierello W, Maruszewska A, Skorka-Majewicz M, Goschorska M, Baranowska-Bosiacka I, Dec K, Styburski D, Nowakowska A, Gutowska I. The influence of polyphenols on metabolic disorders caused by compounds released from plastics - Review. Chemosphere 2020; 240: 124901.
- [18] Merchant Research (2017). Bisphenol a (BFA): World market outlook and forecast up to 2027. <https://mcgroup.co.uk/researches/bisphenol-a-BFA>, Accessed date: 27/10/2021.
- [19] Vilarinho F, Sendón R, Van der Kellen A, Vaz MF, Sanches Silva A. Bisphenol A in food as a result of its migration from food packaging. Trends in Food Science & Technology 2019; 91: 33-65.
- [20] Pedersen, G. A., Hvilsted, S., & Petersen, J. H. (2015). Migration of bisphenol A from polycarbonate plastic of different qualities: Danish Ministry of the Environment. Miljøprojekter No. 1710 <http://mst.dk/service/publikationer/publikationsarkiv/2015/maj/migration-of-bisphenol-a-from-polycarbonate-plastic>.
- [21] FDA, Bisphenol A (BFA): Use in Food Contact Application. <https://www.fda.gov/newsevents/publichealth/Cus/ucm064437.htm>. Accessed date: 20.04.2017.
- [22] Lorber M, Schechter A, Paepke O, Shropshire W, Christensen K, Birnbaum L. Exposure assessment of adult intake of bisphenol A (BFA) with emphasis on canned food dietary exposures. Environment International 2015a; 77: 55-62. <https://doi.org/10.1016/j.envint.2015.01.008>.
- [23] Kang JH, Katayama Y, Kondo F. Biodegradation or Metabolism of Bisphenol A from mikroorganism to mammals. Toxicology 2006; 217: 81-90.
- [24] Nakanishi J, Miyamoto KI, Kawasaki H. Bisphenol A Risk Assessment Document, AIST Risk Assessment Document 2007; Series No. 4.
- [25] National Center for Biotechnology Information. PubChem Compound Database; CID=6623, <https://pubchem.ncbi.nlm.nih.gov/compound/6623>. Accessed date: 27/10/2021.
- [26] Anonymous, Occupational Safety and Health Administration Bisphenol A [https://www.osha.gov/dts/chemicalsampling/data/CH\\_221000.html](https://www.osha.gov/dts/chemicalsampling/data/CH_221000.html). Accessed date: 27/10/2021.
- [27] Cwiek-Ludwicka K. Bisphenol A (BFA) in food contact materials - new scientific opinion from EFSA regarding public health risk. Roczniki Państwowego Zakładu Higieny 2015: 66(4):299-307.
- [28] Tox Town. Bisfenol A, [https://toxtown.nlm.nih.gov/text\\_version/chemicals.php](https://toxtown.nlm.nih.gov/text_version/chemicals.php), Accessed date: 27/10/2021.
- [29] The Environmental Working Group, Bisphenol A- toxic plastics chemical in canned food <http://www.ewg.org/research/bisphenol>. Accessed date: 27/10/2021.

- [30] Pielichowski K, Michalowski S. Nanostructured flame retardants: Performance, toxicity, and environmental impact. *Polymer Nanocomposites and Other Materials Containing Nanoparticles* 2014; 251-277. <https://doi.org/10.1533/9780857096678.3.251>. Accessed date: 27/10/2021.
- [31] EFSA CEF Panel (EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids), 2015. Scientific opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs: Executive summary. *EFSA Journal* 2015a; 13(1): 3978.
- [32] Health Canada, Bureau of Chemical Safety, Food Directorate, Health Products and Food Branch, Survey of Bisphenol A in Canned Drink Products (2009): 1-8.
- [33] Hoekstra EJ, Simoneau C. Release of Bisphenol A from Polycarbonate-A Review, *Critical Reviews in Food Science and Nutrition* 2013; 53(4): 386-402.
- [34] Joint Fao/WHO Expert Meeting (2010). Toxicological and health aspects of bisphenol a Joint. Ottawa, Canada: World Health Organization.
- [35] Janet R. Concerned about BFA: Check your receipts <https://www.sciencenews.org/blog/science-public> 20.04.2017.
- [36] Akyüz S, Yarat A, Egil E. Bisfenol A içerikli dental materyallere güncel yaklaşım. *Marmara Üniversitesi Dış Hekimliği Fakültesi MÜSBED* 2011; 1(3): 190-195.
- [37] Altuntaş Ü. Türkiye’de satışı sunulan bazı gıdalarda ambalaj materyallerinden migrasyonun ölçülmesi. Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, Türkiye, 2014.
- [38] Yavuz HT. Migration of dipropylene and tripropylene glycol diacrylate from packaging materials and screening of potential risks in paper packagings. MSc, Istanbul Technical University Graduate School of Science Engineering And Technology, İstanbul, Turkey, 2013.
- [39] Barnes KA, Sinclair CR, Watson DH. *Chemical Migration and Food Contact Materials* 2007; Chapter 1: 1-12.
- [40] Cao XL, Coriveau J. Migration of Bisphenol A from polycarbonate baby and water bottles into water under severe condition. *J. Agric. Food Chem.* 2008; 56: 6378-6381.
- [41] Sungur Ş. Köroğlu M, Özkan A. Determination of bisphenol a migrating from canned food and beverages in markets. *Food Chemistry* 2014; 142: 87-91.
- [42] Nam SH, Seo YM, Kim MG, Bisphenol A Migration from polycarbonate baby bottle with repeated use. *Chemosphere* 2010; 79: 949-952.
- [43] Ye X, Wong LY, Bishop AM, et al. Variability of Urinary Concentrations of Bisphenol A in Spot Samples, First Morning Voids, and 24-Hour Collections *Environ Health Perspect* 2011; 119: 983-988.
- [44] Michalowicz J. Bisphenol A-Sources toxicity and Biotransformation. *Environmental Toxicology and Pharmacology* 2014; 37: 738-758.
- [45] EFSA (2015b). Scientific Opinion on the risks to public health related to the presence of bisphenol A (BFA) in foodstuffs. *EFSA Journal*, 13(1), 1-621. <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2015.3978>. Accessed date: 27/10/2021.
- [46] Rykowska I, Wasiak W. Properties, Threats and methods of analysis of bisphenol A and its derivatives. *Acta Chromatographica* 2006; 16: 7-20.
- [47] Yıldırım Y, Ertaş Onmaz N, Gönülalan Z et al. Bisfenoller ve fitalatların halk sağlığı üzerine etkileri. *Erciyes Üniv Vet Fak Derg* 2020;17(1): 68-75. <https://doi.org/10.32707/ercivet.655008>.
- [48] Ferrer E, Santoni E, Vittori S, Font G, Mañes J, Sagratini G. Simultaneous determination of bisphenol A, Octylphenol, and nonylphenol by pressurised liquid extraction and liquid chromatography-tandem mass spectrometry in powdered milk and infant formulas. *Food Chemistry* 2011; 126(1): 360-367. <https://doi.org/10.1016/j.foodchem.2010.10.098>.
- [49] Keller and Heckman LLP, France Releases Guidance on BFA Ban and Food-Contact Legislation, <http://www.packaginglaw.com/news/france-releases-guidance-BFA-ban-and-food-contact-legislation>. Accessed date: 27/10/2021.
- [50] Food Standards. Australia New Zealand, Regulation and monitoring of BFA <http://www.foodstandards.gov.au/consumer/chemicals/BFA/pages/regulationandmonitor5377.aspx>. Accessed date: 27/10/2021.
- [51] Commission Regulation (EU), 2018. Commission Regulation (EU) 2018/213 of 12 February 2018 on the Use of Bisphenol A in Varnishes and Coatings Intended to Come into Contact with Food and Amending Commission Regulation (EU) No 10/2011 as Regards the Use of that Substance in Plastic Food Contact Materials. Accessed date: 27/10/2021.
- [52] T.C. Gıda Tarım ve Hayvancılık Bakanlığı, Türk Gıda Kodeksi Gıda İle Temas Eden Madde ve Malzemelere Dair Yönetmelik, 5 Nisan 2018, Sayı: 30382. <https://www.resmigazete.gov.tr/eskiler/2018/04/20180405-2.htm>. Accessed date: 27/10/2021.
- [53] T.C. Gıda Tarım ve Hayvancılık Bakanlığı, Gıda Kodeksi Bebek Formülleri Ve Devam Formülleri Tebliği (Tebliğ No: 2019/14). <https://www.resmigazete.gov.tr/eskiler/2019/07/20190702-5.htm>. Accessed date: 27/10/2021.
- [54] <https://www.resmigazete.gov.tr/eskiler/2019/12/20191225M1-10-1.pdf> 29/06/2021. Accessed date: 27/10/2021.



## Türkçe Ses Kayıt Verilerinin CountVectorizer ve TF-IDFVectorizer Yöntemleri ile BERT Modelleri Olarak Google Colab Platformunda ve RapidMiner’da Makine Öğrenmesi Algoritmalarıyla Analizi

Abdülkadir TEPECİK<sup>1\*</sup>, Engin DEMİR<sup>2</sup>

<sup>1,2</sup> Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Yalova Üniversitesi, Yalova, Türkiye

\*<sup>1</sup> atepecik@yalova.edu.tr, <sup>2</sup> engindmr7@gmail.com

(Geliş/Received: 11/09/2021;

Kabul/Accepted: 01/01/2022)

**Öz:** Duygu, insanın ruh halinde içsel ve çevresindeki etkilerle etkileşiminden doğan fiziksel değişimlerdir. Bireyler duygularını, beden dilinin yanı sıra sesli iletişim vasıtalarıyla da diğer bireylere aktarabilirler. Özellikle beden dilinin yetersiz olduğu durum ve zamanlarda bireyler için sesli iletişim önem kazanmaktadır. Çalışmamızda da Türkçe ses kayıtlarını içeren veri seti üzerinde Python programlama dili aracılığıyla öncelikle verilerin duygu etiketlerinin tespiti yapılmış olup, sonrasında literatür çalışmalarında en çok kullanılan beş makine öğrenmesi algoritmasıyla analizler gerçekleştirilmiştir. Belirlenen metriklerle yapılan analizler hem RapidMiner hem de Python programlama dili aracılığıyla gerçekleştirilmiştir. Çalışmada, Python programlama dili aracılığıyla yapılan analizlerde hem CountVectorizer hem de TF-IDFVectorizer yöntemleri, RapidMiner ile yapılan analizlerde TF-IDFVectorizer yöntemi kullanılmıştır. Sonuç kısmında ise Python programlama dilinde en iyi doğruluk oranını %70 oranla Naive Bayes makine öğrenmesi algoritması CountVectorizer yöntemiyle elde etmiştir. RapidMiner’da ise en iyi doğruluk oranını %69,60 oranla Destek Vektör Makinesi makine öğrenmesi algoritması elde etmiştir. Çalışmamızla beraber ortaya yeni bir Türkçe duygu veri seti çıkmıştır. Ayrıca çalışmamız Türkçe ses kayıtlarından elde edilen verilerin BERT modeli ile duygu tespiti yapılan özgün bir çalışmadır.

**Anahtar kelimeler:** Türkçe Ses Kayıtları, Türkçe Duygu Analizi, RapidMiner, Python, Makine-Öğrenmesi.

### Analysis of Turkish Voice Recordings Data with CountVectorizer and TF-IDF Vectorization Methods as BERT Models on Google Colab Platform and RapidMiner with Machine Learning Algorithms

**Abstract:** Emotions are physical changes in a person's mood resulting from his interaction with internal and environmental influences. Individuals can convey their feelings to other individuals by means of voice communication as well as body language. Especially in situations and times when body language is insufficient, voice communication becomes important for individuals. In our study, the emotion labels of the data were firstly determined by the Python programming language on the data set containing the Turkish voice recordings, and then the analyses were carried out with the five most used machine learning algorithms in the literature studies. The analyses made with the determined metrics were carried out using both RapidMiner and Python programming language. In the study, both CountVectorizer and TF-IDF vectorization methods were used in analyses performed through the Python programming language, and TF-IDF vectorization method was used in analyses performed with RapidMiner. As a result, the best accuracy rate in the Python programming language was achieved by the Naive Bayes machine learning algorithm CountVectorizer method with 70%. At RapidMiner, the Support Vector Machine machine learning algorithm achieved the best accuracy rate of 69.60%.with our study, a new Turkish emotion dataset has emerged. In addition, our study is an original study in which emotion detection is made with the BERT model of the data obtained from Turkish audio recordings.

**Key words:** Turkish Audio Recordings, Turkish Sentiment Analysis, RapidMiner, Python, Machine-Learning.

#### 1. Giriş

İnsanların ruhsal etkenlerle ve çevrelerindeki olaylara göre verdikleri tepkilerin tanımına his veya duygu denir [1]. Duyguyu ifade etmenin birçok yolu vardır. Bunlardan biri beden dilinin yetersiz kaldığı durumlardaki ifade yöntemi olan sesteki duygu ifadesidir.

İnsanların olumlu ve olumsuz duygularını rahatlıkla ifade edebilmeyi imkân buldukları andan itibaren duygu analizine ilgi gün geçtikçe çoğalmış ve bu konudaki çalışma sayısı artmıştır. Duygu tespiti, duygu analizi gibi terimler, makine öğrenmesiyle beraber daha çok ön plana çıkmış ve veri bilimi ile ilgili araştırma alanlarının odak noktası haline gelmiştir. Duygu tespiti sonrası makine öğrenmesi algoritmaları ile yapılan analizler, yapılacak olan birçok çalışma için de yeni fikirlerin ortaya çıkmasına katkı sağlamıştır.

\* Sorumlu yazar: [atepecik@yalova.edu.tr](mailto:atepecik@yalova.edu.tr). Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-3842-7873, <sup>2</sup> 0000-0002-4546-3581

Makine öğrenmesi literatürde, *insan zekâsını ve algısını taklit ederken, kişinin yorumlayıp elle gireceği kurallara ihtiyaç duymayan algoritmalar bütünü* olarak ifade edilmektedir [2].

Yapılan çalışmalarda; S. Tuzcu, çevrimiçi bir kitap satış sitesinin kullanıcı yorumları üzerinde duygu analizi için öncelikle Python programlama dili kullanarak Çok Katmanlı Algılayıcı (Multi-Layer Perceptron MLP) algoritması kullanmayı amaçlamıştır. Çalışmasında, Naive Bayes (NB), Destek Vektör Makinesi (DVM) ve Lojistik Regresyon (LR) algoritmaları kullanılmış ve sonuçları karşılaştırılmıştır. Sonuçlarda; DVM, pozitif yorumları sınıflandırmada tüm algoritmalar içinde en iyi neticeyi vermiş fakat negatif yorumları sınıflandırmada diğer algoritmalarından oldukça geride kalmıştır. NB ise bu üç algoritmadan farklı olarak negatif yorumları pozitif yorumlara göre daha iyi sınıflandırmış fakat genel sınıflandırma başarısı bu algoritmaların altında kalmıştır [3].

J. Gondohanindijo ve arkadaşları, duygusal konuşmayı tanıma problemini çözmeyi, ses işleme yoluyla sınıflandırmayı, duyguların sınıflandırılmasına dayanan ses kalıplarını tanımayı ve kullanılan bazı sınıflandırıcıların doğruluk düzeyini karşılaştırmayı amaçlamışlardır[4]. Çalışmada 639 veriden oluşan veri seti üzerinde DVM, Rastgele Orman (RO), NB ve Sinir Ağı sınıflandırıcıları kullanılmıştır. DVM sınıflandırıcısının en iyi doğruluk ve hassasiyet seviyesine sahip olduğu ve F1-skorunda diğer üç sınıflandırıcıya göre daha iyi olduğu belirtilmiştir.

O. M. Nezami ve arkadaşları, Farsça için ShEMO adı verilen duygusal konuşma veri tabanı sunmuşlardır. Veri tabanında sınıflandırıcılar karşılaştırılarak, çıkan sonuçlar diğer dillerle kıyaslanmıştır[5]. Veri tabanı, 3 saat 25 dakikalık konuşma verilerine eşdeğer 3000 yarı doğal ifadeyi ve beş duyguyu içermektedir. Çalışmada, sınıflandırma algoritmalarından DVM, K-En Yakın Komşu (KEYK) ve Karar Ağacı (KA) kullanılmıştır. DVM modelinin en iyi sonuçları verdiği, gelecekteki çalışmalar için Gizli Markov Modeli ve Derin Sinir Ağları gibi sınıflandırma yöntemlerini kullanarak daha etkili bir tespit elde etmeyi amaçladıklarını belirtmişlerdir.

R. Matin ve D. Valles, otizm spektrum bozukluğu olan çocuklara, sosyal etkileşimdeki duyguları tanımlamak için bir konuşma duygu tanıma modeli geliştirmeyi amaçlamışlardır[6]. DVM modeline dayalı önerilen model Python programlama dili ile geliştirilmiştir. Çalışmada RAVDESS veri setiyle beraber DVM, Karar Ağacı ve LR sınıflandırma algoritmaları kullanılmıştır. Yapılan deneylerin sonucunda DVM modelinin %77 oranda doğrulukta ve önerilen modelin arka plan gürültüsü ile ses için iyi sınıflandırma doğruluğu sağladığı belirtilmiştir.

S. Dani ve arkadaşları, seslerdeki duyguyu tespit etmeyi amaçlamışlardır[7]. KEYK ve Karar Ağacı makine öğrenmesi tekniklerini, Toronto Duygusal Konuşma Seti (TESS) adlı veri seti üzerinde 7 duygu için uygulamışlardır. KEYK tekniğinin %98, Karar Ağaçlarının %92 ve Ekstra-Ağaç sınıflandırıcısının %99 doğruluk oranı sağladığını belirtmişlerdir.

M. G. Pinto ve arkadaşları, sesteki duygu analizinin tespiti için Konvolüsyonel Sinir Ağlarına dayanan bir sınıflandırma modeli sunmuşlardır[8]. Çalışmada 8 duygunun tespiti için RAVDESS veri seti kullanılmıştır. F1-skoruna göre değerlendirme yapılmış, en yüksek skorun kızgın duygusuna, en düşüğün ise üzgün duygusuna ait olduğunu belirtmişlerdir.

Bu makalede Türkçe bir duygu veri tabanı oluşturmak, kullanılan metot ve materyallerle yapılacak olan çalışmalara öncü olmak, duygu tespitinde kullanılan BERT modelinin Türkçe yapısını geliştirmek ve makine öğrenmesi algoritmalarının Türkçe dil yapısı üzerindeki başarı oranlarını ölçmek amaçlanmıştır.

## 2. Veri seti ve Yöntemler

Veri seti, duygu tespiti için kullanılan BERT modeline ve analiz için kullanılan makine öğrenmesi algoritmaları aşağıdaki alt bölümlerde verilmiştir.

### 2.1. Veri seti

Veri seti, Mozilla tarafından başlatılan ses ve konuşma tanıma yazılımları için ücretsiz bir veri tabanı oluşturulmak üzere geliştirilen Common Voice platformundan alınmıştır. İnsanların nasıl konuştuğunu makinelere öğretmek amacıyla oluşturulan bu platformda çeşitli diller için ses kayıtları ve bu ses kayıtlarına ait metin dokümanları bulunmaktadır. Türkçe ses kayıtlarını içeren 592 MB boyutundaki dosya yaklaşık 20.760 ses verisini içermektedir. Ses kayıtlarının içerikleri ise; haber videolarından elde edilen sesler ve platform üzerinden yapılan kayıtlar oluşturmaktadır. Bu veri setine ait bilgiler ise Tablo 1'de verilmiştir.



**Tablo 1.** Veri seti yaş aralıkları ve oranları

Yaş Aralığı	Oranı
<19	%4
19-29	%47
30-39	%23
40-49	%3
50-59	%1

Bu veri seti için %71 erkek , %6 kadın şeklinde bir cinsiyet oranı bilgisi de paylaşılmıştır. Birçoğu tekrar eden kayıtlardan oluşan bu ses kayıtlarının tüm metin dokümanları birleştirilmiştir. Elde edilen metin dokümanları üzerinde eleme işlemi gerçekleştirilmiştir. Bir anlam ifade etmeyen, tekrar eden ve sadece sayı belirten veriler ayrıştırıldıktan sonra 5001 adet metin verisi elde edilmiştir. Bu veri seti, üzerinde anlamsız verilerin ayrıştırılmasından sonra çalışmanın içeriğini oluşturan duygu tespiti ve analiz işlemlerine hazır hale getirilmiştir. Çalışmanın deney kısımlarında veri setinin bir kısmı, öğrenme-test için kullanılmıştır. Denemeler sonucu test veri seti oranı %15 olarak belirlenmiştir. Veri seti üzerinde, verilerin duygu etiketleri BERT modeli aracılığıyla negatif ve pozitif olarak belirlenmiştir. Literatür taramasında en çok kullanılan beş makine öğrenmesi algoritması belirlenmiştir. Bu algoritmalar; Naive Bayes, Destek Vektör Makinesi, Rastgele Orman, Karar Ağacı ve K-En Yakın Komşu'dur. Python programlama dili aracılığıyla gerçekleştirilen çalışmada yapılan analizlerin daha geniş çerçevede değerlendirilmesi için hem CountVectorizer hem de TF-IDFVectorizer yöntemleri kullanılmıştır. Her iki vektörizasyon yöntemi, çalışmadaki makine öğrenmesi algoritmalarının hepsinde kullanılmıştır. RapidMiner platformuyla yapılan analizler ise sadece TF-IDFVectorizer yöntemi ile değerlendirilmiştir.

Sonuçlar ise doğruluk (accuracy), duyarlılık değeri (precision), anma değeri (recall) ve F1-skoru (F1-score) gibi metriklerle değerlendirilmiştir. Çalışma hem RapidMiner platformunda hem de Python programlama dili ile Google'ın Colab ortamında gerçekleştirilmiştir.

