



**MOBİL UYGULAMA AĞ TRAFİĞİNİN DERİN ÖĞRENME
YÖNTEMLERİ KULLANILARAK SINIFLANDIRILMASI**

İmren DAŞDEMİR YAŞAR

**YÜKSEK LİSANS TEZİ
BİLGİ GÜVENLİĞİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2023

ETİK BEYAN

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

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

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

İmren DAŞDEMİR YAŞAR

06/07/2023

**MOBİL UYGULAMA AĞ TRAFİĞİNİN DERİN ÖĞRENME YÖNTEMLERİ
KULLANILARAK SINIFLANDIRILMASI
(Yüksek Lisans Tezi)**

İmren DAŞDEMİR YAŞAR

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
Temmuz 2023**

ÖZET

Bilişim alanında ortaya çıkan hızlı ilerlemeler ticari ve bireysel internet kullanımını her geçen gün daha da artırmaktadır. Ayrıca, internet kullanım karakteristiği sabit masaüstü iş istasyonlarından mobil yöne doğru kaymaktadır. Tez çalışmasında, son yıllarda genel internet kullanımında ağırlığı gittikçe artan mobil uygulama ağ trafiğinin sınıflandırılması üzerine odaklanılmıştır. Bu kapsamda, mobil uygulama ağ trafiğinin makine öğrenmesi, derin öğrenme ve özellik seçme yöntemleri ile otomatik sınıflandırılması gerçekleştirilmiştir. Tez çalışmasında, sırasıyla 2019 ve 2022 yılında yayınlanarak araştırmacıların kullanımına sunulan MIRAGE-2019 ve MIRAGE-COVID-CCMA-2022 olmak üzere iki güncel veri seti kullanılmıştır. MIRAGE-2019 veri seti AccuWeather, Comics, Dropbox, Duolingo, Facebook, FourSquare, Groupon, Messenger, OneFootball, Pinterest, Slitter.io, Spotify, Subito, Trello, TripAdvisor, Twitter, Viber, Waze, Wish ve Youtube olmak üzere toplam 20 adet mobil uygulamaya ait 317 731 adet mobil ağ trafiği verisi içermektedir. MIRAGE-COVID-CCMA-2022 veri seti ise Discord, GotoMeeting, Meet, Messenger, Skype, Slack, Teams, Webex ve Zoom olmak üzere toplam 9 adet iletişim ve iş birliği mobil uygulamasını ait 263 619 adet mobil ağ trafiği verisi içermektedir. Bu kapsamda, tez çalışmasının birinci sınıflandırma uygulamasında MIRAGE-2019 veri seti üzerinden 20 sınıflı, ikinci sınıflandırma uygulamasında ise MIRAGE-COVID-CCMA-2022 veri seti üzerinden 9 sınıflı otomatik sınıflandırma gerçekleştirilmiştir. Sınıflandırma uygulamalarının deneylerinde makine öğrenmesi olarak Destek Vektör Makinesi, k-En Yakın Komşu, Adaptif Güçlendirme ve Rastgele Orman sınıflandırıcılardan; derin öğrenme olarak ise Evrişimli Sinir Ağları, Uzun Kısa Süreli Bellek, Çift Yönlü Uzun Kısa Süreli Bellek ve Kapılı Tekrarlayan Birim sınıflandırıcılardan yararlanılmıştır. Özellik seçme işlemleri ise Minimum Fazlalık Maksimum Uygunluk, k-En Yakın Komşu ve Rastgele Orman olmak üç farklı yöntem ile gerçekleştirilmiştir. Sınıflandırma uygulama sonuçları MIRAGE-2019 veri seti için %90'ın ve MIRAGE-COVID-CCMA-2022 veri seti için %95'in üzerinde bir genel doğruluk ve ağırlıklı F1-Skoru başarısı ile otomatik sınıflandırma yapılabileceğine işaret etmektedir. Tez çalışması sonucunda elde edilen bir diğer önemli çıkarım ise özellik seçme aşamasının sınıflandırma sürecine dahil edilmesinin hemen hemen bütün sınıflandırıcılar için sonuçları olumlu yönde etkilediğidir. Ayrıca gerek çalışma süresi gerek sınıflandırma başarısı açısından en uygun sınıflandırma yapısının Minimum Fazlalık Maksimum Uygunluk ile seçilen özelliklerin Rastgele Orman sınıflandırıcı ile sınıflandırılması olduğu sonucuna varılmıştır.

Bilim Kodu : 92403

Anahtar Kelimeler : Akış özellikleri, derin öğrenme, makine öğrenmesi, mobil ağ trafik sınıflandırması, özellik seçme

Sayfa Adedi : 52

Danışman : Prof. Dr. Aysun COŞKUN

CLASSIFICATION OF MOBILE APPLICATION NETWORK TRAFFIC USING DEEP LEARNING METHODS

(M. Sc. Thesis)

İmren DAŞDEMİR YAŞAR

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2023

ABSTRACT

Rapid advances in the field of informatics are increasing commercial and individual internet usage day by day. Also, internet usage characteristic is shifting from fixed desktop workstations to mobile. In this thesis, we focus on the classification of mobile application network traffic, which has become increasingly dominant in general internet usage