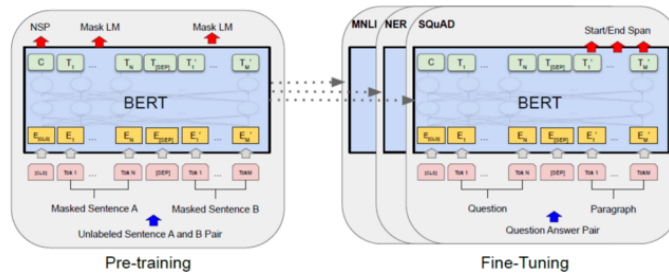
## 2.2. Metin Verilerini Kodlama Yöntemleri

Veri setinde bulunan cümleler üzerinde işlem yapabilmek için belirteçleme (tokening) adı verilen yöntemle cümledeki kelimelerin ayrıştırılması gerekir. Daha sonra makine öğrenim algoritmasına giriş olarak kullanılmak üzere kelimelerin, vektörleştirme (özellik ayıklama feature-extraction) işlemi ile tamsayılar veya kayan nokta değerleri olarak kodlanması gerekir. Klasik olarak 3 metot kullanılmaktadır. Bunlar; metni kelime sayısı vektörlerine dönüştüren CountVectorizer, kelime frekansı vektörlerine dönüştüren TF-IDFVectorizer ve benzersiz tamsayılarla (hashing) dönüştüren HashingVectorizer'dir.

Bu çalışmada ilk iki sayısallaştırma metodu kullanılmıştır.

## 2.3. BERT

BERT (Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri), çok çeşitli doğal dil işleme görevleri hakkında en gelişmiş sonuçları elde eden yeni bir eğitim öncesi dil temsili yöntemidir. 2018 yılında Google'da Jacob Devlin ve arkadaşları tarafından geliştirilen bu model yapısı cümleyi hem sağdan sola hem de soldan sağa değerlendirmektedir. Böylelikle kelimelerin birbirleriyle olan ilişkilerini daha iyi ortaya koymaktadır. Modelin yapısında, BookCorpus ve Wikipedia veri setleri bulunmaktadır[9]. BERT model yapısı Şekil 1'de verilmiştir.

**Şekil 1.** BERT model yapısı [10]

Model, Masked Language Modeling (MLM) ve Next Sentence Prediction (NSP) yöntemleri ile eğitilmektedir. MLM tekniğinde, maskelenen kelime, açık şekilde beslenen kelimelerden yararlanılarak tahmin edilmeye çalışılır. Daha çok kelimeler arasında ilişkiler üzerinde durulmaktadır. Bir diğer yöntem NSP'de ise cümleler arasındaki ilişki üzerinde durulmaktadır. NSP ile model eğitim aşamasındayken iki cümle yapısı arasındaki ilişkiye bakılarak ikinci cümlenin ilkinin devamı olup olmadığına bakılır ve tahmin yapılmaya çalışılır. Eğitim aşamasında gerçekleştirilen optimizasyon ile bu iki yöntem kullanılırken ortaya çıkan kaybın minimuma indirilmesi amaçlanmıştır [9]. BERT model yapısı verilen metin dokümanı hem sağdan hem soldan incelemesi, yapısında bulunan yöntemlerle eğitim aşamasında daha iyi bir öğrenme sağlamaktadır.

## 2.4. Naive Bayes

Naive Bayes algoritması, Bayes teoremine dayalı bir öğrenme algoritmasıdır. Tembel yapıya sahip olmasına rağmen düzensiz yapıdaki veri setlerinde de çalışabilmektedir. Bayes Teoreminin matematiksel ifadesi (1) gibidir

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A|B)$ ; B olayı gerçekleştiği durumda A olayının meydana gelme olasılığı,  $P(B|A)$ ; A olayı gerçekleştiği durumda B olayının meydana gelme olasılığı ve  $P(A)$  ve  $P(B)$ ; A ve B olaylarının olasılıklarıdır.

NB sınıflandırma algoritması, tüm koşullu olasılıkların çarpımıdır [12]. Bu sınıflandırma algoritmasıyla verilen eğitim ve test verilerinin doğru bir şekilde sınıflandırılması yapılabilmektedir. NB sınıflandırma algoritmasına ait formül (2)'de verilmiştir.

$$P(Class_j | x) = \frac{P(x | Class_j)P(Class_j)}{P(x)} \quad (2)$$

Burada verilmiş bir x tahminci olasılığı, j sınıfına ait olasılık ve eğitim kümesinden bulunan bu sınıf tahminci olasılığı ile x tahmincisinin bu sınıfa ait olma olasılığı hesaplanabilir. Bu algoritma spam filtreleme, duyarlılık analizi, çok sınıflı tahmin, öneri sistemleri ve benzeri alanlarda kullanılmaktadır [11].

## 2.5. Rastgele Orman

Rastgele Orman, birden fazla karar ağacının avantajlarını kullanarak daha uyumlu ve daha iyi sonuçlar üreten modeller ortaya koyarak sınıflandırma işlemini en iyi şekilde yapmaya çalışan bir sınıflandırma algoritmasıdır. RO çoğu zaman büyük bir sonuç üreten, esnek, kullanımı kolay bir makine öğrenmesi algoritmasıdır. RO algoritmasının en önemli avantajı ise hem sınıflandırma hem de regresyon işlemlerinde kullanılabilmesidir[13]. RO algoritmasına ait matematiksel ifade (3)'te verilmiştir. (3) denkleminde

$f_i \text{ sub}(i)$ :  $i$  özelliğinin önemi ve  $n_i \text{ sub}(j)$ :  $j$  düğümünün önemi olmak üzere;

$$\hat{f}_i = \frac{\sum_{j:i \text{ özelliğine ayrılan } j \text{ düğümü}} n_i}{\sum_{k \in \text{bütün düğümler}} n_k} \quad (3)$$

Bunlar bütün özellik önem değerlerinin toplamına bölünerek 0 ve 1 arasında bir değere normalize edilirse (4) ifadesi yazılabilir yani;

$$\text{norm } \hat{f}_i = \frac{\hat{f}_i}{\sum_{j \in \text{bütün özellikler}} \hat{f}_j} \quad (4)$$

RO düzeyindeki nihai özelliğin önemi, tüm ağaçların ortalamasıdır. Her ağaçtaki özelliğin önem değerinin toplamı hesaplanır ve toplam ağaç sayısına bölünürse (5) denklemi elde edilir:

$$RFf_i = \frac{\sum_{j \in \text{bütün ağaçlar}} \text{norm}f_{ij}}{T} \quad (5)$$

Burada,  $RFf_i$  RO modelindeki tüm ağaçlardan hesaplanan özelliğinin önemi,  $\text{norm}f_{ij}$   $j$  ağacındaki  $i$  için normalleştirilmiş özellik önemi,  $T$  ise toplam ağaç sayısıdır. Bu algoritma aynı zamanda sınıflandırıcı kategorik veriler için de değerlendirilebilmektedir. RO algoritması ayrıca over-fitting (aşırı uyum) problemlerini de gidermektedir.

## 2.6. Destek Vektör Makinesi

Destek Vektör Makinesi, 1995 yılında Vapnik tarafından istatistiksel öğrenme teorisi ve VC-boyut teorisinden türetilmiştir. DVM, ikili sınıflandırma problemini çözmek için geliştirilmiştir. DVM, karmaşık verilerin yüksek doğrulukla işleme özelliği olarak da tanımlanmaktadır. DVM'deki esas amaç eğitilmiş veri örneklerini önceden tanımlanmış sayıda sınıfa ayıran hiper düzlem bulmaktır [14-15]. DVM'nin matematiksel ifadeleri (6) ve (7) de verilmiştir.

$$\begin{aligned} \bar{x}_i * \bar{w} + b &= 1; \text{ pozitif sınıflar için} \\ \bar{x}_i * \bar{w} + b &= -1; \text{ negatif sınıflar için} \end{aligned} \quad (6)$$

Sabit marjlı DVM, hiper düzlemi geçen destek vektörleri konusunda çok kesindir. Hiper düzlemin marjını maksimize etmek için, sabit marjlı DVM optimizasyon problemine dönüşür yani;

$$\begin{aligned} \min_{\bar{w}, b} \quad & \frac{1}{2} \|\bar{w}\|^2 \\ \text{s.t.} \quad & y_i(\bar{x}_i \cdot \bar{w} + b) \geq 1; \\ & i = 1, \dots, n \end{aligned} \quad (7)$$

Bu sınıflandırma algoritmasının avantajları olarak şunlar söylenebilir; yüksek boyutlu uzayda diğer algoritmalara göre daha etkililerdir, amaç fonksiyonu için farklı çekirdek fonksiyon yapıları kullanılmaktadırlar ve amaç noktalarında kullandıkları eğitim fonksiyonları sayesinde belleğin daha verimli kullanılmasına olanak sağlamaktadırlar [16].

Algoritma kendi içinde de doğrusal ve doğrusal olmayan olarak ikiye ayrılmaktadır. Doğrusal olmayan yöntemde çekirdek fonksiyonları yöntemi kullanılmaktadır. DVM algoritması, metin ve görüntü sınıflandırmada, biyolojik bilim alanlarında ve elle yazılmış karakterlerin tanınması gibi alanlarda kullanılmaktadır.

## 2.7. Karar Ağacı

Karar Ağacı, özellik, hedef ve sınıflamaya göre karar düğümlerinden ve yaprak düğümlerinden oluşan ağaç şeklinde bir model yapısı oluşturan gözetimli sınıflandırma algoritmasıdır [17]. Yapısında bulunan en üst düğüme root (kök) ve diğer düğüm yapılarına da leaf (yaprak) adı verilmektedir. Karar Ağacı algoritmaları veri setini küçük parçalara bölerek geliştirilmektedir. Böylelikle algoritma yapısında büyük kayıpların önüne geçilerek daha küçük kayıpların olması sağlanmıştır.

Algoritma yapısı, ağaç yapısına benzetildiği için kullanılan matematiksel ifadelerde de bu durum göz önünde bulundurulmuştur. Belirli bir  $T=t$  ağacı ve  $D=d$  eğitim veri seti için,  $t$ 'nin  $d$  üzerinde ne kadar iyi çalıştığına dair olasılıksal tahmin yani  $P(T=t|D=d)$  bulunabilir. İdeal ağaç ise maksimum  $P(T=t|D=d)$  değerine sahip olmalıdır.

$$P(Y = y_1 | X = x, D = d) = \sum_{\text{Bütün Ağaçlar}} P(Y = y_1 | T = t, X = x, D = d) P(T = t | D = d) \quad (8)$$

Buna göre, bir ağaç yapısı eğitim verilerini ne kadar iyi ayırırsa  $P(T=t|D=d)$  ifadesi de maksimum değeri alır.

Karar ağaçlarının avantajları ise şöyledir; hem kategorik hem de sayısal verileri işleyebilirler, birden fazla çıktısı olan problemler için de çözüm üretebilirler, yorumlanması ve anlaşılması kolaydır ve kullanılan ağaç yapıları görsel halinde sunulabilir [18].

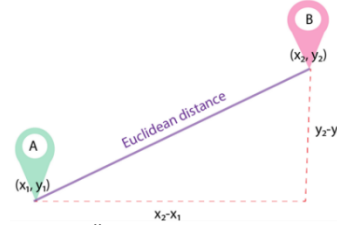
## 2.8. K-En Yakın Komşu

K-En Yakın Komşu algoritması, eklenecek olan yeni verinin mevcut veri kümelerine uzaklığını hesaplayarak, K sayıda yakın komşuluğa bakarak sınıflandırma işlemini gerçekleştirmektedir. Uzaklık hesaplamaları için, Öklid (Euclidean), Minkowski ve Manhattan uzaklık hesaplamalarını kullanmaktadır. Algoritma yapısı eski, gürlü eğitim verilerine karşı dirençli olduğundan günümüzde halen kullanılmakta olan popüler bir sınıflandırma algoritmasıdır. Algoritmanın işleyişi ise şu şekildedir; ilk olarak k parametresi belirlenir, uzaklık hesaplamaları gerçekleştirilir, komşular bulunur ve veri etiketlemesiyle süreç tamamlanır.

Algoritmanın hesaplamalarında kullanılan Öklid mesafe fonksiyonuna göre  $(x_1, y_1)$  ve  $(x_2, y_2)$  noktaları arasındaki mesafe (9)’daki formüle göre hesaplanır.

$$dist((x_2, y_2), (x_1, y_1)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (9)$$

Şekil 2’de bu formülün iki boyutta grafiği verilmiştir.



Şekil 2. Öklid mesafe fonksiyonu

KEYK algoritmasında, bir K değeri için, algoritma veri noktasının K’ya en yakın komşularını en fazla veriye sahip noktalara atar. Matematiksel olarak denklem (10)’daki gibi verilir.

$$P(y = j | X = x) = \frac{1}{K} \sum_{i \in A} I(y^{(i)} = j) \quad (10)$$

KEYK algoritması, hem regresyon hem de birçok alanda sınıflandırma işlemlerinde kullanılmaktadır [19].

## 3. Uygulamalar ve Sonuçları

### 3.1. Google Colab’da BERT Modeli ile Uygulama

Türkçe duygu tespitinde BERT modelindeki gibi cümleler pozitif ve negatif olarak değerlendirilmektedir. İlk olarak gerekli kütüphanelerin ve veri setinin yüklenmesi gerçekleştirilmiştir. Daha sonra cümle yapıları belirteçlerine ayrıldıktan sonra modele aktarılır ve burada cümlelerin pozitiflik veya negatiflik skorları ve etiketleri belirlenir. Kod kısmında pipeline nesnesi aracılığıyla veriler, peş peşe seri biçimde tanımlanmış birtakım işlemlerden geçmektedir. Bunlar sırasıyla verilerin tokenlerine ayrılması ve sonrasında yüklenen BERT modeline aktarılmasıdır.

Veri setinde bulunan cümlelerin karşısında duygu etiketleri (pozitif - negatif) ve skorları yazdırılmıştır. Gerçekleştirilen duygu tespiti sonrasında 5001 cümleden 2956 negatif ve 2045 pozitif duygu etiketine sahip cümle elde edilmiştir. Veri seti içinde oransal olarak %59 negatif ve %41 pozitif cümle bulunmaktadır. Çıktı sonuçlarının bir kısmı Şekil 3’te verilmiştir.

```

[{'label': 'positive', 'score': 0.9095399975776672, 'sentence': '"Etkinlik yirmi iki Mayıs'a kadar açık kalacak.'}]
[{'label': 'negative', 'score': 0.9859170317649841, 'sentence': '"İnsan doğasının kusurundan doğan bir durum.'}]
[{'label': 'negative', 'score': 0.9882176518440247, 'sentence': '"Kasedin etkisi büyük oldu.'}]
[{'label': 'negative', 'score': 0.9988734126091003, 'sentence': '"Bunun değişmesini beklemiyorum.'}]
[{'label': 'negative', 'score': 0.9816184043884277, 'sentence': '"Aralarında pek çok mülteci var.'}]
[{'label': 'positive', 'score': 0.5698550343513489, 'sentence': '"Bir sonraki ay bir anayasa kabul edildi.'}]
[{'label': 'positive', 'score': 0.7273528575897217, 'sentence': '"Biz realiteyi dikkate alıyoruz.'}]
[{'label': 'positive', 'score': 0.6747391223907471, 'sentence': '"Barış müzakereleri hiçbir zaman kolay olmaz.'}]
[{'label': 'positive', 'score': 0.9280513525009155, 'sentence': '"Yaşanan gelişmeler pek çok kimseyi şaşırtıyor.'}]
[{'label': 'positive', 'score': 0.5782375931739807, 'sentence': '"Dört yıl sonra dava gün ışığına çıktı.'}]
[{'label': 'positive', 'score': 0.5698100328445435, 'sentence': '"Bağış önerileri açık rekabetle seçiliyor.'}]
[{'label': 'negative', 'score': 0.8981885313987732, 'sentence': '"Saldırının sorumluluğunu henüz üstlenen olmadı.'}]
[{'label': 'negative', 'score': 0.8290243744850159, 'sentence': '"Proje Mart ayında başlatıldı.'}]
[{'label': 'negative', 'score': 0.5087615251541138, 'sentence': '"Erdoğan eleştirileri reddetti.'}]
[{'label': 'positive', 'score': 0.8988512754440308, 'sentence': '"Anlaşma bir Ocak'ta yürürlüğe girecek.'}]
[{'label': 'positive', 'score': 0.7391859889030457, 'sentence': '"Projeler altı bilimsel alanı kapsıyorlar.'}]

```

Şekil 3. BERT modeli ile yapılan duygu tespiti sonuçları

Duygu etiketleri belirlenen verilerden oluşan veri seti üzerinde makine öğrenmesi algoritmalarıyla belirlenen metrikler üzerinden analizler gerçekleştirilmiştir. Bu analizler, Python programlama dili aracılığıyla hem CountVectorizer hem de TF-IDFVectorizer yöntemiyle yapılmıştır.

Her iki yöntemde, çalışmada kullanılan tüm makine öğrenmesi algoritmaları uygulanmıştır. Veri seti, %85 eğitim verisi ve %15 test verisi olarak ayrılmıştır. Analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 431 negatif, 320 pozitif olmak üzere 751 veri bulunmaktadır.

Tablo 2'deki sonuçlarda duyarlılık değeri, anma değeri, F1-skoru ve doğruluk metriklerine ait sütunlar ikiye ayrılmıştır. Sol taraftaki veriler CountVectorizer yöntemine, sağ tarafta bulunan veriler ise TF-IDFVectorizer yöntemiyle elde edilen sonuçlara aittir. Mavi renk ile belirtilen hücrelerdeki değerler negatif veriler arasında en yüksek değeri, turuncu renkteki hücreler ise pozitif veriler arasındaki en yüksek değeri belirtmektedir.

Tablo 2. CountVectorizer (sol) ve TF-IDFVectorizer (sağ) yöntemleriyle elde edilen sonuçlar

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skoru		Doğruluk	
Naive Bayes	Pozitif	0.68	0.71	0.55	0.44	0.61	0.54	%70	%68
	Negatif	0.71	0.68	0.81	0.87	0.75	0.76		
RO	Pozitif	0.69	0.64	0.42	0.49	0.52	0.56	%67	%67
	Negatif	0.66	0.68	0.86	0.80	0.75	0.73		
DVM	Pozitif	0.63	0.67	0.56	0.56	0.59	0.61	%68	%69
	Negatif	0.70	0.71	0.76	0.79	0.73	0.75		
KA	Pozitif	0.59	0.54	0.42	0.51	0.49	0.52	%63	%61
	Negatif	0.65	0.65	0.78	0.68	0.71	0.67		
KEYK	Pozitif	0.68	1.00	0.17	0.07	0.27	0.13	%61	%60
	Negatif	0.60	0.59	0.94	1.00	0.74	0.74		

Tabloya göre;

Doğruluk oranında en iyi sonucu NB algoritması CountVectorizer ile %70 olarak elde etmiştir. En düşük doğruluk oranını ise %60 ile KEYK algoritması, TF-IDFVectorizer ile elde etmiştir.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %71 ile en iyi duyarlılık değeri oranını NB elde etmiştir. TF-IDFVectorizer ile en iyi duyarlılık oranını %71 ile DVM elde etmiştir. Pozitif veriler için ise %69 ile en iyi duyarlılık oranını RO elde etmiştir. TF-IDFVectorizer ile en iyi duyarlılık oranını %100 oranla KEYK elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %94 ile en iyi anma değeri oranını KEYK elde etmiştir. TF-IDFVectorizer ile en iyi anma değeri oranını %100 oranla (1.00) KEYK elde etmiştir. Pozitif veriler için ise %56 ile en iyi anma değeri oranını DVM elde etmiştir. TF-IDFVectorizer ile en iyi anma değeri oranını %56 oranla DVM elde etmiştir.

F1-skoru için ilk değerlendirme negatif veriler üzerinde ve CountVectorizer ile elde edilen sonuçlar için yapılmıştır. %75 ile hem NB hem de RO elde etmiştir. TF-IDFVectorizer ile en iyi F1-skorunu %71 oranla DVM elde etmiştir. Pozitif veriler için ise %61 ile en iyi F1-skoru oranını NB elde etmiştir. TF-IDFVectorizer ile en iyi F1-skorunu %61 oranla DVM elde etmiştir.

Makine öğrenmesi algoritmaları ile yapılan analizden karmaşıklık matrisleri de elde edilmiştir.

Karmaşıklık matrisi, Python programlama dilinde yapılan analizle her iki vektörizasyon işlemi sonucunda da elde edilmiştir. Karmaşıklık matrisinde negatif ve pozitif duygu etiketlerinde gerçek değerler ve tahmin değerleri gösterilmiştir. Sonuçlar Tablo 3'te verilmiştir.

**Tablo 3.** CountVectorizer (sol) ve TF-IDFVectorizer (sağ) yöntemlerinin karmaşıklık matrislerine ait sonuçlar

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler			
			Negatif		Pozitif	
NB		Negatif	348	374	83	57
		Pozitif	143	180	177	140
RO		Negatif	371	343	60	88
		Pozitif	187	162	133	158
DVM		Negatif	328	341	103	90
		Pozitif	141	141	179	179
KA		Negatif	337	295	94	196
		Pozitif	184	158	136	162
KEYK		Negatif	405	431	26	0
		Pozitif	266	297	54	23

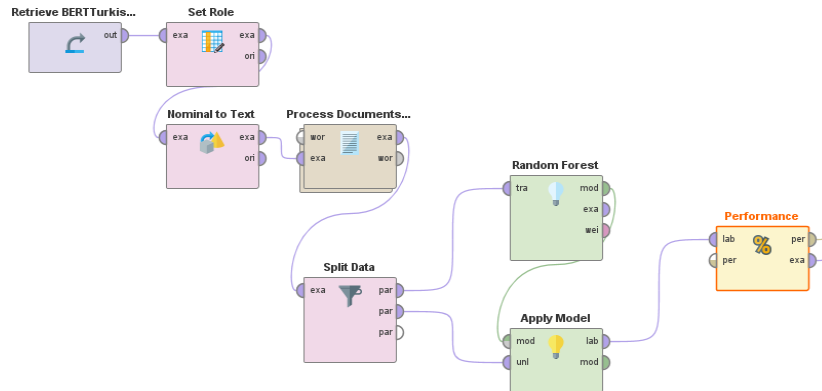
Özet olarak en iyi sonuçlara ulaşan algoritmalar Tablo 4'te verilmiştir.

**Tablo 4.** CountVectorizer (sol) ve TF-IDFVectorizer (sağ) ile elde edilen sonuçlarda en iyi oranlara ulaşan algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri		Anma Değeri		F1-Skoru		Doğruluk	
Negatif	NB	DVM	KEYK	KEYK	NB, RO	DVM	NB	DVM
Pozitif	RO	KEYK	DVM	DVM	NB	DVM		

### 3.2. RapidMiner'da Uygulama

İkinci uygulama olarak RapidMiner platformuyla, aynı veriler üzerinde analizler gerçekleştirilmiştir. Vektörizasyon yöntemi olarak sadece TF-IDFVectorizer yöntemi kullanılmıştır. Veri seti, %85 eğitim verisi ve %15 test verisi olarak ayrılmıştır. Analizler test veri seti üzerinde gerçekleştirilmiştir. Her makine öğrenmesi algoritması için operatörlerden oluşan model yapısı oluşturulmuştur. Bu model yapısında her seferinde sadece kullanılan makine öğrenmesi algoritması değişmiştir. Örnek olarak RO algoritmasına ait model, Şekil 4'te verilmiştir.



**Şekil 4.** RapidMiner'da RO algoritması için oluşturulan model yapısı

Oluşturulan model yapıları sonrasında çıkan sonuçlar Tablo 5’te verilmiştir.

**Tablo 5.** TF-IDFVectorizer yöntemiyle elde edilen sonuçlar

Algoritmalar	Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
NB(Kernel)	Pozitif	%62.35	%51.79	%67.47
	Negatif	%70.10	%78.33	
RO	Pozitif	%0	%0	%59.07
	Negatif	%59.07	%100	
DVM	Pozitif	%90.29	%39.74	%69.60
	Negatif	%68.38	%73.94	
KA	Pozitif	%85.71	%3.91	%60.40
	Negatif	%59.92	%99.55	
KEYK	Pozitif	%64.94	%53.09	%69.07
	Negatif	%71.14	%80.14	

Pozitif ve negatif duygu etiketine sahip olan veri seti üzerinde yapılan işlemler sonrasında en yüksek doğruluk oranını DVM elde etmiştir. Doğruluk oranı %69.60’dır. Analiz sonuçlarında her bir makine öğrenmesi algoritması için karmaşıklık matrisi elde edilmiştir. Bu karmaşıklık matrislerinin sonuçları Tablo 6’da verilmiştir.

**Tablo 6.** TF-IDFVectorizer yöntemiyle elde edilen karmaşıklık matrisleri

Algoritmalar	Gerçek Değerler	Veri Duygu Durumu	Tahmin Edilen Değerler	
			Negatif	Pozitif
NB(Kernel)		Negatif	347	148
		Pozitif	96	159
RO		Negatif	443	307
		Pozitif	0	0
DVM		Negatif	400	185
		Pozitif	43	122
KA		Negatif	441	295
		Pozitif	2	12
KEYK		Negatif	355	144
		Pozitif	88	163

RapidMiner’da yapılan duygu tespiti sonucunda pozitif ve negatif veriler / cümleler elde edilmiştir. Yapılan analizler test veri seti üzerinde gerçekleştirilmiştir. Test veri setinde 443 negatif ve 307 pozitif olmak üzere 750 veri bulunmaktadır.

Duyarlılık değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %71.14 oranla en iyi duyarlılık değeri oranını KEYK elde etmiştir. Pozitif veriler için %85.71 oranla en iyi duyarlılık oranını Karar Ağacı elde etmiştir.

Anma değeri için ilk değerlendirme negatif veriler üzerinde yapılmıştır. %100 oranla en iyi anma değeri oranı RO elde etmiştir. Pozitif veriler için %53.09 oranla en iyi anma değeri oranını KEYK elde etmiştir. En iyi sonuçlara ulaşan algoritmalar Tablo 7’de verilmiştir.

**Tablo 7.** TF-IDFVectorizer yöntemiyle elde edilen sonuçlarda en iyi oranlara ulaşan algoritmalar

Veri Duygu Durumu	Duyarlılık Değeri	Anma Değeri	Doğruluk
Negatif	KEYK	RO	DVM
Pozitif	KA	KEYK	

#### 4. Değerlendirme ve Öneriler

Teknolojideki gelişmeler, veri ve veri bilimindeki çalışmaların daha gelişmiş ortamlarda ve imkânlarla yapılmasına olanak sağlamıştır. Özellikle veri analizi son zamanlarda popülerliğini gittikçe arttırmıştır. Çalışmamızda da Türkçe ses kayıtlarının verilerini içeren veri seti üzerinde duygu tespiti ve makine öğrenmesi algoritmalarıyla analizler yapılmıştır. Çalışma sonucunda Türkçe bir duygu veri seti elde edilmiştir.

Python programlama dili ile yapılan analizde, iki farklı vektörizasyon işleminin kullanılması çalışmanın değerlendirmelerine farklı bakış açısı sağlamıştır. CountVectorizer yöntemiyle elde edilen sonuçlarda NB algoritmasının daha çok ön plana çıktığı görülmektedir. TF-IDFVectorizer yönteminde ise DVM ikili sınıflandırmadaki etkinliğini ortaya koymuştur. Türkçe verilerle yapılan analizlerde bu iki algoritmanın kullanımıyla daha etkin sonuçlar alınabileceği söylenebilir.

RapidMiner platformuyla elde edilen sonuçlara bakıldığında doğruluk oranında DVM iyi bir sonuç elde etmesine rağmen aynı durum değerlendirilen metrikler üzerinde gerçekleşmemiştir. Hem duyarlılık değeri hem de anma değeri metriklerinde KEYK algoritması öne çıkmıştır. RapidMiner'da vektörizasyon yöntemindeki eksiklikler ve ön işlem adı verilen işlemlerin yeterli olmaması sonuçları etkileyen unsurlar olmuştur. RapidMiner'da bu yapıların geliştirilmesi analiz sonuçlarının daha başarılı olmasında etkili olacaktır.

Kullanılan platformların yetkinlikleri de çalışmadaki sonuçlara etki etmiştir. Python programlama diliyle yapılan uygulamada kullanılan Google'ın Colab ortamı, RapidMiner platformuna göre daha geniş olanaklar sunmuştur. RapidMiner'a göre daha hızlı sonuçlar elde edilmiştir ve birden fazla yöntemin kullanılmasını sağlamıştır. Bu avantaj ise çalışmanın daha iyi analizini sağlamış ve sonuçlar hakkında daha uygun değerlendirmelerin yapılmasına katkı sağlamıştır.

BERT modelinde Türkçe çalışmalar için daha iyi sonuçlar elde edilebilmesi adına BERT modelinin yapısına Türkçe durak kelimeler, Türkçe kelimelerin ve skor yapısının eklenmesi ve geliştirilmesi için çalışmalar yapılabilir.

#### Kaynaklar

- [1] Duygu - Vikipedi. <https://tr.wikipedia.org/wiki/Duygu>
- [2] Yapay Zeka, Makine Öğrenmesi ve Derin Öğrenme Kavramları Arasındaki Fark Nedir? <https://evrimagaci.org/yapay-zeka-makine-ogrenmesi-ve-derin-ogrenme-kavramlari-arasindaki-fark-nedir>
- [3] TUZCU S. Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi. 2020;1(2):1-5.
- [4] Gondohanindjo J, Noersongko E, Fanani AZ, Basuki RS. Comparison Method in Indonesian Emotion Speech Classification. In2019 International Seminar on Application for Technology of Information and Communication (iSemantic) 2019 Sep 21 (pp. 230-235). IEEE.
- [5] Nezami OM, Lou PJ, Karami M. ShEMO: a large-scale validated database for Persian speech emotion detection. Language Resources and Evaluation. 2019 Mar;53(1):1-6.
- [6] Matin R, Valles D. A speech emotion recognition solution-based on support vector machine for children with autism spectrum disorder to help identify human emotions. In2020 Intermountain Engineering, Technology and Computing (IETC) 2020 Oct 2 (pp. 1-6). IEEE.
- [7] Mande AA, Dani S, Telang S, Shao Z. EMOTION DETECTION USING AUDIO DATA SAMPLES. International Journal of Advanced Research in Computer Science. 2019 Nov 1;10(6).
- [8] de Pinto MG, Polignano M, Lops P, Semeraro G. Emotions understanding model from spoken language using deep neural networks and mel-frequency cepstral coefficients. In2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS) 2020 May 27 (pp. 1-5). IEEE.
- [9] BERT Modeli ile Türkçe Metinlerde Sınıflandırma Yapmak <https://medium.com/@toprakucar/bert-modeli-ile-turkce-metinlerde-siniflandirma-yapmak>.
- [10] Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. Oct 112018.
- [11] Machine Learning Classification Naive Bayes <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-naive-bayes>.
- [12] Naive Bayes Classifier - Erdinç Uzun. [https://erdincuzun.com/makine\\_ogrenmesi/naive-bayes-classifier](https://erdincuzun.com/makine_ogrenmesi/naive-bayes-classifier)
- [13] Kirasich K, Smith T, Sadler B. Random forest vs logistic regression: binary classification for heterogeneous datasets. SMU Data Science Review. 2018;1(3):9.
- [14] He LM, Kong FS, Shen ZQ. Multiclass SVM based land cover classification with multisource data. In2005 International Conference on Machine Learning and Cybernetics IEEE 2005 Aug 18 (Vol. 6, pp. 3541-3545). .
- [15] Kulkarni AD, Lowe B. "Random Forest Algorithm for Land Cover Classification", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), March 16 Volume 4 Issue 3,2016, PP: 58 - 63.



- [16] Makine Öğrenimi Bölüm-4 (Destek Vektör Makineleri). <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-4-destek-vektor-makineleri>.
- [17] Decision Tree (Karar Ağacı): ID3 Algoritması [https://erdincuzun.com/makine\\_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama](https://erdincuzun.com/makine_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama).
- [18] Makine Öğrenimi Bölüm-5 (Karar Ağaçları). <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-5-karar-agacları>.
- [19] Makine Öğrenimi Bölüm-2 (k-En Yakın Komşuluk). <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-2>



## Siber Saldırıları için Rastgele Orman Algoritması Kullanılarak Öznitelik Seçimi

Abdulkadir BİLEN<sup>1\*</sup>, Ahmet Bedri ÖZER<sup>2</sup>

<sup>1</sup> Emniyet Genel Müdürlüğü, Ankara, Türkiye

<sup>2</sup> Bilgisayar Mühendisliği, Mühendislik Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye

\*<sup>1</sup> abdulkadir.bilen82@gmail.com, <sup>2</sup> bozer@firat.edu.tr

(Geliş/Received: 02/11/2021;

Kabul/Accepted: 29/11/2021)

**Öz:** Veri boyutlarındaki artışla birlikte araştırmacılar analiz aşamasını daha kolay hale getirmek için çeşitli yöntemlere ihtiyaç duymuşlardır. Veri boyutunu indirmek ve analiz doğruluğunu artırmak önem arz etmektedir. Veri analiz edilirken gereksiz alanlarla uğraşmamak ve daha az girdi ile daha doğru sonuç çıkarmak gerekmektedir. Öznitelik seçimi veri analizindeki en önemli aşamalardan birisidir. Öznitelik seçerken çeşitli makine öğrenmesi yöntemleri kullanılmaktadır. Çalışmada Tek Değişkenli Öznitelik seçimi, Özyinelemeli Öznitelik Eleme, Ağaç Tabanlı Öznitelik Seçimi ve Temel Bileşen Analizi yöntemleri kullanılmıştır. Bu yöntemlerle veri setindeki 13 öznitelik içinden en önemli olanları tespit edilmiştir. En önemli 6, 5 ve 4 öznitelik ayrı ayrı girdi olarak Rastgele Orman algoritması ile siber saldırı yöntemi tahmini yapılmıştır. Öznitelik sayısı 4'e indirildiğinde en yüksek doğruluk oranı olan %97.24 elde edilmiştir. Bu oran, tahmin yapılırken öznitelik seçiminde ilişkili özniteliklerin kullanılmasının boyut ve hız açısından önemli olduğunu göstermiştir. Elde edilen sonuçlarla birlikte öznitelik seçiminin veri üzerindeki önemi bir kez daha ortaya konulmuştur.

**Anahtar kelimeler:** Öznitelik Seçimi, Rastgele Orman, Makine Öğrenmesi, Veri Analizi

### Feature Selection Using Random Forest Algorithm to Cyber Attack

**Abstract:** With the increase in data sizes, researchers needed various methods to make the analysis process easier. It is important to reduce the data size and increase the analysis accuracy. When analyzing data, it is necessary not to deal with unnecessary fields and to produce more accurate results with less input. It is one of the most important first steps in feature selection and data analysis. Various machine learning methods are used for feature selection. Univariate Feature Selection, Recursive Feature Elimination, Tree-Based Feature Selection and Principal Component Analysis methods were used in the study. With these methods, the most important ones among the 13 features in the data set were determined. The most important 6, 5, and 4 attributes were separately input, and the cyber-attack method was predicted with the Random Forest algorithm. When the number of features was reduced to 4, the highest accuracy rate of 97.24% was obtained. It has been concluded that the inclusion of related features in the estimation is important in terms of size and speed in this ratio feature selection. With the results obtained, the importance of feature selection on the data has been demonstrated once again.

**Key words:** Feature Selection, Random Forest, Machine Learning, Data Analysis

### 1. Giriş

Verilerin boyutları arttıkça makine öğrenmesine dayalı yöntemlerin başarısı ve hızı önemli ölçüde etkilenmektedir. Öznitelik seçimi, istatistik, örüntü tanıma, veri madenciliği gibi birçok alanda zorlu bir araştırma konusu olmuştur. Veri ön işleme stratejisi olan öznitelik seçimi, daha basit ve anlaşılır modeller oluşturmada, daha anlaşılabilir veri hazırlamada ve sonuç performansını artırmada faydalar sağlamaktadır [1]. Öznitelik seçimi, gürültülü, ilgisiz ve gereksiz özellikleri ortadan kaldırmak için görüntü tanıma [2], kısaltmalar, yanlış yazımlar ve eş anlamlı kelimeler gibi dildeki zorlukları gidermek için metin madenciliğinde [3], görüntüyü sınıflandırma ve bölümlenme için görüntü analizinde [4], işlem maliyetini azaltmak, depolamayı en aza indirmek ve test verilerinin daha iyi anlaşılmasını sağlamak için izinsiz giriş tespitinde [5], veri analizini ve yorumlanmasını önemli ölçüde azaltmak için [6] uzun yıllardır metodolojide ve pratikte kullanılmaktadır.

Veri işlemede, aşırı öğrenmeyi önlemek ve sonuçların genelleştirilmesini değerlendirmek için eğitim, doğrulama ve test kümelerine ayrılan verilerin çapraz doğrulaması yapılmaktadır. Öznitelik seçimi modelin önemli bir ön adımıdır çünkü değerlendirmesi kötü yapılan ve alakasız veriyi içeren öznitelik seçimi modelin başarısını ciddi şekilde etkilemektedir [7]. Girdi boyutundaki azalma, ya öğrenme hızını ve model karmaşıklığını azaltarak ya da genelleştirme kapasitesini ve sınıflandırma doğruluğunu artırarak performansı geliştirebilmektedir. Uygun

\* Sorumlu yazar: [abdulkadir.bilen82@gmail.com](mailto:abdulkadir.bilen82@gmail.com). Yazarların ORCID Numarası: <sup>1</sup> 0000-0003-2359-8829, <sup>2</sup> 0000-0002-8005-7386

özniteliklerin seçilmesi aynı zamanda ölçüm maliyetini de azaltmaktadır. Bu gibi avantajlarından dolayı öznitelik seçimi, çoğunlukla sınıflandırma ve regresyon gibi gerçek dünya problemlerinde aktif olarak kullanılmaktadır [8].

Etiket bilgilerin kullanılabilirliği açısından öznitelik seçimi tekniği, denetimli yöntemler, yarı denetimli yöntemler ve denetimsiz yöntemler olmak üzere üç ana kategoriye ayrılmaktadır. Arama stratejilerine dayalı olarak öznitelik seçimi ise, filtre yöntemleri, sarmalayıcı yöntemler ve gömülü yöntemler olarak sınıflandırılmaktadır [9]. Denetimli öznitelik seçimi, sınıflardan ve regresyon hedeflerinden veri örneklerini ayırt ederek sınıf etiketlerinin korelasyonu değerlendirilmektedir. Denetimsiz öznitelik seçimi, genellikle kümeleme görevlerinde veri benzerliği, ayırt edici bilgi ve yeni veri oluşturma hatası gibi kriterlere göre öznitelikliğini önemini değerlendirmektedir. Sıralama yöntemlerine dayalı olan filtre yöntemleri herhangi bir öğrenme yöntemine dayanmadığı için daha verimlidir [10]. Sarmalayıcı yöntemler, öznitelikliğini ilgi düzeyini değerlendirmek için önceden tanımlanmış modelin öğrenme performansını kullanmaktadır. Gömülü yöntemler, model öğrenme ile öznitelik seçimini gömerek filtre ve sarmalayıcı yöntemler arasında değişim çözümü sağlamaktadır [10]. Öznitelik seçimi on yıllardır uygulanan bir makine öğrenmesi alanı olmasına rağmen büyük veri ve yeni senaryolardan dolayı hala gündemdedir. Verinin yalnızca büyük hacimli veya akan veri olması değil aynı zamanda dengesiz sınıflara, belirsiz etiketlere ve durağan olmayan dağılımlara sahip olması da öznitelik seçimini kullanmayı gerektirmektedir [11].

Öznitelik seçimi yapılırken literatürde çeşitli makine öğrenmesi yöntemleri kullanılmaktadır. Bu yöntemlerden Rastgele Orman (RO) algoritması yaygın olarak seçilmektedir. Kumar ve Shaikh tarafından RO sınıflandırıcısı ile birlikte, Relief öznitelik seçim, Rastgele orman seçim, Recursive öznitelik seçim ve Boruta öznitelik seçim algoritmaları kullanılmıştır. Araştırılan bu yöntemlerle tüm öznitelik seçim yaklaşımlarında öğrenme algoritmasının performansının arttığı gözlemlenmiştir. Çalışmanın sonucunda ise Rastgele ormana dayalı Boruta öznitelik seçim yöntemi en iyi sonucu vermiş olsa da dört öznitelik seçiminde de 9 özellik kümesi üretilmiştir [12]. Yadav ve Pal tarafından yapılan çalışmada 14 öznitelikli kalp hastalığı ile ilgili veriler kullanılmıştır. Sınıflandırma doğruluğu, kesinliği ve hassasiyeti analiz edildikten sonra Pearson korelasyonu ve Rastgele orman topluluğuna sahip Lasso düzenleme öznitelik seçim yöntemlerinde en iyi sonuç alınmıştır [13].

Hasan ve arkadaşları tarafından girdi özniteliklerini azaltarak izinsiz giriş tespit performansını artırmaya yönelik saldırı tespit sistemi için rastgele orman modeli önerilmiştir. 42 öznitelik ve 25 öznitelik kullanılarak yapılan bu çalışmada rastgele orman algoritmasıyla 25 öznitelikliğin işleme süresi daha kısa bulunmuştur. Önerilen yaklaşımda sınıflandırma için yararlı olan önemli ve alakalı öznitelikleri seçtiği ve bununda süreyi azaltmanın yanı sıra doğruluğu da artırdığı gözlemlenmiştir [14]. Li ve arkadaşları tarafından sınıflandırma doğruluğunu ispatlamak ve eğitim süresini azaltmak için rastgele orman algoritması tabanlı otomatik kodlayıcı saldırı tespit sistemi önerilmiştir. Öznitelik seçimi, öznitelik gruplandırması ve eğitim kümesinden oluşan bu yöntemle oluşturulan modelin algılama süresini azalttığı, tahmin doğruluğunu da artırdığı görülmüştür [15]. Yapılan araştırmalarda öznitelik seçiminin makine öğrenmesi yöntemlerine fayda sağladığı ve özellikle rastgele orman algoritmasının başarılı olduğu tespit edilmiştir. Çalışmanın ilk bölümünde öznitelik seçiminin literatür ve hangi alanlarda uygulandığı, ikinci bölümde öznitelik seçimi yöntemlerinin neler olduğu, üçüncü bölümde örnek veri setinde öznitelik seçiminin başarı durumu, son bölümde ise sonuçlar verilmiştir.

## 2. Öznitelik seçimi

Kümeleme ve sınıflandırma yöntemlerinin performansı, veri boyutlarının artmasından önemli derecede etkilenmektedir. Yüksek boyutlu verinin model oluşturma süresi ve analizi zordur. Öznitelik seçimi büyük veri kümesinden özniteliklerin alt kümesini belirlemek ve gereksiz özellikleri kaldırarak modeli daha doğru hale getirmek için kullanılan bir ön işlem sürecidir. Öznitelik seçme yöntemleri, bir arama tekniğine ve alt kümelerin performans değerlendirmesine dayanmaktadır. Seçilen özniteliklerin sayısını azaltmak veya ortadan kaldırmak ve çıktı doğruluğu performansını artırmak gibi iki önemli amacı gerçekleştirmelidir [16]. Öznitelik seçiminde denetimli, denetimsiz ve bu ikisi arasında yer alan yarı denetimli yöntemler bulunmaktadır. Öznitelik seçiminde bir model çıkarmak için kullanılan tümevarım öğrenme yöntemi arasındaki ilişkiye göre üç ana yaklaşım bulunmaktadır. Verinin genel karakterine bağlı olan ve tümevarım algoritmasından bağımsız olan filtreler, özniteliklerin alt kümesini değerlendirmek için sınıflandırıcıyla elde edilen tahmini kullanan sarmalayıcılar ve eğitim sürecinde öznitelik seçimi gerçekleştiren ve öğrenme makinelerine özgü olan gömülü yöntemlerdir [17].

## 2.1. Denetimli

Denetimli öznitelik seçimi genellikle sınıflandırma problemlerine yönelik olup öznitelik ile sınıf etiketi arasındaki ilişkiyi veya korelasyonu temel ilke olarak kullanmaktadır. Özniteliklerin önemi, uyguluk ölçütleri ile değerlendirilmektedir. Bir öznitelik kümesi  $X = (x_1, x_2, \dots, x_n)$  ve sınıf etiketi  $C$ 'ye sahip belirli bir  $D = (X, C)$  veri kümesi için, denetimli model sınıflandırma doğruluğunu maksimuma çıkaran  $\hat{S}(|\hat{S}| = \hat{k})$  optimal öznitelik altkümesini bulmayı amaçlamaktadır [18].

## 2.2. Denetimsiz

Denetimsiz öznitelik seçim yöntemleri, verinin doğan sınıflandırmasını kapsamayı ve kümeleme veya değerlendirme kriterlerine bağlı öznitelik altkümesini bularak kümeleme doğruluğunu artırmayı amaçlamaktadır. Küme algoritmalarına bağlı olup olmadıklarına göre denetimsiz filtre veya sarmalayıcı olabilmektedir. Bu yöntem özniteliklerin kümeleme yeteneğine göre en önemli özniteliklerden optimal özniteliğin altkümesini aramaktadır. Öznitelik araması, yeni seçilen özniteliğin mevcut kümeleme sonuçlarını değiştiremeye kadar devam etmektedir. Bu yöntemin amacı öznitelik seçiminin geçerliliğinin ve tahmin doğruluğunu geliştirmektir [18].

## 2.3. Yarı denetimli

$D = (D_l, D_u)$  veri kümesi göz önüne alındığında,  $D_l$  sınıf etiketlerine sahip örnek kümesidir ve  $D_u$ , sınıf etiketi olmayan örnek kümesidir. Yarı denetimli öznitelik seçimi,  $D_l$  tarafından eğitilen öğrenme modelinin performansını iyileştirmek için  $D_u$ 'yu kullanmaktadır. Başlıca filtre modeli olan yarı denetimli öznitelik seçim yöntemleri, yarı denetimli öğrenmede kilit rol oynamaktadır. Puan işlevleri yarı denetimli öznitelik seçim yöntemlerinin çoğunda uygulanmaktadır ve varyans puanı, Laplacian puanı, Fisher puanı ve kısıtlama puanı olmak üzere dört kategoride puanlama yapmaktadır [18].

## 2.4. Filtreleme

Filtre öznitelik seçim yaklaşımında, özniteliğin uygunluğu veri setinin istatistiksel ve gerçek karakteri aracılığıyla değerlendirilmektedir. Bu özelliklere dayanarak uygun bir öznitelik makine öğrenmesi ve veri madenciliğinde seçilmektedir. Filtre metodu öznitelik seçimi için temel kriter olarak sıralama tekniklerini kullanmaktadır. Uygun sıralama kriteri öznitelikleri puanladıktan sonra kullanılmaktadır. Filtre yöntemi daha az alakalı özellikleri filtrelemek için sıralamadan önce kullanılmaktadır. Benzersiz bir öznitelik veriyle ilgili alakalı bilgiyi içermektedir [19].

## 2.5. Sarmalayıcı

Sarmalayıcı yöntemde, öznitelikleri altkümesi oluşturulmakta ve belli sınıflandırıcıların altındaki değerlendirmeye alınmaktadır. Genellikle bu yöntemde öznitelik altkümesini değerlendirmek için belirli bir amaç fonksiyonu kullanılmaktadır.  $2^N$  alt kümelerini değerlendirirken NP zor bir problem haline geldiğinden optimal altkümeler, altkümeleri sezgisel olarak bulmak için arama algoritmalarıyla araştırılmaktadır. Birçok arama algoritması ilişkili amaç fonksiyonunu minimize veya maksimize ederek öznitelik altkümesini bulmak için kullanılmaktadır. Sarmalayıcı yöntemler, belli sınıflandırıcılar ile ilişkili olduğundan onun hesaplama karmaşıklığı genellikle diğer iki kategorideki öznitelik seçim yöntemlerinden daha yüksektir [19].

## 2.6. Gömülü

Gömülü yöntemde, optimizasyon problemleri iz düşünüm alt uzayında gömülü öznitelikleri sıralamak için kullanılmaktadır. Bu sıralama prosedürüne bağlı olarak, öznitelikler özel uygulamalar için seçilmektedir. bu yöntem farklı öznitelik alt kümelerini yeniden sınıflandırmak için harcanan zamanı azaltmayı amaçlamaktadır. Temel amaç öğrenme sürecinin bir parçası olarak öznitelik seçimini dahil etmektir [19].

## 2.7. Rastgele Orman Algoritması

RO, sınıflandırma ve regresyon görevleri için çok popüler ve oldukça hassas olan ve model toplama fikrine dayalı bir öğrenme algoritmasıdır. RO algoritmasının arkasındaki ana fikir, her ağacın bir sınıf için oy kullandığı

ve ormanın, ormandaki tüm ağaçlar arasından en çok oyu alan sınıflandırmayı seçen değiştirme yöntemiyle eğitim verilerinin rastgele örneklerinden çok sayıda tarafsız karar ağacını üretmektir. Yöntemin en önemli avantajlarından biri, RO'ların sınıf tahminine ilişkin her özneliğin etkisini öğrenmek için özneliklerin önem puanını ölçebilmesidir. Fakat yüksek boyutlu problem için, özneliklerin sayısı çok büyük olabilmektedir bu da öznelik önem puanlarının manuel olarak araştırılmasını ve sınıflandırma için en uygun özneliklerin seçimini çok zorlaştırmaktadır. Bu açıdan önem puanına bağlı otomatik öznelik seçim prosedürü ilgili, etkili ve ayırt edici özneliklerin seçilmesine yol açmaktadır. RO, rastgele seçilen karar ağaçlarına dayanan topluluk öğrenicisidir, topluluktaki her ağaç orijinal eğitim verisinin değiştirilmesiyle orijinal verinin rastgele bir örneğinden oluşturulmaktadır. Karar ağaçlarını oluşturmak ve her ağaçtaki son sınıfı belirlemek için gini indeksi kullanılmaktadır. Dolayısıyla  $v$  düğümündeki gini indeksi  $Gini(v)$ ,  $v$ 'nin saflığını ölçmektedir. Denklem 1'deki formülle ifade edilmektedir [20].

$$Gini(v) = \sum_{i=1}^I f_i(1 - f_i) \quad (1)$$

Burada  $f_i$ ,  $v$  düğümünde kaydedilen  $i$  sınıfının kesridir. Yine  $v$  ağaç düğümünü bölmek için  $X_i$  özneliğinin gini bilgi kazancı Denklem 2'deki formülde tanımlanmıştır.

$$gain(X_i, v) = Gini(X_i, v) - (W_L Gini(X_i, v^L) + (W_R Gini(X_i, v^R))) \quad (2)$$

Burada  $Gini(X_i, v)$ ,  $v$  düğümündeki kirliliktir.  $W_L$  ve  $W_R$ , sırasıyla  $v$  düğümünün sol ve sağ alt düğümleridir.  $W_L$  ve  $W_R$ , sol ve sağ alt düğüme atanan örneklerin kesridir. Son olarak, kirlilikteki azalmayı maksimize eden öznelik ayırma özneliği olarak kullanılmaktadır.  $X_i$  özneliği için önem puanı Denklem 3'teki gibi  $gain(X_i, v)$ 'den hesaplanmaktadır.

$$Imp_i = \frac{1}{n_{tree}} \sum_{k \in S_{X_i}} gain(X_i, v) \quad (3)$$

Burada  $n_{tree}$ , RO'daki veya topluluk boyutundaki ağaçların sayısıdır ve  $k \in S_{X_i}$ , bölünmüş düğümler kümesidir. Önem skorunun normalleştirilmesi Denklem 4'deki gibidir.

$$Imp_{norm} = \frac{Imp_i}{Imp_{max}} \quad (4)$$

Burada  $Imp_i$ , RO'dan  $X_i$ 'nin önem puanını temsil etmektedir.  $Imp_{max}$ , maksimum önemi temsil etmektedir ve normalleştirilmiş önem puanı  $0 \leq Imp_{norm} \leq 1$  arasında yer almaktadır [20].

### 3. Bulgular ve Tartışma

#### 3.1. Veri öznelikleri

Çalışmada Elazığ ilinde meydana gelen siber suç verileri kullanılmıştır. Veri setinde 14 farklı öznelik bulunmaktadır. Meydana gelen olayla ilgili olan bu öznelikler mağdurun cinsiyeti, yaşı, yaş aralığı, eğitimi, mesleği, geliri, medeni hali, suçun türü, suçun yılı, saldırının amacı, saldırının zararı, saldırının yöntemi, failin bilinme durumu ve failin yakalanma durumundan oluşmaktadır.

#### 3.2. Rastgele Orman Yöntemi ile Tahmin

Veri seti öncelikle %80 eğitim ve %20 test verisi olarak ayrılmıştır. Mağdurun cinsiyeti, yaşı, yaş aralığı, eğitimi, mesleği, geliri, medeni hali, suçun türü, saldırının amacı, saldırının zararı, failin bilinme durumu ve failin yakalanma durumu öznelikleri girdi olarak verilmiş, saldırı yöntemi tahmin edilmiştir. Veri setindeki 13 öznelik girdi olarak verildiğinde, saldırı yönteminin tahmin edilmesinde rastgele orman sınıflandırması ile %95.03 doğru tahmin yapılmıştır.

Öznelikler, Tek Değişkenli Öznelik seçimi, Özyinelemeli Öznelik Eleme, Ağaç Tabanlı Öznelik Seçimi ve Temel Bileşen Analizi yöntemlerine göre öznelik seçimi yapılmıştır. En güçlü öznelikleri tespit etmek için yapılan Tek Değişkenli Öznelik seçiminde en yüksek puan alan  $k$  özelliği dışındaki tüm özellikleri kaldıran

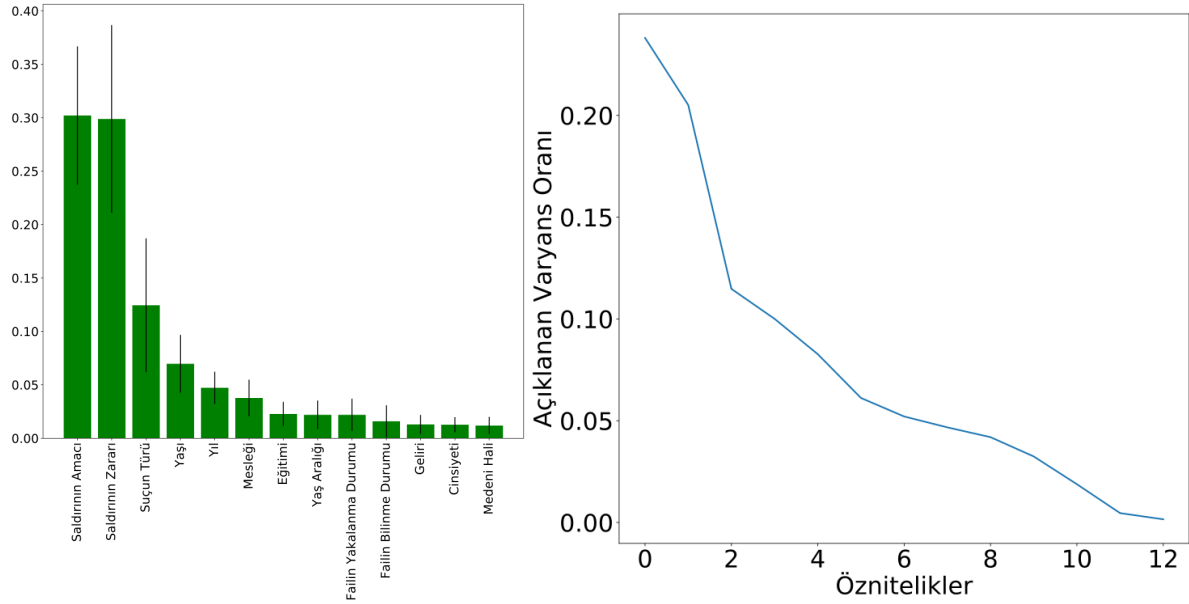
SelectKBest yöntemi kullanılmıştır ve özniteliklerin puanları Tablo3.1’de verilmiştir. Saldırının amacı, mağdurun yaşı, suçun türü, saldırının zararı ve mağdurun mesleği öznitelikleri 100 puanın üzerinde puanlanmıştır.

**Tablo 3.1.** Model sonuçları

Saldırının Amacı	Mağdurun Yaşı	Suçun Türü	Saldırının Zararı	Mağdurun Mesleği	Suçun Yılı	Yaş Aralığı
442	379	293	222	138	82	58
Failin Bilinme Durumu	Mağdurun Cinsiyeti	Mağdurun Medeni Hali	Mağdurun Geliri	Failin Yakalanma Durumu	Mağdurun Eğitimi	
31	27	16	15	14	12	

Özyinelemeli öznitelik eleme (Recursive feature elimination-RFE) yönteminde, bir modele uyan ve belirlenen öznitelik sayısına ulaşılan kadar en zayıf özniteligi (veya öznitelikleri) ortadan kaldıran bir öznitelik seçme yöntemidir. Bu yöntemde Saldırının amacı, saldırının türü, saldırının zararı, mağdurun yaşı, mağdurun mesleği ve suçun yılı en güçlü 6 öznitelik olarak elde edilmiştir.

Ağaç tabanlı öznitelik önem sıralaması Şekil 3.2A’da görüldüğü üzere yine aynı öznitelikler sıralamada önde yer almıştır. Temel Bileşen Analizinde yüksek miktar ve boyuttaki verilerin yüksek ilişkili olanları belirlemektedir ve verideki temel öznitelikleri tespit ederek daha az sayıya indirgemektedir. Bu analize göre saldırının amacı, suçun türü, saldırının zararı ve mağdurun yaşı öznitelikleri ön plana çıkmaktadır ve grafik Şekil 3.2B’de verilmiştir. Her iki şekilde de öznitelik açısından önemli olanlar grafiklerde ön sırada yer almıştır. Saldırı yöntemi ile saldırının amacı, zararı ve türü arasında diğer özniteliklere göre önemli derecede ilişki bulunmaktadır.

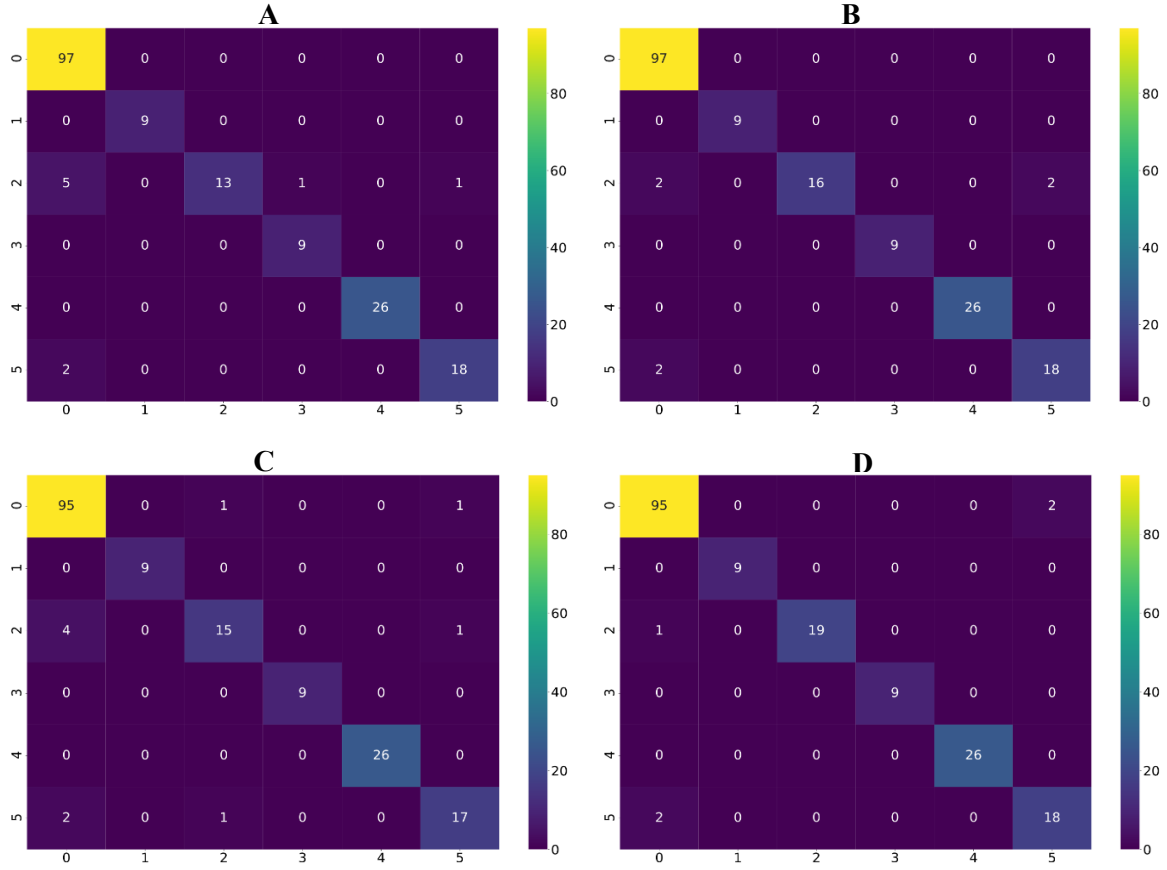


**Şekil 3.1.** A) Öznitelik önem sıralaması B) Temel bileşen analizi sıralaması

**Tablo 3.2.** Model sonuçları

RO Tahmini	Eğitim Doğruluğu (%)	Eğitim Süresi (Sn)	Test Doğruluğu (%)	Test Süresi (Sn)
13 Öznitelik	99.58	0.06249159	95.03	0.01563142
6 Öznitelik	99.30	0.04688907	96.69	0.01561856
5 Öznitelik	98.75	0.04687858	94.48	0.01559711
4 Öznitelik	99.00	0.04686189	97.24	0.01558018

Dört farklı öznitelik seçme yöntemi ile öznitelik sayıları azaltılmıştır. RO algoritması ile saldırı yöntemi tahmin edilmiştir. Tablo 3.2’de gösterildiği gibi toplam 13 öznitelik girdi olarak verildiğinde doğruluk oranı %95.03; ilk 6 öznitelik için %96.69; ilk 5 öznitelik için %94.48; ve ilk 4 öznitelik için %97.24 olarak elde edilmiştir. Saldırının amacı, suçun türü, saldırının zararı ve mağdurun yaşı girdi olarak verildiğinde saldırı yöntemini en yüksek oranda tahmin edilmiştir. Yine eğitim süresi ve test süresi açısından değerlendirildiğinde aralarındaki süre miktarları çok küçük olsa bile öznitelik sayısı azaldıkça sürelerinde azaldığı tespit edilmiştir. Bu değerlerde öznitelik seçiminin önemini bir kez daha ortaya koymuştur. Şekil 3.1 (A, B, C, D’deki) hata matrisine bakıldığında hata oranlarının oldukça düşük olduğu ve tahmin oranlarının birbirine yakın olduğu görülmektedir.



**Şekil 3.2. A) 13 Öznitelikle yapılan tahminin hata matrisi B) 6 Öznitelikle yapılan tahminin hata matrisi C) 5 Öznitelikle yapılan tahminin hata matrisi D) 4 Öznitelikle yapılan tahminin hata matrisi (Burada 0 = Hack Araçları veya Zararlı Yazılım Kullanarak, 1 = Kart Kopyalama, Üretme Cihazlarını Kullanarak, 2 = Phishing (oltalama) Saldırısı Kullanarak, 3= Sahte Alışveriş Sitesi Oluşturarak, 4= Sosyal Medyadaki Herkese Açık Verilerini Alarak, 5= Sosyal Mühendislik Kullanarak olarak sayısallaştırılmıştır.)**

#### 4. Sonuçlar

Öznitelik seçimi veri boyutu arttıkça daha da önem kazanmaktadır. Çok yüksek boyutlu veri ile çalışılırken eğitim süresinde veri boyuna göre artmaktadır. Eğitim süresini azaltmaya ve doğruluk oranlarını da artırmaya yarayan öznitelik seçiminin faydaları literatür çalışmalarından da görülmektedir. Öznitelik seçiminde kullanılan yöntemler arasında bazı farklılıklar olsa da genellikle aynı öznitelikler ön plana çıkmıştır. Çalışmada kullanılan veri setindeki en önemli öznitelikler saldırının amacı, suçun türü, saldırının zararı ve mağdurun yaşı tespit edilmiştir. RO algoritması ile seçilen öznitelik yönteminin başarılı olduğu görülmüştür. RO algoritması ile yapılan tahminde en önemli 4 öznitelikle %97.24 doğruluk oranıyla en başarılı tahmin yapılmıştır. Yapılan uygulama ile



öznitelik seçiminin hayati öneme sahip olduğu ve doğruluk oranını %2 civarında artırdığı tespit edilmiştir. Birçok sınıflandırma probleminde öznitelik seçim yöntemlerinin kullanılabileceği değerlendirilmektedir.

### Kaynaklar

- [1] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6), 1-45.
- [2] Zhang, Y., Wang, X., Cai, Z., Zhou, Y., & Philip, S. Y. (2021, July). Tensor-Based Unsupervised Multi-View Feature Selection for Image Recognition. In *2021 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE.
- [3] Hossny, A. H., Mitchell, L., Lothian, N., & Osborne, G. (2020). Feature selection methods for event detection in Twitter: a text mining approach. *Social Network Analysis and Mining*, 10(1), 1-15.
- [4] Bolón-Canedo, V., & Remeseiro, B. (2020). Feature selection in image analysis: a survey. *Artificial Intelligence Review*, 53(4), 2905-2931.
- [5] Alazzam, H., Sharieh, A., & Sabri, K. E. (2020). A feature selection algorithm for intrusion detection system based on pigeon inspired optimizer. *Expert systems with applications*, 148, 113249.
- [6] Liang, S., Ma, A., Yang, S., Wang, Y., & Ma, Q. (2018). A review of matched-pairs feature selection methods for gene expression data analysis. *Computational and structural biotechnology journal*, 16, 88-97.
- [7] Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85, 189-203.
- [8] Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in biology and medicine*, 112, 103375.
- [9] Miao, J., & Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, 91, 919-926.
- [10] Li, J., & Liu, H. (2017). Challenges of feature selection for big data analytics. *IEEE Intelligent Systems*, 32(2), 9-15.
- [11] Bolón-Canedo, V., Sánchez-Marroño, N., & Alonso-Betanzos, A. (2016). Feature selection for high-dimensional data. *Progress in Artificial Intelligence*, 5(2), 65-75.
- [12] Kumar, S. S., & Shaikh, T. (2017, September). Empirical evaluation of the performance of feature selection approaches on random forest. In *2017 international conference on computer and applications (ICCA)* (pp. 227-231). IEEE.
- [13] Yadav, D. C., & Pal, S. A. U. R. A. B. H. (2020). Prediction of heart disease using feature selection and random forest ensemble method. *International Journal of Pharmaceutical Research*, 12(4), 56-66.
- [14] Hasan, M. A. M., Nasser, M., Ahmad, S., & Molla, K. I. (2016). Feature selection for intrusion detection using random forest. *Journal of information security*, 7(3), 129-140.
- [15] Li, X., Chen, W., Zhang, Q., & Wu, L. (2020). Building auto-encoder intrusion detection system based on random forest feature selection. *Computers & Security*, 95, 101851.
- [16] El-Hasnony, I. M., Barakat, S. I., Elhoseny, M., & Mostafa, R. R. (2020). Improved feature selection model for big data analytics. *IEEE Access*, 8, 66989-67004.
- [17] Bolón-Canedo, V., & Alonso-Betanzos, A. (2019). Ensembles for feature selection: A review and future trends. *Information Fusion*, 52, 1-12.
- [18] Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70-79.
- [19] Zhang, R., Nie, F., Li, X., & Wei, X. (2019). Feature selection with multi-view data: A survey. *Information Fusion*, 50, 158-167.
- [20] Uddin, M. T., & Uddiny, M. A. (2015, May). A guided random forest based feature selection approach for activity recognition. In *2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)* (pp. 1-6). IEEE.



## Yüz Tanıma Sistemlerinde Kullanılan ESA, YGH-DVM ve DSA Algoritmalarının Performans Testleri

Faruk AYATA<sup>1\*</sup>, Hayati ÇAVUŞ<sup>2</sup>

<sup>\*1</sup> Bilgisayar Teknolojileri Bölümü, Başkale Meslek Yüksekokulu, Van Yüzüncü Yıl Üniversitesi, Van, Türkiye

<sup>2</sup> BÖTE, Eğitim Fakültesi, Van Yüzüncü Yıl Üniversitesi, Van, Türkiye

<sup>\*1</sup> farukayata@yyu.edu.tr, <sup>2</sup> hayatiicavus@gmail.com

(Geliş/Received: 20/12/2021;

Kabul/Accepted: 03/02/2022)

**Öz:** Yüz tanıma sistemleri, kriminoloji, güvenlik sistemleri gibi görüntü içeren alanlarda veriyi işlemek için görüntü işleme, makine öğrenmesi ve derin öğrenme algoritmaları kullanılmaktadır. Yüz tanıma sistemlerinde sıklıkla kullanılan makine öğrenmesi, bireyleri ayırt etmek için temelde YSA (Yapay Sinir Ağları) olan teknikleri kullanılmaktadır. Bu çalışmada makine öğrenmesi tekniklerinden; ESA (Evrişimsel Sinir Ağları), YGH-DVM (Yönelimli Gradyan Histogramı - Destek Vektör Makineleri) ve DSA (Derin Sinir Ağları) yöntemleri kullanılarak FEI, CelebA ve Aile yüz veri seti üzerinde analizler yapılmıştır. Bu testler sonucunda ESA yöntemi ile; FEI Veri Setinde %98.86, CelebA Veri Setinde %99.89, Aile Veri Setinde ise %100, YGH-DVM yöntemi ile; FEI Veri Setinde %97.71, CelebA Veri Setinde %97.75, Aile Veri Setinde ise %95.67, DSA yöntemi ile de; FEI Veri Setinde %97.61, CelebA Veri Setinde %95.79, Aile Veri Setinde ise %91.83 başarılı yüz bulma tahminine ulaşılmıştır.

**Anahtar kelimeler:** Yüz Tanıma, Evrişimsel Sinir Ağları, Derin Sinir Ağları, Aile Yüz Veri Seti.

### Performance Tests of ESA, YGH-DVM and DSA Algorithms Used in Face Recognition Systems

**Abstract:** Image processing, machine learning and deep learning algorithms are used to process data in fields such as face recognition systems, criminology, security systems. Machine learning, which is frequently used in face recognition systems, uses techniques based on ANN (Artificial Neural Networks) to distinguish individuals. In this study, one of the machine learning techniques; Analyzes were performed on FEI, CelebA and Aile face dataset using ESA (Convolutional Neural Networks), YGH-DVM (Directional Gradient Histogram - Support Vector Machines) and DSA (Deep Neural Networks) methods. As a result of these tests, with the ESA method; 98.86% in FEI Data Set, 99.89% in CelebA Data Set, 100% in Family Data Set, with YGH-DVM method; 97.71% in FEI Data Set, 97.75% in CelebA Data Set, 95.67% in Family Data Set, and with DSA method; In the FEI Data Set 97.61%, in CelebA Data Set 95.79%, in Family Data Set 91.83% successful face detection estimates were reached.

**Key words:** Face Recognition, Convolutional Neural Networks, Deep Neural Networks, Family Face Data Set.

### 1. Giriş

Görüntü işleme alanı; dijital görüntü işleme, yüz tanıma, nesne tanıma, desen tanıma, kod tanıma ve optik karakter tanıma gibi teknolojileri barındırmaktadır [3]. Bu teknolojilerden yüz tanıma; güvenlik sistemlerinde, personel, üye ve öğrenci takip sistemlerinde, geçiş kontrollerinde, kayıp ve ya aranan bireylerin bulunmasında kullanılabilir.

Son birkaç yılda, yüz tanıma ve kişi tanımlama sistemleri, güvenlik endişesi nedeniyle birçok araştırmacıyı cezbetmektedir; bu nedenle, birçok ilginç ve faydalı araştırma gösterileri ve ticari uygulamalar geliştirilmiştir. Herhangi bir yüz tanıma veya görme temelli kişi tanımlama sisteminin ilk adımı, yüzün görüntüdeki yerini belirlemektir. Olası yüzün yerini belirledikten sonra araştırmacılar, yüzü doğru bir şekilde tespit etmek için yüz özelliklerini (gözler, burun, burun delikleri, kaşlar, ağızlar, sıçramalar vb.) algılama yöntemini kullanır [26]. Yüz tanıma veya kişi tanımlama sistemleri, bir giriş yüz görüntüsünü veya görüntü özelliklerini bilinen bir yüz veri tabanı ile karşılaştırır ve buna göre bir sonuç geri gönderir.

Görüntüleri bilgisayarda işleyebilen birçok algoritma mevcuttur. Bu algoritmalar, makine öğrenmesi, derin öğrenme ve YSA başlıkları altında toplanmaktadır. Bu algoritmaların uygulamadaki başarı oranları çok yüksek

\* Sorumlu yazar: [farukayata@yyu.edu.tr](mailto:farukayata@yyu.edu.tr). Yazarların ORCID Numarası: <sup>1</sup> 0000-0003-2403-3192, <sup>2</sup> 0000-0001-5602-5221

değildir, çünkü kameranın açısı, aydınlatma, poz ve yüz ifadesi gibi değişkenler yüzün tanınmasında zorluklar çıkarabilir [24]. Yüksek başarı oranının elde edilmek istenildiği durumlarda YSA tabanlı sınıflandırma algoritmalarının tercih edilmesi daha doğru olacaktır [17]. YSA insan beyninin çalışma sisteminin taklit edilmesiyle ortaya çıkmış ve 1980-2000 yılları arasında büyük başarı sağlamış bir yöntemdir [7]. YSA'ların geliştirilerek daha iyi bir beyin modellenme amacına sahip, gizli katman sayısının artırıldığı, mimarileri DSA olarak adlandırılmaktadır. ESA ise daha çok görüntü içeren verilerin sınıflandırılmasında kullanılan YSA tabanlı bir yöntemdir. ESA temelde, ham olarak verilen görüntünün hangi görüntü sınıfına ait olduğu ya da o görüntüyü tanımlayan en iyi sınıf olasılığının çıkartılması işlemini yapar [2].

İmgelerden öznitelik çıkarmak için sıklıkla kullanılan bir özellik tanımlayıcısı olan YGH, bir görüntü algılama penceresinin veya ilgilenilen bölgenin lokalize bölümlerinde gradyan oryantasyonunun oluşumlarını sayarak özellik çıkarımı yapmaktadır [13-18]. Çıkarılan öz nitelikleri kullanarak sınıflandırma yapabilen birçok yöntem vardır. Bu yöntemlerden DVM, iki sınıfa ait verileri en uygun şekilde birbirinden ayırmada kullanılır. Bu ayırma işlemini hiper düzlemler ile yapar. İki gruba da en yakın noktalar üzerinden paralel çizgiler çizilir ve bu çizgileri birbirine yaklaştırarak ortak tek bir sınır çizgisi çizilerek veriler sınıflandırılır [23].

## 2. Literatür Taraması

Tek bir görüntüden veya bir görüntü dizisinden yüz algılama; poz, boyut, yön, renk, ifade ve aydınlatma koşullarındaki değişkenlik nedeniyle zor bir iştir. İnsan yüzlerinin bulunduğu görüntülerden bilgi alan tam otomatik bir sistem oluşturmak ve insan yüzlerini algılamak oldukça zor bir süreçtir [11]. Yüz tanıma için birçok yaklaşım vardır. Yüz algılama araştırmacıları, yüz algılama çalışmasını dört kategoride özetlediler: şablon eşleştirme yaklaşımları, özellik değişmez yaklaşımlar, görünüme dayalı yaklaşımlar ve bilgiye dayalı yaklaşımlar [26]. Aşağıda yüz bulma ve tanıma alanındaki geçmişte gelecek vaat eden birkaç çalışma özetlemektedir.

Ayata ve ark. (2020), ESA kullanarak yüz tanıma işlemini gerçekleştirip istemediğiniz kişilerin kişisel bilgisayarınızı kullanmasını kısıtlayan ve bu kısıtlamaya ek olarak kişisel bilgisayarınızı kullanmak isteyen kişinin fotoğrafını çekerek bu fotoğrafı sistemde daha önce tanımlanmış olan ve bilgisayar sahibine ait cep telefonuna mesaj olarak gönderen bir uygulama geliştirmişlerdir. Testler sonucunda en kötü açı ve ışık değerinde % 76.31 ve en iyi açı ve ışık değerinde de %99.15 başarı sağlanmışlardır. Geliştirdikleri sistem ile ESA yöntemini yüz tanımadaki kişisel bilgisayarların güvenliğini arttırmışlardır [4].

Vinay ve ark. (2017), yaptıkları çalışma iki basit ama etkili ESA mimarisi kullanarak yüz tanıma işlemi yapmaktadır. Her iki mimaride de gabor ve frangi filtrelerini, veri seti olarak CMU, Grimace, Yale, Face 95 ve FEI kullanmışlardır. Analiz sonucunda veri setlerine göre başarı oranının %90 ile %100 arasında değiştiğini belirtmişlerdir. Görüntülerden yüzleri tanıma noktasında ESA yönteminin ne kadar etkili olduğunu ortaya koymuşlardır [25].

Coşkun ve ark. (2017), geliştirdikleri sistemde ESA katmanlarından ikisine iki normalleştirme işlemi ekleyerek yeni bir mimari önermişlerdir. Bu mimari sayesinde ağırlık hızlanmasını sağlamışlardır. Ayrıca yüzleri sınıflandırmak için softmax sınıflandırıcı kullanmışlar ve test için Georgia veri seti kullanarak analiz sonuçlarında yüksek başarı oranı elde ettiklerini belirtmişlerdir. Geliştirdikleri sistemle yüz tanıma işlemlerindeki zamansal probleme çözüm üretmişlerdir.

Tanrıverdi (2017), öğretmenlerin akıllı cihazlarına yükleyebileceği otomatik yüz bulma ve tanıma işlemi gerçekleştirebilen bir sistem geliştirmiştir. Yüz tanıma işlemi için ÖYYTS (Özyüz Yaklaşımıyla Yüz Tanıma Sistemleri), FYYTS (Fisher Yaklaşımıyla Yüz Tanıma Sistemleri) ve YİDYTS (Yerel İkili Desen Yüz Tanıma Sistemleri) tekniklerini kullanmıştır. Analizler sonucunda yüz tanıma başarı oranını %84.81 bulmuştur. Geliştirdikleri sistem ile sınıflardaki manuel yoklama sisteminde karşılaşılabilecek sorunları ortadan kaldıracaklarını belirtmişlerdir [20].

Holat (2016), yüz tanıma için tasarladığı gerçek zamanlı sistemde PCA (Temel Bileşen Analizi), LDA (Doğrusal Ayraç Analizi) ve LBP (Yerel İkili Örüntü) yüz tanıma yöntemlerini kullanmıştır. Geliştirdiği sistemde yüz bölgesinin tespiti için Adaboost algoritması ve testler içinde Yale veri seti kullanmıştır. Analizler sonucunda yüz tanıma işleminde %91'e yakın başarı oranı elde ettiğini belirtmiştir [12].

Bayrakdar ve ark. (2016), yüz eylemlerini algılayarak sınıflandıracak otomatik bir yüz ifade analiz sistemi önermişlerdir. Çalışmalarında 150'nin üzerinde eylem birimini ya da kombinasyonunu kullanan 24 farklı denek ile 1100'den fazla görüntü dizisi içeren bir veritabanı oluşturmuşlar ve görüntülerdeki yüz eylemlerini sınıflandırmak için üç farklı yaklaşımı karşılaştırmışlardır: gri-seviye görüntülerin temel bileşenler tabanlı bütüncül mekânsal analizi (holistic spatial analysis based on principal components of gray level images), kırışıklıklar gibi yerel görüntü özelliklerinin kesin ölçümü (explicit measurement of local image features such as wrinkles) ve hareket akışı alanları ile şablon eşleştirme (template matching with motion flow fields). Sonuç olarak

bu metotlar ile 6 farklı eylem birimi ve 20 denek içeren bir veri seti üzerinde, sırasıyla %89, %57 ve %85 başarımla elde etmişlerdir. Bu üç metodu birlikte kullandıklarının da ise performansı %92'lere kadar çıkarmayı başarmışlardır [5].

Güzel Turhan, (2014) yüz tanıma sistemlerinde yaygın olarak kullanılan TBA tabanlı yöntemi geliştirerek sınıf bazında yeni bir TBA yaklaşımı sunmuştur. Geliştirdiği sistem Java platformunda çalışmakta olup video görüntüleri üzerinde tanıma işlemini gerçekleştirmiştir. Analizler sonucunda önerilen sistemin var olan klasik sistemlere oranla daha başarılı olduğu göstermiştir. %87.03 başarılı tanıma oranı elde etmiştir.

Farahani ve ark. (2013), farklı yaş gruplarındaki göz ve ağız özelliklerinden duygu tanınması için bulanık tabanlı yeni bir yöntem sunmuşlardır. Farklı renk alanlarının birleşiminden gözleri ve ağız algılandıktan sonra bulanık analizleri için dört parametre seçmişlerdir. Bu parametreler gözün açık olması, ağzın açık olması, göz açıklık/genişlik oranı ve ağız genişliği olarak belirlemişlerdir. Daha sonra bulanık mantık çıkarım yöntemlerinden biri olan Mamdani yöntemi kullanılarak yüz özellikleri ve onların duygu uzayındaki karşılıklarını kodlamışlardır. Bu yöntem ile Ebner'in yüz ifadesi veritabanı üzerinde ortalama %78.8 doğruluk oranı elde etmişlerdir. Ayrıca önerilen yöntemi, Cohn- Kanade veritabanını kullanılan Yerel İkili Örüntü (LBP- Local Binary Pattern), DVM, AdaBoost ve Yükseltilmiş LBP (Boosted-LBP) ile karşılaştırmışlardır. Mutluluk, korku ve üzüntü ifadeleri için en iyi sonuçlar önerilen yöntemde elde edilmesine rağmen normal ve iğrenme ifadeleri için sonuçlar diğer yöntemlere kıyasla daha düşük kaldığını belirtmişlerdir [9].

Taşova (2011), yaptığı çalışmada görüntüdeki eğim değeri kullanılarak öznitelik çıkarımı yapmıştır. Daha sonra elde edilen öznitelikler YSA kullanılarak yüz tanıma işlemini gerçekleştirmiştir. Çalışmanın testleri için iki farklı veri tabanı kullanmıştır ve 10 kişilik veri tabanında %98.15'lik bir başarı oranı elde etmiştir. İnsan yüzündeki çok sayıda belirgin noktayı YSA'yı beslemek için birer özellik olarak kullanıp başarılı sonuçlar elde etmiştir [21].

Erdoğan (2010), yüz tanıma işleminde temel iki yöntem olan Özyüz ve Fisher yüz algoritmalarını inceleyerek kıyaslama işlemi yapmıştır. Analizler sonucunda Fisher yüz tanıma algoritmasının, Özyüz tanıma algoritmasına karşın daha başarılı olduğunu belirtmiştir [8].

Özbudak, (2009) görünüş tabanlı yöntem kullanarak cinsiyete göre sınıflandırma işlemini gerçekleştirmiştir. Sınıflandırma işlemi için Fisher Lineer Ayırıştırma (FLA) ve KYK algoritmalarını kullanmıştır. Analiz sonucunda FLA'nın %7 ve KYK'nın ise %9 hata oranıyla sınıflandırma yaptığını belirtmiştir. Buna göre cinsiyet sınıflandırmada FLA'nın KYK'ya göre daha iyi başarımla gösterdiğini ayrıca cinsiyet ayırımında yüz organlarından burunun etkin bir rol aldığını belirtmişlerdir [16].

### 3. Veri Setleri

#### 3.1 FEI veri seti

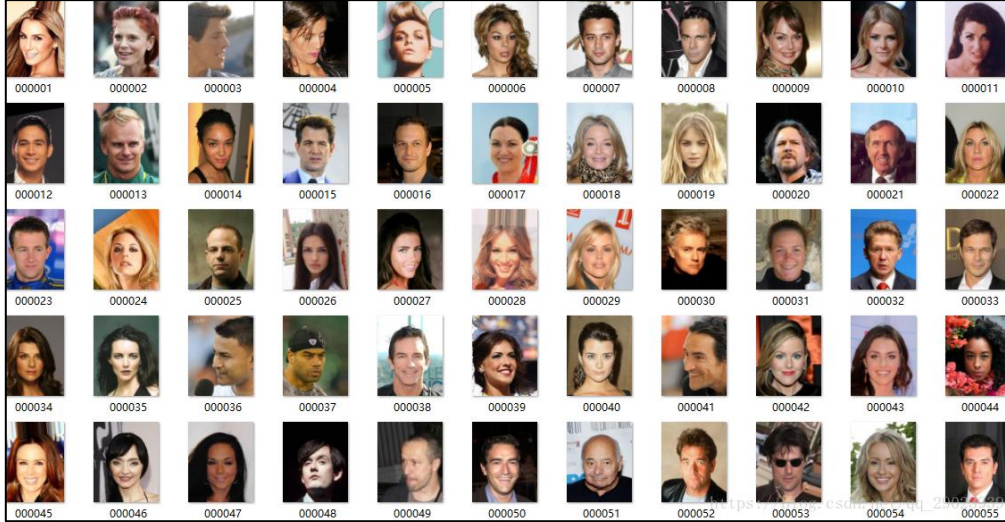
Brezilya'daki FEI Yapay Zekâ Laboratuvarı'nda Haziran 2005 ile Mart 2006 arasında 19-40 yaşlar aralığındaki öğrenciler ve personel tarafından farklı görünüm, saç modeli ve aksesuarlar tercih edilerek çekilmiş bir dizi yüz görüntüsü içeren Brezilyalı yüz veri tabanıdır. FEI yüz veri tabanında 200 kişinin 14 farklı poz ve ışık değerinden oluşan, toplamda 2800, yüz görüntüsü bulunmaktadır. Her görüntünün orijinal boyutu 640x480 pikseldir (Şekil 1).



Şekil 1. FEI veri setinden bir kesit

### 3.2 CeleBA veri seti

Bu veri seti 40 öznitelik ek açıklaması olan, 200.000'den fazla ünlü resmine sahip, büyük ölçekli bir yüz veri setidir. 10.177 bireyin sarı saç, gülümseme, keçasakalı ve gözlük gibi çeşitli özellikleri vardır [14]. Her görüntünün boyutu 178x218 olarak belirlenmiştir (Şekil 2). Çalışmada bu veri setinin bir kısmı kullanılmıştır.



Şekil 2. CeleBA veri setinden bir kesit

### 3.3 Aile yüz veri seti

Aile yüz veri seti; 39 çekirdek ailedeki bireylerin, çoğunluğu biyometrik resim (kimlik, pasaport ve ehliyet gibi belgelerde kullanılan, en az 600dpi olacak şekilde yüksek kaliteye sahip, arka fonu beyaz ve başka objelerden arındırılmış vesikalık fotoğraf türü) olan, yüz görüntülerinden oluşmaktadır. Bu veri setinde her ailenin 4-9 arasında değişen bireylerinin yüz görüntüleri bulunmaktadır (Şekil 3). Toplam yüz görüntü sayısı 208'dur. Bu veri seti ilk kez tarafımızca oluşturulmuştur.



Şekil 3. Aile yüz veri setinden örnek bir aile

## 4. Metot

### 4.1 YGH-DVM yöntemi

Bu yöntemde özellik çıkarıcı olarak YGH, sınıflandırıcı olarak da DVM kullanılmıştır. YGH özellik çıkarıcısında degradelerin ( $x$  ve  $y$  türevleri) yönlerinin dağılımı özellik olarak kullanılır. Degradelerin büyüklüğü kenarlarda ve köşelerde (ani yoğunluk değişimlerinin olduğu bölgeler) büyük olduğundan nesnenin şekli hakkında daha çok bilgi içermektedir [15]. YGH'ların, bloklar ve hücrelerden oluşan, resimlerin sıkıştırılmış ve kodlanmış halleri olduğu söylenebilir. Her pikselin parlaklık yoğunluğu ve yönleri hakkında bilgi içerirler.

DVM'ler iki sınıfa ait verileri en uygun şekilde birbirinden ayırmak için kullanılan bir makine öğrenimi algoritmasıdır. Bu ayırma işlemini hiper düzlemler ile yapar. İki gruba da en yakın noktalar üzerinden paralel çizgiler çizilir ve bu çizgileri birbirine yaklaştırarak ortak tek bir sınır çizgisi çizilir [23].

YGH algoritmasından gelen özellik vektörü, DVM modelinde her bir etiketle giriş vektörü için bir eşleşme puanı belirlemek için kullanılır. DVM, eğitim verisi içerisinde en yakın eşleşmeye ait maksimum puanlı etiketi döndürür [22-19].

YGH-DVM yöntemi uygulanan görüntüdeki yüzün tespiti için hazırlanan programın Pseudo kodu aşağıda verilmiştir.

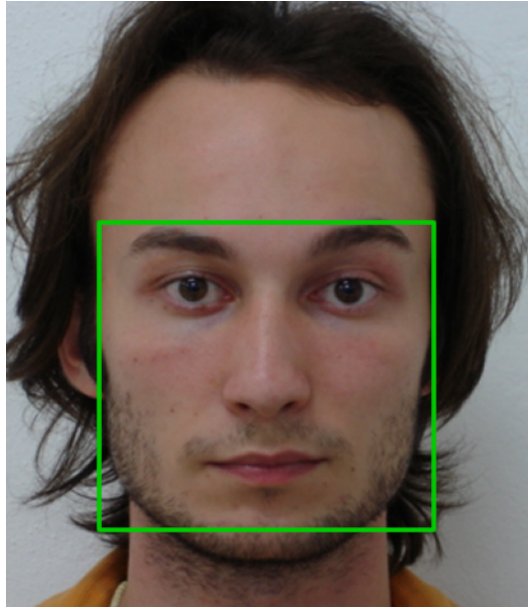
### **Sözde Kod**

**Giriş:** images [ ] All images of the FEI data set

**Çıkış:** Coordinate points (x,y) of the face rectangle and predicted face points

1. Predictor = Load predictor library from dlib(shabe predictor)
2. Face\_detector = Load detector library from frontal\_face\_detection model
3. images [ ] FEI dataset images
4. For j in images [ ] do:
5.     image = get j<sup>th</sup> image from images dataset
6.     image = resize image
7.     gray = convert image to gray mode
8.     face\_hog = detect faces of the gray
9.     for i in face\_hog do:
10.         x,y,w,h get coordinates of the i<sup>th</sup> face
11.         draw rectangle according to face coordinate x,y,w,h
12.         shape = predict face object point via dlib predictor function
13.         for (x,y) in shape do:
14.             generated\_image = draw object of the face
15. return generated\_image

YGH-DVM yöntemi uygulanan görüntüdeki yüzün tespiti Şekil 4'te gösterilmiştir.



Şekil 4. YGH-DVM yöntemi ile yüzün tespiti

## **4.2 DSA yöntemi**

DSA, büyük eğitim verilerini işleyebilen ve çok sayıda gizli katmana sahip olan makine öğrenme modelleri ile çalıştığı için çok daha faydalı özellikleri öğrenebilir, sınıflandırabilir ve tahmin doğruluğunu artırabilir.



DSA yöntemi uygulanan görüntüdeki yüzün tespiti için hazırlanan programın Pseudo kodu aşağıda verilmiştir.

### **Sözde Kod**

**Giriş:** images [ ] All images of the FEI data set

**Çıkış:** Coordinate points (x,y) of the face rectangle and predicted face points

1. Predictor = load predictor library from dlib shape predictor
2. p = dnn\_protocal\_text
3. m = dnn\_model
4. face\_detector = load detector library from dnn according to p,m parameters
5. images [ ] = FEI dataset images
6. for j in images [ ] do:
7.     image = get j<sup>th</sup> image from images dataset
8.     gray = convert image to gray mode
9.     img\_height, img\_witdh = get image properties
10.    face\_dnn = detect face of the gray
11.    for i in face\_dnn do:
12.       prediction\_score = get face score of the i<sup>th</sup> faces
13.       if prediction\_score < 0.6 then
14.          contiunue
15.       draw rectangle according to face coordinate
16.       shape = predict face object point via dlib predictor function
17.       for (x,y) in shape do:
18.          generated\_image = draw object of the face
19. return generated\_image

DSA yöntemi uygulanan görüntüdeki yüzün tespiti Şekil 5’de gösterilmiştir.



Şekil 5. DSA yöntemi ile yüzün tespiti

### **4.3 ESA yöntemi**

Eğitilebilen birçok katmandan oluşan ve bu sayede yüksek miktarda öznetelik üretebilen, ileri beslemeli, ESA mimarisi, derin öğrenmeye göre daha hızlı öğrenir. Çünkü derin öğrenmede girdi boyutu nedeniyle bağlantı büyüklüğü artar ve bu da yavaş öğrenmeye sebep olur ama ESA yapısında eğitim verisinin az tutulması ile bu dezavantaj ortadan kaldırılır [1-6-10].



ESA yöntemi uygulanan görüntüdeki yüzün tespiti için hazırlanan programın Pseudo kodu aşağıda verilmiştir.

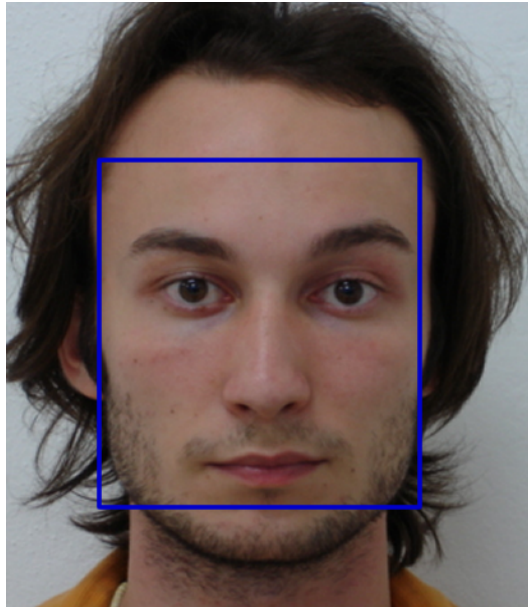
### **Sözde Kod**

**Giriş:** images [ ] All images of the FEI data set

**Çıkış:** Coordinate points (x,y) of the face rectangle and predicted face points

1. Predictor = Load predictor library from dlib(shabe predictor)
2. Face\_detector= Load detector library from cnn\_face\_detection model
3. images [ ] FEI dataset images
4. For j in images[ ] do:
5.     image = get j<sup>th</sup> image from images dataset
6.     image = resize image
7.     gray = convert image to gray mode
8.     face\_cnn = detect faces of the gray
9.     for i in face\_cnn do:
10.         x,y,w,h get coordinates of the i<sup>th</sup> face
11.         draw rectangle according to face coordinate x,y,w,h
12.         shape = predict face object point via dlib predictor function
13.         for (x,y) in shape do:
14.             generated\_image = draw object of the face
15. return generated\_image

ESA yöntemi uygulanan görüntüdeki yüzün tespiti Şekil 6’da gösterilmiştir.



**Şekil 6.** ESA yöntemi ile yüzün tespiti

#### **4.4 Kullanılan başarımlı ölçüm yöntemi**

Hazırlanan modelin başarısını ölçmek için Accuracy metriği kullanılmıştır. Accuracy değeri modelde doğru tahmin edilen alanların toplam veri değerine oranı ile hesaplanmaktadır (Denklemler 1). Çalışmada FEI ve CelebA veri setlerinde 2800, Aile veri setinde ise 208 adet görüntü bulunmaktadır.

$$\text{Doğruluk Oranı} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

## 5. Bulgular

Yapılan çalışma kapsamında farklı veri setleri üzerinde yüzleri bulmak için ESA, DSA, YGH-DVM teknikleri kullanılmıştır. Bu teknikler kullanılarak veri setleri üzerinde yapılan testlerin sonuçları Tablo 1, Tablo 2, Tablo 3 ve Tablo 4.'te gösterilmiştir.

**Tablo 1.** ESA yöntemiyle FEI, CelebA ve Aile Yüz veri setleri üzerinde yapılan test sonuçları

	ESA B.O	Zaman Değerleri
FEI Veri Seti	98,86	03:22:20
CelebA Veri Seti	99,89	01:40:01
Aile Yüz Veri Seti	100,00	00:07:19

ESA yöntemi; FEI, CelebA ve Aile veri setlerine uygulanarak Tablo 1'de gösterilen sonuçlar elde edilmiştir. En yüksek başarımın Aile veri seti üzerinde elde edildiği görülmektedir.

**Tablo 2.** DSA yöntemiyle FEI, CelebA ve Aile Yüz veri setleri üzerinde yapılan test sonuçları

	DSA B.O.	Zaman Değerleri
FEI Veri Seti	97,61	00:10:33
CelebA Veri Seti	95,79	00:03:41
Aile Yüz Veri Seti	91,83	00:00:35

DSA yöntemi; FEI, CelebA ve Aile veri setlerine uygulanarak Tablo 2'de gösterilen sonuçlar elde edilmiştir. En yüksek başarımın FEI veri seti üzerinde elde edildiği görülmektedir.

**Tablo 3.** YGH-DVM yöntemiyle FEI, CelebA ve Aile Yüz veri setleri üzerinde yapılan test sonuçları

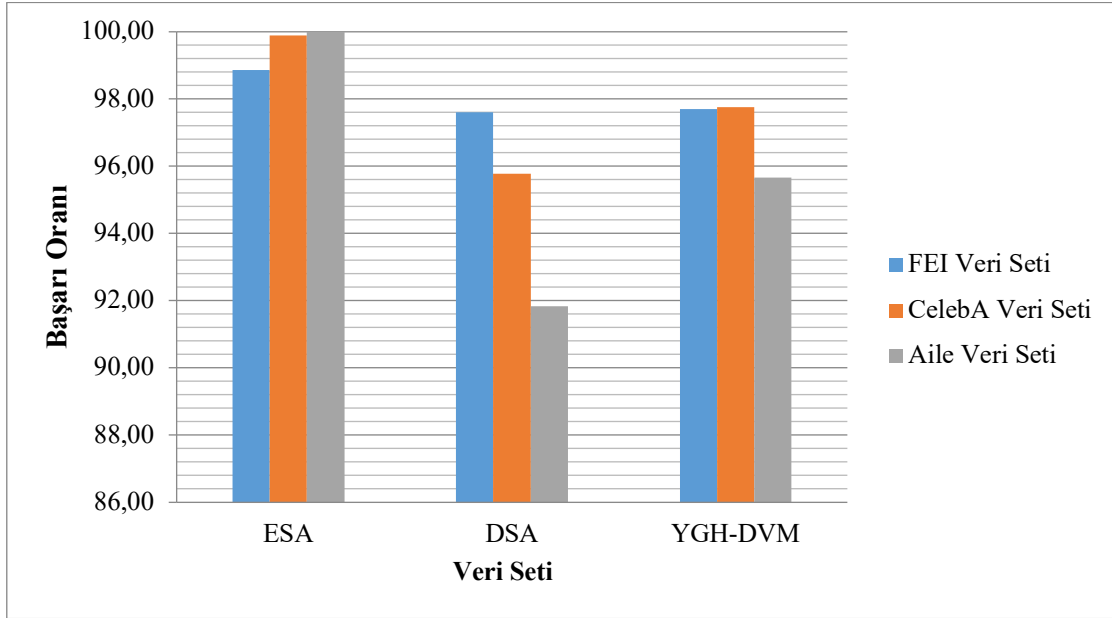
	YGH-DVM B.O.	Zaman Değerleri
FEI Veri Seti	97,71	00:13:40
CelebA Veri Seti	97,75	00:01:23
Aile Yüz Veri Seti	95,67	00:00:28

YGH-DVM yöntemi; FEI, CelebA ve Aile veri setlerine uygulanarak Tablo 3'te gösterilen sonuçlar elde edilmiştir. En yüksek başarımın CelebA veri seti üzerinde elde edildiği görülmektedir.

**Tablo 4.** Uygulanan yöntemlerin FEI, CelebA ve Aile Yüz veri setlerine göre başarı oranları

		ESA	DSA	YGH-DVM
FEI Veri Seti	Toplam Başarı	2768	2733	2736
	Başarı Oranı	98,86	97,61	97,71
	Zaman Değeri	03:22:20	00:10:33	00:13:40
CelebA Veri Seti	Toplam Başarı	2797	2682	2737
	Başarı Oranı	99,89	95,79	97,75
	Zaman Değeri	01:40:01	00:03:41	00:01:23
Aile Yüz Veri Seti	Toplam Başarı	208	191	199
	Başarı Oranı	100,00	91,83	95,67
	Zaman Değeri	00:07:19	00:00:35	00:00:28

Yüz tahmin yöntemlerini karşılaştırma adına yapan bu çalışmanın analiz sonuçları Tablo 4'te gösterilmiştir. Başarı oranları incelendiğinde en yüksek başarının ESA tekniğiyle Aile Yüz veri setinde ama zamansal başarıya bakıldığında ise YGH-DVM tekniğiyle Aile Yüz veri setinde elde edildiği görülmektedir. Bu sonuçlara dayanarak hazırlanan başarı oran grafiği de Şekil 7'de gösterilmiştir.



Şekil 7. Uygulanan yöntemlerin veri setlerine göre başarı oran grafiği

## 6. Sonuçlar

Bu çalışmada makine öğrenmesi tekniklerinden ESA, DSA ve YGH-DVM yöntemlerinin yüz tahmin performansları kıyaslanmıştır. Analizler için FEI, CelebA ve tarafımızca hazırlanmış olan Aile Yüz veri setleri kullanılmıştır.

Tablo 4.4'te gösterilen analiz sonuçları incelendiğinde; ESA yöntemi ile; FEI Veri Setinde %98.86, CelebA Veri Setinde %99.89, Aile Veri Setinde ise %100, YGH-DVM yöntemi ile; FEI Veri Setinde %97.71, CelebA Veri Setinde %97.75, Aile Veri Setinde ise %95.67, DSA yöntemi ile de; FEI Veri Setinde %97.61, CelebA Veri Setinde %95.79, Aile Veri Setinde ise %91.83 başarılı yüz bulma tahminine ulaşılmıştır. En yüksek başarı oranını ESA yöntemi ile, %100 başarılı tahmin ile, Aile Yüz veri seti kullanılarak elde edilmiştir. Diğer yöntemlerin ESA yöntemi kadar başarılı yüz tahmini yapamadığı ama zamansal anlamda ESA yönteminden daha başarılı oldukları görülmüştür.

Yüz tanıma sistemleri üzerinde çalışma yapan Tanrıverdi [20] ve Holat [12] LBP yöntemi kullanarak sırasıyla %84.41 ve %91 başarı elde etmişlerdir. Taşova [21], YSA kullanarak yaptığı çalışmada 10 kişilik bir veri seti üzerinde yaptığı analiz sonuçlarına göre %98.15 başarılı yüz tahmini gerçekleştirmiştir. Farklı yöntemlerin ve veri setlerinin kullanılması başarı oranını etkilediği görülmektedir. Ayrıca Vinay ve ark. [25] ve Coşkun ve ark. 'nın [6] ESA yöntemi ile geliştirdikleri sistemlerde hem zamansal anlamda hem de başarı anlamında ilerleme kaydetmişlerdir. Bu sayede doğrudan görüntü üzerinde çalışan ESA yöntemin ne kadar etkili olduğunu göstermişlerdir.

## Kaynaklar

- [1] Akay, S., 2018. Facial Action Unit Detection In Videos Using Deep Neural Networks (Yüksek Lisans Tezi). BÜ. Fen Bilimleri Enstitüsü. İstanbul.
- [2] Akpınar, B., 2019. Görüntü Sınıflandırma İçin Derin Öğrenme İle Bayesçi Derin Öğrenme Yöntemlerinin Karşılaştırılması (Yüksek Lisans Tezi). AKÜ, Fen Bilimleri Enstitüsü. Afyon.
- [3] Anonim-1., 2020. <http://www.prowmes.com/blog/makine-ogrenmesi/>. Erişim tarihi:09.04.2020.
- [4] Ayata, F., Çavuş, H., İnan, M., Seyyarer, E., Biçek, E., & Kına, E., 2020. Dostroajan: Facial Recognition Based System Input Control Agent. AJIT-e: Online Academic Journal of Information Technology, 11(40), 82–96.

- <https://doi.org/10.5824/ajite.2020.01.005.x>.
- [5] Bayrakdar, S., Akgün, D., & Yücedağ, İ., 2016. Yüz ifadelerinin otomatik analizi üzerine bir literatür çalışması A survey on automatic analysis of facial expressions. SAÜ Fen Bilimleri Dergisi, 20(2), 383–398.
- [6] Coşkun, M., Uçar, A., Yıldırım, Ö., & Demir, Y., 2017. Face Recognition Based on Convolutional Neural Network. 2017 International Conference on Modern Electrical and Energy Systems (MEES). <https://doi.org/10.13810/j.cnki.issn.1000-7210.2019.05.024>.
- [7] Doğan, G., 2010. Yapay Sinir Ağları Kullanarak Türkiye'deki Özel Bir sigorta şirketinde portföy Değerlendirmesi (Yüksek Lisans Tezi). HÜ, Fen Bilimleri Enstitüsü. Ankara.
- [8] Erdoğan, A. Y., 2010. Yüz Tanımda Özyüz Ve Fisher Yüz Algoritmalarının İncelenmesi. (Yüksek Lisans Tezi). AÜ. Fen Bilimleri Enstitüsü. Ankara.
- [9] Farahani, F. S., Sheikhan, M., & Farrokhi, A., 2013. A fuzzy approach for facial emotion recognition. 13th Iranian Conference on Fuzzy Systems, IFSC 2013, 1–4. <https://doi.org/10.1109/IFSC.2013.6675597>.
- [10] Haniççi, A., 2019. Evrimsel Sinir Ağları Kullanılarak Ekg Ve Yüz Tabanlı Biyometrik Tanıma (Yüksek Lisans Tezi). BTÜ, Fen Bilimleri Enstitüsü. Bursa.
- [11] Hasanuzzaman, M. ve Ueno, H. (2007). Face and Gesture Recognition for Human-Robot Interaction. 10.5772/4836.
- [12] Holat, R., 2016. Yüz Bulma Ve Tanıma Sistemleri Kullanarak Kimlik Tespitinin Yapılması (Yüksek Lisans Tezi). DÜ, Fen Bilimleri Enstitüsü. Düzce.
- [13] Lai, C. Q. ve Teoh, S. S., 2016. An Efficient Method of HOG Feature Extraction Using Selective Histogram Bin and PCA Feature Reduction. Advances in Electrical and Computer Engineering. 16. 101-108. 10.4316/AECE.2016.04016.
- [14] Liu, S., Song, Z., Liu, G., Xu, C., Lu, H., ve Yan, S., 2012. Cross-Scenario Clothing Retrieval Via Parts Alignment And Auxiliary Set. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 3330–3337. IEEE, 2012.
- [15] Mallick, S., 2016. Anonim-6. <https://www.learnopencv.com/histogram-of-oriented-gradients/>. Erişim tarihi: 29.04.2020.
- [16] Özbudak, Ö., 2009. Yüz Resimlerinden Cinsiyet Tayini (Yüksek Lisans Tezi). İTÜ, Fen Bilimleri Enstitüsü. İstanbul.
- [17] Pişkin, M., 2020. Yüz Tanıma | Mesut Pişkin. <http://mesutpiskin.com/blog/yuz-tanima.html>. Erişim tarihi: 03.03.2020.
- [18] Singh, A., 2019. HOG Özellik Tanımlayıcısına Değerli Bir Giriş. <https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/>. Erişim Tarihi: 14.10.2020.
- [19] Soydaş, M., 2017. Anonim-8. <http://mehmetsoydas.com/2017/05/27/real-time-face-detection-recognition-YGH-features-svm/>. Erişim tarihi: 29.04.2020.
- [20] Tanrıverdi, M., 2017. Yüz Bulma Ve Tanıma Tabanlı Otomatik Sınıf Yoklama Yönetim Sistemi (Yüksek Lisans Tezi). AÜ, Fen Bilimleri Enstitüsü. Ankara.
- [21] Taşova, O., 2011. Yapay Sinir Ağları İle Yüz Tanıma (Yüksek Lisans Tezi). DEÜ, Fen Bilimleri Enstitüsü. İzmir.
- [22] Thaware, R., 2018. Anonim-7. <https://www.eeweb.com/profile/rajeevthaware/articles/real-time-face-detection-and-recognition-with-svm-and-YGH-features>. Erişim tarihi: 29.04.2020.
- [23] Ulgen, K., 2017. Anonim-10. <https://medium.com/@k.ulgen90/makine-ogrenimi-bolum-4-destek-vektor-makineleri-2f8010824054>. Erişim tarihi: 29.04.2020.
- [24] Varol, A., ve Cebe, B., 2011. Yüz Tanıma Algoritmaları Algorithms Of Face Recognition. 5th International Computer & Instructional Technologies Symposium, September, 22–24.
- [25] Vinay, A., Reddy, D. N., Sharma, A. C., Daksha, S., Bhargav, N. S., Kiran, M. K., Murthy, K. N. B., & Natrajan, S., 2017. G-CNN and F-CNN: Two CNN Based Architectures For Face Recognition. Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDAI 2017, 6, 23–28. <https://doi.org/10.1109/ICBDACI.2017.8070803>.
- [26] Yang, M.H., Hand Gesture Recognition and Face Detection in Images, Ph.D Thesis, University of Illinois, Urbana-Champaign, 2000.

## Kentsel İçme Suyu Şebekelerinde Suyun Kalite Seviyesinin Yapay Zekâ Yaklaşımları Kullanılarak Belirlenmesi

Abdullah ŞENER<sup>1\*</sup>, Burhan ERGEN<sup>2</sup>, Hamit MIZRAK<sup>3</sup>

<sup>1</sup> Bingöl Üniversitesi, Genç Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Bingöl, Türkiye

<sup>2</sup> Firat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Elazığ, Türkiye

<sup>3</sup> Malatya Turgut Özal Üniversitesi, Malatya, Türkiye

\*<sup>1</sup> asener@bingol.edu.tr, <sup>2</sup> bergem@firat.edu.tr, <sup>3</sup> hamitmizrak@gmail.com

(Geliş/Received: 02/01/2022;

Kabul/Accepted: 11/02/2022)

**Öz:** Su, bütün canlıların yaşam dengesini sağlayan ve faaliyetlerini ayakta tutan hayati bir öneme sahiptir. İçme suyunun kalitesi suyun kimyasal, biyolojik, fiziksel ve radyolojik ölçütleriyle bağlantılıdır. Su kalitesi insan sağlığını ve ekolojik sistemi doğrudan etkileyen bir faktördür. Kentsel şebekelerde içme suyunun kalitesini ölçebilmek için birçok teknik ve model geliştirilmiştir. Dünya nüfusunun hızlı artışı ve yaşamsal alanlardaki içme suyu şebekelerinin sayısını da artırmaktadır. Bu tür ihtiyaçlara anlık cevap verebilmek insanlar için zaman alıcı bir süreçtir. Son zamanlardaki teknolojik gelişmeler bu tür problemlerin çözümünde etkin rol üstlenmektedir. Bu çalışmada su bileşenlerini içeren açık veri kümesi kullanılarak yapay zekâ tabanlı analizler gerçekleştirilmiştir. Gerçekleştirilen analizlerde suyun kaliteli olup olmadığı tespit edilmiştir. Deneysel analizlerde makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları kullanılmıştır. En iyi performans sonucu torbalama yöntemi ile elde edilmiştir. Bu yöntem ile %96,44 oranında genel doğruluk başarıları sağlanmıştır.

**Anahtar kelimeler:** Yapay Zekâ, Makine Öğrenme, Su Kalite Ölçümü, Sınıflandırma.

## Determining Water Quality in Urban Drinking Water Networks Using Artificial Intelligence Approaches

**Abstract:** Water has vital importance as it provides balance to all living things and keeps their activities alive. The quality of drinking water depends on the chemical, biological, physical and radiological criteria of water. Water quality is a factor that directly affects human health and ecological system. Many techniques and models have been developed to measure the quality of drinking water in urban networks. With the rapid increase in the world population, the number of drinking water networks in vital areas is also increasing. It is a time-consuming process for people to be able to respond to such needs instantly. Recent technological developments are playing an active role in solving such problems. In this study, artificial intelligence-based analyzes were performed on an open data set with water components. The analyzes performed determined whether the water was of good quality. Machine learning methods and ensemble learning algorithms were used in the experimental analysis. The best performance result was obtained using the bagging method. With this method, an overall accuracy of 96.44% was achieved.

**Key words:** Artificial Intelligence, Machine Learning, Water Quality Measurement, Classification.

### 1. Giriş

Su, canlıların yaşam döngüsünü sağlayan, tatsız ve kokusuz bir bileşiktir [1]. Tatlı su, çevreyle ilgili sistemin gereksinimini karşılayan ve tarımsal kalkınmanın sürdürülebilirliğini sağlayan sınırlı bir kaynaktır. İnsanlar tarafından tüketilen suyun kalitesinin düşük olması birçok hastalığın oluşmasına neden olabilir [2]. Canlılar yaşam kalitelerini koruyabilmeleri ve sağlıklı bir şekilde sürdürebilmeleri için tükettikleri suyun kalitesi önemlidir. Denizler, göller, nehirler gibi su yataklarının kalite seviyeleri belirli işlem adımlarından geçirilerek ölçülebilmektedir. Bununla ilgili birçok teknolojik altyapı sistemi tasarlanmış ve kullanılmıştır. Suyun kalitesi, suyun biyolojik, fiziksel, kimyasal gibi etmenlerine bağlıdır [3]. Suyun kalite ölçümü son derece hayati önem taşımaktadır ve kalite standardına uymayan içme suları insan sağlığını doğrudan olumsuz etkileyebilmektedir [4]. Kalitesiz suyun sonuçları çevreyi, alt yapı sistemlerini veya insanları doğrudan etkileyebilir. Nitekim Birleşmiş Milletler 'in raporuna göre her yıl 1,5 milyon insan kirli ve kalitesi düşük olan içme sularını kullandığından dolayı çeşitli hastalıklara yakalanmakta ve bu hastalıklardan dolayı ölmektedir [5].

Günümüzde suyun kalitesini ölçebilecek birçok model ve yaklaşım geliştirilmiştir. Son zamanlarda teknolojik tabanlı sistemler, yapay zekâ yaklaşımları ile birleştirilerek analizler yapılmaya başlanmıştır [3]. Bu durum karar verme süreçlerinin performansını olumlu etkilemiştir. Literatürde suyun kalite ölçümünü yapay zekâ yaklaşımları

\* Sorumlu yazar: [asener@bingol.edu.tr](mailto:asener@bingol.edu.tr). Yazarların ORCID Numarası: <sup>1</sup> 0000-0002-8927-5638, <sup>2</sup> 0000-0003-3244-2615, <sup>3</sup> 0000-0002-4795-3007

ile gerçekleştiren birçok çalışma bulunmaktadır. Bu çalışmalardan bazıları incelenirse; Aldhyani ve ark. [3], su kalitesini sınıflandırma işleminde doğrusal olmayan otoregresif sinir ağı modeli ile uzun-kısa süreli bellek (LSTM) modelini kullanarak analizler gerçekleştirdi. Ayrıca makine öğrenme yöntemleri de deneysel analizlerde kullandılar. Aldhyani ve ark. [3] çalışmasında en iyi performansı doğrusal olmayan otomatik regresif sinir ağı modeli ile sağladılar ve bu modelden elde ettikleri genel doğruluk başarıları ise %96,17'dir. Khan ve ark. [6], su kalitesini sınıflandırma işleminde temel bileşenler analizi (PCA) yöntemi ile gradyan artırma yöntemini kullandılar. Khan ve ark. [6] çalışmalarında sırasıyla %95 ve %100 genel doğruluk başarıları elde ettiler. Radhakrishnan ve ark. [7] su kalitesini sınıflandırma sürecinde makine öğrenme yöntemlerini kullandılar. Radhakrishnan ve ark. [7] destek karar makineleri (DVM), karar ağaçları (KA), naif bayes (NB) yöntemlerini kullanarak analizlerini gerçekleştirdiler. En iyi sınıflandırma sonucunu KA yöntemi ile sağladılar ve bu yöntem ile %98,50 oranında genel doğruluk başarıları elde ettiler. Venkata ve ark. [8] su kalitesini sınıflandırma sürecinde sinir ağı modellerinden yararlandılar. Venkata ve ark. [8] tekrarlayan sinir ağı (RNN) ve LSTM modellerini deneysel analizlerde kullandılar. En iyi performans sonucunu LSTM yöntemi ile elde ettiler ve bu yöntem ile %94 oranında genel doğruluk başarıları sağladılar.

Bu çalışma şehir şebeke sularının kalite seviyesinin ölçülmesinde makine öğrenme yöntemleri ve topluluk öğrenme algoritmalarının etkinliğini ölçebilmek için gerçekleştirilmiştir. Çalışmanın diğer bölümleri hakkında özet bilgiler şu şekildedir; materyal, makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları hakkında bilgiler ikinci bölümde yer almıştır. Deneysel analizler ve sonuçları hakkında açıklamalara üçüncü bölümde yer verilmiştir. Yazarların önerdikleri yaklaşım hakkındaki yorumlarına tartışma bölümünde yer verilmiştir. Analiz sonuçlarının değerlendirilmesi, gelecek çalışmalar hakkındaki bilgilere sonuç bölümünde yer verilmiştir.

## 2. Materyal ve Yapay Zekâ Yaklaşımları

Bu bölümde veri kümesi, makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları hakkında genel bilgiler verilmiştir.

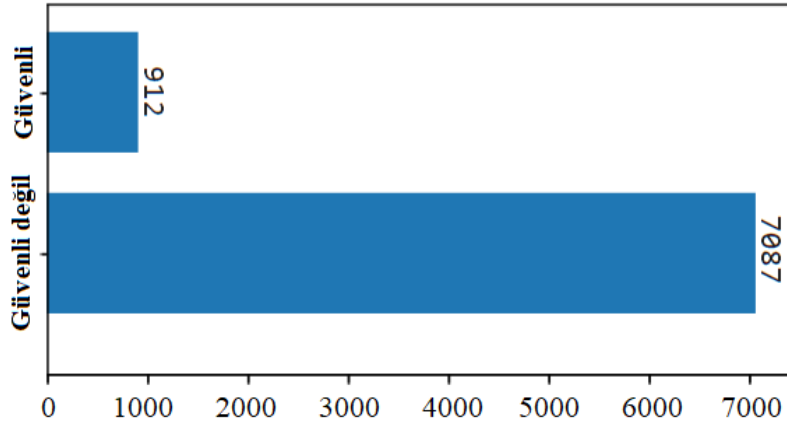
### 2.1. Veri kümesi

Veri kümesi açık erişimlidir ve 'csv' uzantılı dosyadan oluşmaktadır. Veri kümesinde toplam 21 öznitelik sütunu yer almaktadır ve kentsel bir yapıda su kalitesini ölçmeye ilişkin verilerden oluşmaktadır. Verilerin tümü sayısal değerlerden oluşmaktadır. 21. öznitelik "güvenli-1" ve "güvenli değil-0" sayısal etiketlerinden oluşmaktadır. Toplam 7999 örnek numune değerinden oluşmaktadır ve veri kümesindeki her bir öznitelik parametresindeki değerler bir litre içerisindeki su seviyesi dikkate alınarak oluşturulmuştur [9]. Diğer öznitelikler ve açıklamaları Tablo 1'de verilmiştir.

**Tablo 1.** Veri kümesinde yer alan öznitelikler ve açıklamaları.

No	Öznitelik	Tehlike eşik değeri	No	Öznitelik	Tehlike eşik değeri
1	alüminyum	2,8	11	virüsler	0
2	amonyak	32,5	12	kurşun	0,015
3	arsenik	0,01	13	nitratlar	10
4	baryum	2	14	nitritler	1
5	kadmiyum	0,005	15	cıva	0,002
6	kloramin	4	16	perklorat	56
7	krom	0,1	17	radyum	5
8	bakır	1,3	18	selenyum	0,5
9	florür	1,5	19	gümüş	0,1
10	bakteri	0	20	uranyum	0,3

Tablo 1'de özniteliklerin açıklaması dikkate alındığında, alüminyum değeri 1 litre suyun içerisinde 2,8 ve üzeri bir değere sahipse tehlikeli bir seviyededir. Bu durumda su kalitesi güvenli değildir. Diğer özniteliklerin tehlike seviyelerini belirten eşik değerleri Tablo 1'de verilmiştir. Güvenli örneklerin sayısı 717'dir ve güvenli olmayan örneklerin sayısı 7087'dir. Sınıf türünü ve sayısını gösteren grafik Şekil 1'de gösterilmiştir.



Şekil 1. Su kalitesi veri kümesinin sınıf türü ve sınıf türüne ait örnek sayısı.

## 2.2. Makine öğrenme yöntemleri

### 2.2.1. Lojistik regresyon

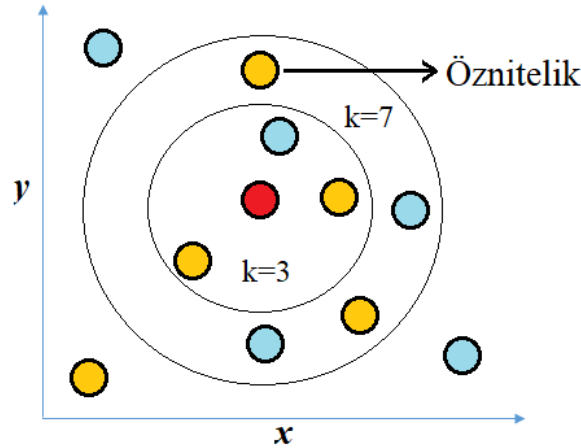
Lojistik regresyon (LR), birden fazla bağımsız değişken parametresini, bağımlı değişkenler üzerinden ilişkisini kurarak tahmin yürütmeyi sağlayan bir makine öğrenme yöntemidir. Tahmin yürütme işleminde ikili bağımlı değişken kullanılır. Yani, iki farklı kategorik değer alabilen bir sonuç üretilir. Örneğin; “doğru-yanlış”, “iyi-kötü”, “1 veya 0” gibi iki değer tahminini yürütür. Burada çıkışın sıfır veya bir olduğu varsayılırsa, bağımsız değişkenin değerinin bir olması bağımlı değişkenin de değerinin bir olma olasılığını artırmaktadır. LR yöntemi doğrusal regresyonun bir uzantısıdır. LR yönteminde çıkış değeri ikiden fazla ise her bir bağımsız değer sırasıyla sabit tutularak ikili sınıflandırma gerçekleştirilir ve çoklu sınıflandırma sürecinde ise tüm bağımsız değer sonuçları karşılaştırılarak girdi verisinin sınıflandırma işlemi gerçekleştirilir [10,11].

Olasılıksal bir değer üretme işleminde bağımsız değişken ile bağımlı bir değişken arasında sigmoidal bir ilişki vardır. Bu sebeple girdi verileri için üretilen tahmin değeri 0 ile 1 arasındadır. Bu durumun matematiksel işlemi Denklem 1'e göre gerçekleştirilir. Denklem 1 incelendiğinde; olasılık ( $p$ ), regresyon katsayıları ( $b_0$ ,  $b_1$ ), bağımsız değişken ( $X$ ) ile temsil edilmektedir [10].

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \times X \quad (1)$$

### 2.2.2. K-En yakın komşu

En yakın komşu (kNN) yöntemi, girdi verisini kendisinden önceki girdi verilerinin yakınlık derecesine bağlı olarak sınıflandırma ve regresyon işlemini gerçekleştirebilen bir makine öğrenme yöntemidir. kNN yöntemi mevcut girdi verilerinin durumunu saklar, yeni veri durumlarını da mevcut veri durumlarını dikkate alarak benzerlik ölçütlerini elde eder. Benzerlik olasılığı hesaplanırken özniteliklerin bulunduğu uzayda iki öznitelik arasındaki mesafeler hesaplanır. Ardından komşu öznitelikler ile mesafeler hesaplanır ve özniteliklerin sınıflandırma sürecinde komşu öznitelikler arasında mesafe uzunluğu en yakın olan özniteliklerin sınıf etiketine ataması gerçekleştirilir. İki öznitelik arasındaki mesafe uzunluğu hesaplanırken genellikle Öklid yöntemi tercih edilir. kNN yönteminde  $k$  değeri  $\{2,3,5,\dots\}$  gibi değerlerden biri seçilerek sınıflandırma işlemi gerçekleştirilir. Örneğin;  $k = 3$  olarak tercih edildiğinde uzay kümesinde mevcut özniteliklerin etrafındaki en yakın üç özniteliklere göre uzaklık mesafeleri ölçülür ve mevcut öznitelik kendisine en yakın özniteliklerin sınıf türüyle etiketlenir [12,13].  $k$  değerinin ikili sınıflandırma işleminde seçimi Şekil 2'de gösterilmiştir.



Şekil 2. kNN yönteminde  $k$  değerine göre özniteliklerin seçimi.

### 2.2.3. Karar ağacı

Karar ağacı (KA) yöntemi, karar ve yaprak düğümlerinden oluşan sınıflandırma ve regresyon işlemi için entropi değerlerini hesaplayan grafiksel tabanlı şematik bir makine öğrenme yöntemidir. Bu yöntemde sınıflandırma ve regresyon işlemleri için (ID3, C4.5, CART, vb.) algoritmalar kullanılır. Algoritma şemasında sırasıyla kök, düğüm ve yapraklar yer alır. Verilerin belirsizlikleri hesaplarken entropi ( $E$ ) değerleri dikkate alınır. Entropi ( $E$ ) değeri olasılık değerlerinin hesaplanmasında kullanılan önemli bir parametredir. Entropi parametresinin hesaplanmasında Denklem 2’de belirtilen formül kullanılır. Denklem 2 incelendiğinde,  $N$  değişkeni toplam öznitelik sayısını temsil eder ve  $P$  değişkeni ise  $i$ . özniteliğin olasılık değerini temsil eder [14,15]. Sınıflandırma sürecinde entropi değeri minimum olan yaprak tercih edilir ve etiketlenir.

$$E = - \sum_{i=1}^N P_i \log_2 P_i \quad (2)$$

### 2.2.4. Rastgele Orman

Rastgele orman (RO) yöntemi, birden fazla karar ağacının bir araya gelmesi ile oluşturulmuş, sınıflandırma ve regresyon işlemlerinde kullanılan bir makine öğrenme yöntemidir. RO yöntemi, KA yöntemine göre algoritmasında birden fazla karar ağacı algoritması içerdiği için hata oranını daha doğru tahmin eder. Bir araya getirilen karar ağaçları “orman” topluluğunu oluşturur ve ağaçlar arasında korelasyon daha düşük seviyededir. Orman topluluğu içerisinde rastgele düğüm seçimi gerçekleştirilir ve ardından en iyi düğümü belirlemek için *Gini* fonksiyonu kullanılır. *Gini* fonksiyonu düğümlerin homojenlik değerini ölçer. Ağaç yapısındaki bir dalın alt düğümünün *Gini* değeri üst düğümünün *Gini* değerinden daha düşükse, düğümlerin bulunduğu dal başarılı sayılır. *Gini* fonksiyonun hesaplanmasında Denklem 3 kullanılır. Bu denklemde, toplam öznitelik sayısı  $N$  değişkeni ile temsil edilir ve seçilmiş öznitelik değeri  $n$  değişkeni gösterilir. Seçilmiş özniteliğin kendisinden küçük öznitelik sayısının kendisinden büyük öznitelik sayısına bölümünün karesi ise  $p_i$  değişkeni ile temsil edilir [16,17].

$$Gini(N) = 1 - \sum_{i=1}^n p_i^2 \quad (3)$$

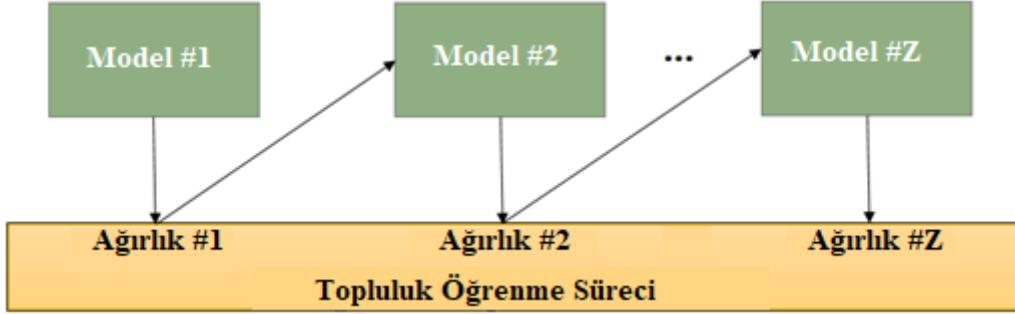
## 2.3. Topluluk öğrenme algoritmaları

### 2.3.1. Uyarlanabilir artırma

Uyarlanabilir artırma (AdaBoost) algoritmasında ağırlık değerleri her bir girdi verisi için yeniden oluşturulmaktadır. Yanlış sınıflandırılan girdi verilerine ise daha yüksek değerlere sahip ağırlıklar atanmaktadır.



Bu işlem döngüsel olarak sonlanana kadar devam etmektedir ve ağırlık değerlerinin artırılması gerçekleştirildiği için bu algoritmaya uyarlanabilir artırma denilmektedir. AdaBoost algoritmasında "Z" sayıda karar ağacı olduğunu varsayarsak, Z sayıda model ve Z sayıda ağırlıklar oluşturulur. Birinci modelde yanlış sınıflandırılan girdi verileri tespit edilir ve ikinci modele kayıtları aktarılır. Bu işlem 3., 4., ..., Z. modele kadar devam eder ve her model arası ağırlık aktarımları da artırılarak gerçekleştirilir. Z sayıda modelin her biri eğitim sürecini güçlendirmek ve performansı artırabilmek için tasarlanmıştır [18]. Bu süreci gösteren tasarım Şekil 3'te gösterilmiştir.



Şekil 3. AdaBoost algoritmasının modeller arasında ağırlık aktarımı.

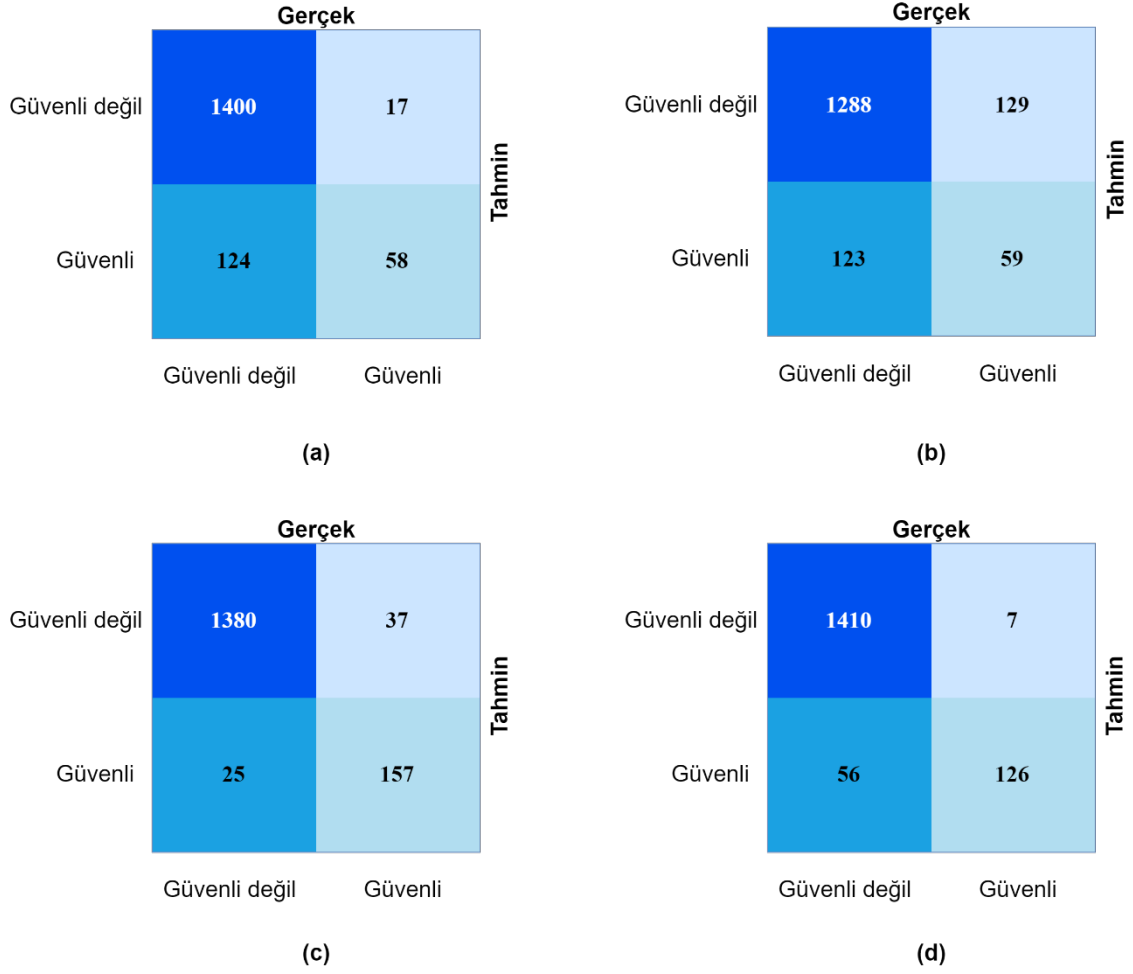
### 2.3.2. Aşırı gradyan artırma

Aşırı gradyan artırma (XGBoost), sınıflandırma ve regresyon tahminlerini gerçekleştiren modelleri içerisinde barındıran topluluk algoritmasıdır. Bu algoritmada AdaBoost algoritması gibi karar ağaçlarını kullanarak her model adımında yanlış girdi verilerini tespit ederek bir sonraki modele aktarılmasını sağlar. Bu işlem süresinde modellerin ağırlıkları gradyan tabanlı artırma ile sürekli güncellenir. XGBoost algoritması, AdaBoost algoritmasının zorlandığı ve hatalı verileri tespit etmede yaşanan güçlükleri tedarik edebilmek için tasarlanmıştır. XGBoost algoritmasının model mimarisi farklıdır ve modellerden çıkabilecek hata kaybını en aza indirebilmek için gradyan iniş algoritmasını kullanmıştır [19].

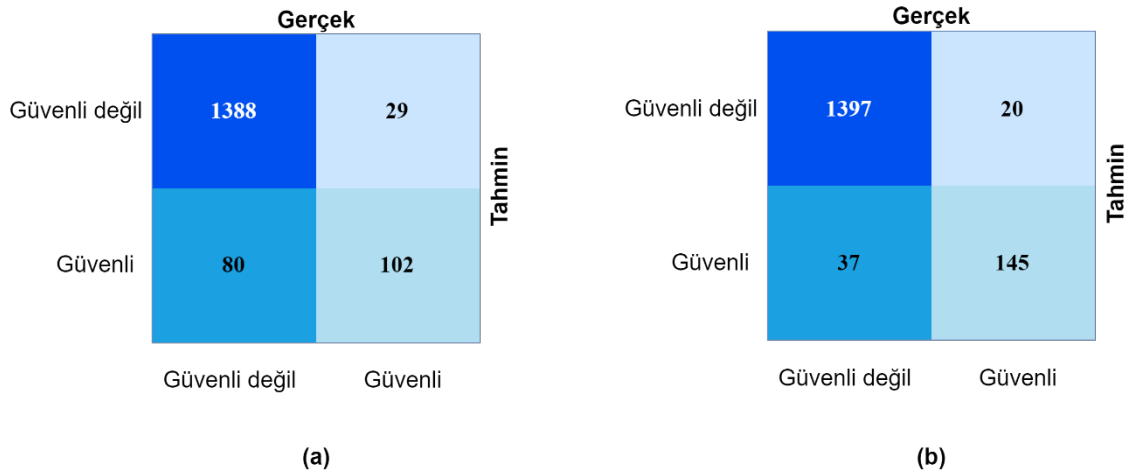
## 3. Deneysel Analizi ve Sonuçları

Bu çalışmanın deneysel analizleri Python yazılım dili ve Jupyter Notebook arayüzü kullanılarak gerçekleştirilmiştir. Yazılım kodlarının derlenmesinde donanım desteği olarak Google Colaboratory sunucusu kullanılmıştır. Google Colaboratory sunucusunun sunmuş olduğu donanım özellikleri; işlemci türü Intel® Xeon® CPU @2.20 GHz, ekran kartı Tesla K80, hafıza kartı 8 GB. Ekran kartının GPU destekleme özelliğinin olması, çalışmaların deneysel analizlerini daha hızlı gerçekleşmesini sağlamıştır [20,21]. Analiz sonuçlarının ölçülmesinde karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisi hesaplanırken; kesinlik (precision), geri çağırma (recall), F1-skor ve doğruluk (accuracy) metrikleri kullanılmıştır [22–24]. Makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları Sklearn Kütüphanesi tarafından desteklendi ve bu yöntemler için varsayılan değerler kabul edilerek kullanılmıştır [25,26].

Deneysel analizlerin tümünde veri kümesinin %20'si test verisi olarak kullanılmıştır. Veri kümesinin %80'i eğitim verisi olarak kullanılmıştır. Veri kümesi sırasıyla LR, kNN, KA, RO makine öğrenme yöntemleri ile analiz edilmiştir. Ardından AdaBoost, XGBoost topluluk öğrenme algoritmaları kullanılarak analizler gerçekleştirilmiştir. Makine öğrenme yöntemlerindeki genel doğruluk başarıları sırasıyla %91.18, %84.24, %96.12, %96.06 elde edilmiştir. Topluluk öğrenme algoritmalarından elde edilen genel doğruluk başarıları sırasıyla %93.18, %96.44'tür. Makine öğrenme yöntemleri ve topluluk öğrenme algoritmalarından elde edilmiş karmaşıklık matris grafikleri Şekil 4 ve Şekil 5'te sırasıyla gösterilmiştir.



Şekil 4. Makine öğrenme yöntemlerinden elde edilmiş karmaşıklık matrisleri; a) LR, b) kNN, c) KA, d) RO.



Şekil 5. Topluluk öğrenme algoritmalarından elde edilmiş karmaşıklık matrisleri; a) AdaBoost, b) XGBoost.

Karmaşıklık matrislerin metrik sonuçları Tablo 2’de verilmiştir. Öğrenme yöntemleri arasında en iyi performansı topluluk öğrenme algoritmalarında yer alan XGBoost vermiştir. XGBoost algoritması ile %96,44 oranında genel doğruluk başarısı elde edilmiştir. Ayrıca veri kümesinin sınıfları arasında dengesiz dağılımdan

dolayı yöntemlerin F1 skorları da karşılaştırılmıştır. F1-skorları arasında en iyi performansı XGBoost algoritması ile edilmiştir ve elde edilmiş F1-skor oranı %98'dir.

**Tablo 2.** Deneysel analizlerden elde edilmiş karmaşıklık matrislerin metrik sonuçları.

Yöntem	Sınıf	Kesinlik(%)	Geri çağırma(%)	F1-skor(%)	Doğruluk (%)
LR	Güvenli değil	92	99	95	91,18
	Güvenli	77	32	45	
kNN	Güvenli değil	91	91	91	84,24
	Güvenli	31	32	32	
KA	Güvenli değil	98	97	98	96,12
	Güvenli	81	86	84	
RO	Güvenli değil	96	100	98	96,06
	Güvenli	95	69	80	
AdaBoost	Güvenli değil	95	98	96	93,18
	Güvenli	78	56	65	
XGBoost	Güvenli değil	97	99	98	96,44
	Güvenli	88	80	84	

#### 4. Tartışma

Yapılan çalışmada suyun kalite değerini sınıflandırabilen yapay zekâ tabanlı yöntemler kullanılmıştır. Çalışmada veri kümesindeki veri sınıflarının dengesiz dağılımı F1-skorların önemini ön plana çıkarmıştır. Nitekim F1-skorları arasında da XGBoost algoritması genel doğruluk başarısında olduğu gibi daha iyi sonuç vermiştir. Ayrıca veri kümesindeki özniteliklerin sayısının 20 olması ön işlem adımlarının ( öznitelik seçme, özellik çıkarma, vb.) uygulanması açısından sağlıklı olmayacağı düşünülmüştür. Sınıflandırma performansına katkı sağlayabilmek için belki evrimsel bir sinir ağı tasarlanabilirdi ve geliştirilmiş derin öğrenme modeline bu çalışmada kullanılmış makine öğrenme yöntemleri sınıflandırma aşamasında kullanılabilirdi. Bu çalışmaya benzer analizler ve sonuçları Tablo 3'te verilmiştir. Bu analizler incelenirse;

**Tablo 3.** Literatürde yer alan benzer çalışmalar ile bu çalışmanın karşılaştırılması.

Makale	Yıl	Yöntem	Doğruluk (%)
Sillberg ve ark. [27]	2021	DVM	95
Shahra ve ark. [28]	2021	Yapay Sinir Ağı (YSA)	94
Al-Adhaileh ve ark. [29]	2021	Nöro-bulanık çıkarım sistemi, İleri beslemeli sinir ağı	100
Bu çalışma	2022	XGBoost	96,44

Sillberg ve ark. [27] su kalitesi indeksi verilerini kullanarak nehir suyunun kalite düzeyini sınıflandırmışlardır. Sillberg ve ark. [27] önerdikleri yaklaşımda öznitelik gerçekleştirme ve DVM yöntemini hibrit bir model tasarlayarak kullanmışlardır. Sınıflandırma sonucunda %95 oranında genel doğruluk başarısı elde etmişlerdir. Sillberg ve ark. [27] DVM dışında farklı sınıflandırma yöntemlerini kullanabilselerdi belki de performans sonucunu artırabilirdi. Shahra ve ark. [28] içme suyu verilerini kullanarak suyun kaliteli olup olmadığını tespit edebilmek için yapay zekâ yaklaşımlarını kullanmışlardır. Shahra ve ark. [28] önerdikleri yaklaşımda YSA modelini ve DVM yöntemini kullanmışlardır. Sınıflandırma sonucunda YSA model ile en iyi performansı elde etmişlerdir ve %94 oranında genel doğruluk başarısı elde etmişlerdir. Shahra ve ark. [28] derin

ağlar ile bir YSA modeli tasarımları performans sonucunu olumlu etkilemiştir. Al-Adhaileh ve ark. [29] içme ve atık su arıtma verilerini kullanarak suyun kaliteli olup olmadığını tespit eden analizler gerçekleştirmişlerdir. Al-Adhaileh ve ark. [29] nöro-bulanık çıkarım sistemi algoritmasını kullandılar ve sınıflandırma sürecinde ileri beslemeli sinir ağı ile kNN yöntemini kullandılar. En iyi performansı nöro-bulanık çıkarım sistemi ile birlikte kullandıkları ileri beslemeli sinir ağından elde etmişlerdir. Al-Adhaileh ve ark. [29] çalışmalarında %100 genel doğruluk başarısı elde etmişlerdir. Onların çalışması bizim çalışmamıza kıyasla daha özgün yaklaşımlar içermiştir ve bunun karşılığında performansı olumlu katkı sağlamıştır.

## 5. Sonuç

Su sağlığımız ve yaşam kalitemiz için önemlidir. Tüketilen suyun kalitesinin düşük ve kirli olması insanların sağlığını tehdit etmektedir. Suyun kalitesinin ölçümü çeşitli teknoloji tabanlı modeller kullanılarak ölçülebilmektedir. Bu çalışmada suyun kalite değerini sınıflandırabilen yapay zekâ tabanlı yöntemler kullanılmıştır. Çalışmanın analizlerinde elde edilen performans sonuçları umut vericidir. Yapılan çalışma sayesinde kentsel içme suyu şebekelerinde suyun kalite seviyesinin belirlenmesinde doğruluk oranı yüksek sonuçlar elde edilmektedir.

Gelecek çalışmada, çeşitli veri kümeleri kullanılarak yeni bir derin öğrenme modeli geliştirilecektir. Derin öğrenme modelinde ön işlem adımları ve son işlem adımlarına yer verilerek performans artışı sağlanacaktır.

## Kaynaklar

- [1] A. Boretti, L. Rosa, Reassessing the projections of the World Water Development Report, *Npj Clean Water*. 2 (2019) 15. doi:10.1038/s41545-019-0039-9.
- [2] R.J. Hogeboom, The Water Footprint Concept and Water's Grand Environmental Challenges, *One Earth*. 2 (2020) 218–222. doi:https://doi.org/10.1016/j.oneear.2020.02.010.
- [3] T.H.H. Aldhyani, M. Al-Yaari, H. Alkahtani, M. Maashi, Water Quality Prediction Using Artificial Intelligence Algorithms, *Appl. Bionics Biomech*. 2020 (2020) 6659314. doi:10.1155/2020/6659314.
- [4] M. Allaire, H. Wu, U. Lall, National trends in drinking water quality violations, *Proc. Natl. Acad. Sci.* 115 (2018) 2078–2083. doi:10.1073/pnas.1719805115.
- [5] Water | United Nations, UN Water. (2021). https://www.un.org/en/global-issues/water (accessed December 29, 2021).
- [6] M.S. Islam Khan, N. Islam, J. Uddin, S. Islam, M.K. Nasir, Water quality prediction and classification based on principal component regression and gradient boosting classifier approach, *J. King Saud Univ. - Comput. Inf. Sci.* (2021). doi:https://doi.org/10.1016/j.jksuci.2021.06.003.
- [7] N. Radhakrishnan, A.S. Pillai, Comparison of Water Quality Classification Models using Machine Learning, in: 2020 5th Int. Conf. Commun. Electron. Syst., 2020: pp. 1183–1188. doi:10.1109/ICCES48766.2020.9137903.
- [8] D. Venkata Vara Prasad, L. Y Venkataramana, P.S. Kumar, G. Prasannamedha, K. Soumya, P. A.J., Water quality analysis in a lake using deep learning methodology: prediction and validation, *Int. J. Environ. Anal. Chem.* (2020) 1–16. doi:10.1080/03067319.2020.1801665.
- [9] P. Smarty, Water Quality Dataset, Kaggle. (2021). https://www.kaggle.com/mssmartypants/water-quality (accessed December 31, 2021).
- [10] P. Schober, T.R. Vetter, Logistic Regression in Medical Research, *Anesth. Analg.* 132 (2021) 365–366. doi:10.1213/ANE.0000000000005247.
- [11] M. Toğaçar, B. Ergen, M.E. Sertkaya, Ztürre Hastalığının Derin Öğrenme Modeli ile Tespiti, *Fırat Üniversitesi Mühendislik Bilim. Derg.* 31 (2019) 223–230.
- [12] M. Jawthari, V. Stoffová, Predicting students' academic performance using a modified kNN algorithm, *Pollack Period.* 16 (2021) 20–26. doi:10.1556/606.2021.00374.
- [13] M.E. Sertkaya, B. Ergen, M. Togacar, Diagnosis of Eye Retinal Diseases Based on Convolutional Neural Networks Using Optical Coherence Images, in: 2019 23rd Int. Conf. Electron., 2019: pp. 1–5. doi:10.1109/electronics.2019.8765579.
- [14] A. Topîrceanu, G. Grossecck, Decision tree learning used for the classification of student archetypes in online courses, *Procedia Comput. Sci.* 112 (2017) 51–60. doi:https://doi.org/10.1016/j.procs.2017.08.021.
- [15] H. Polat, M. Turkoglu, O. Polat, Deep network approach with stacked sparse autoencoders in detection of DDoS attacks on SDN-based VANET, *IET Commun.* 14 (2020) 4089–4100. doi:10.1049/iet-com.2020.0477.
- [16] V. Tümen, B. Ergen, Intersections and crosswalk detection using deep learning and image processing techniques, *Phys. A Stat. Mech. Its Appl.* 543 (2020) 123510. doi:10.1016/j.physa.2019.123510.
- [17] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning, *Stata J. Promot. Commun. Stat. Stata*. 20 (2020) 3–29. doi:10.1177/1536867X20909688.
- [18] V. Chang, T. Li, Z. Zeng, Towards an improved Adaboost algorithmic method for computational financial analysis, *J. Parallel Distrib. Comput.* 134 (2019) 219–232. doi:10.1016/j.jpdc.2019.07.014.
- [19] A. Ibrahim Ahmed Osman, A. Najah Ahmed, M.F. Chow, Y. Feng Huang, A. El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, *Ain Shams Eng. J.* (2021).

- doi:<https://doi.org/10.1016/j.asej.2020.11.011>.
- [20] T. Carneiro, R.V.M. Da Nóbrega, T. Nepomuceno, G. Bian, V.H.C. De Albuquerque, P.P.R. Filho, Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications, *IEEE Access*. 6 (2018) 61677–61685. doi:10.1109/access.2018.2874767.
- [21] E. Başaran, Z. Cömert, A. Şengür, Ü. Budak, Y. Çelik, M. Toğaçar, Chronic Tympanic Membrane Diagnosis based on Deep Convolutional Neural Network, in: 2019 4th Int. Conf. Comput. Sci. Eng., 2019: pp. 1–4. doi:10.1109/ubmk.2019.8907070.
- [22] M. Toğaçar, B. Ergen, Deep Learning Approach for Classification of Breast Cancer, in: 2018 Int. Conf. Artif. Intell. Data Process., 2018: pp. 1–5. doi:10.1109/idap.2018.8620802.
- [23] M. Toğaçar, B. Ergen, Biyomedikal Görüntülerde Derin Öğrenme ile Mevcut Yöntemlerin Kıyaslanması, *Fırat Üniversitesi Mühendislik Bilim. Derg.* 31 (2019) 109–121.
- [24] K.-S. Cheng, J.-Y. Ling, T.-W. Lin, Y.-T. Liu, Y.-C. Shen, Y. Kono, Quantifying Uncertainty in Land-Use/Land-Cover Classification Accuracy: A Stochastic Simulation Approach, *Front. Environ. Sci.* 9 (2021) 46. doi:10.3389/fenvs.2021.628214.
- [25] Scikit-learn developers, Machine learning, Scikit-Learn. (2019). [https://scikit-learn.org/stable/supervised\\_learning.html#supervised-learning](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning) (accessed January 2, 2022).
- [26] Scikit-learn developers, Ensemble methods, Scikit-Learn. (2019). <https://scikit-learn.org/stable/modules/ensemble.html> (accessed January 2, 2022).
- [27] C. Sillberg, P. Kullavanijaya, O. Chavalparit, Water Quality Classification by Integration of Attribute-Realization and Support Vector Machine for the Chao Phraya River, *J. Ecol. Eng.* 22 (2021) 70–86. doi:10.12911/22998993/141364.
- [28] E.Q. Shahra, W. Wu, S. Basurra, S. Rizou, Deep Learning for Water Quality Classification in Water Distribution Networks, in: 2021: pp. 153–164. doi:10.1007/978-3-030-80568-5\_13.
- [29] M. Hmoud Al-Adhaileh, F. Waselallah Alsaade, Modelling and Prediction of Water Quality by Using Artificial Intelligence, *Sustainability*. 13 (2021) 4259. doi:10.3390/su13084259.





## Fırat Üniversitesi Fen Bilimleri Dergisi

### Telif Hakkı Devir Formu / Copyright Form

Biz aşağıda gerekli bilgileri bulunan,

[Yazarların Adı / Author names]

.....tarafımızdan yazılmış,

[Makale Adı / Article title]

.....

başlıklı makale içeriği, sunduğu sonuç, bilgi, bulgu ve yorumları konusunda, Fırat Üniversitesi Fen Bilimleri Dergisi Editörlüğü'nün hiç bir sorumluluk taşımadığını kabul ederiz.

Sunduğumuz makalenin orijinal olduğunu; herhangi bir başka dergiye yayınlanmak üzere gönderilmediğini; daha önce yayınlanmadığını; tümüyle yada bir bölümü herhangi bir kongre/sempozyum da sunulup fakat tam metin olarak basılmadığını ve orijinal telif hakkı devri formu ile birlikte Fırat Üniversitesi Fen Bilimleri Dergisi Editörlüğü'ne gönderildiğini garanti ederiz.

Makalenin telif hakkından feragat etmeyi kabul ederek sorumluluğu üstlenir ve imza ederiz. Bu vesileyle makalenin telif hakkı Fırat Üniversitesi Fen Bilimleri Enstitüsü' ne devredilmiştir. Fırat Üniversitesi Fen Bilimleri Dergisi Editörlüğü makalenin yayınlanabilmesi konusunda yetkili kılınmıştır. Bununla birlikte yazarların aşağıdaki hakları saklıdır:

Aşağıdaki ve diğer bütün durumlarda makalenin Fırat Üniversitesi Fen Bilimleri Dergisi tarafından yayımlandığına dair referans verilmelidir.

1. Telif Hakkı dışında kalan patent v.b. bütün tescil edilmiş haklar.
2. Makaleyi satmamak koşulu ile kendi amaçları için çoğaltma hakkı.
3. Yazarın gelecekteki kitaplar ve dersler gibi çalışmalarında; makalenin tümü ya da bir bölümünü ücret ödemeksizin kullanma hakkı.

**Bütün yazarlar tarafından imzalanması gerekiyor:** (Sorumlu yazar/Corresponding author\*)

Yazarın Adı ve Soyadı / Author Names

İmzası / Signature

Tarih / Date

1 .....

2 .....

3 .....

4 .....

5 .....

**NOT:** Eksik imza durumunda sorumluluk imzalayan yazarlara aittir. Lütfen formu doldurunuz, imzalayınız ve aşağıdaki adrese metinle birlikte gönderiniz.

Adres: Fırat Üniversitesi Fen Bilimleri Dergisi Editörlüğü, F.Ü. Fen Bilimleri Enstitüsü, 23119-Elazığ.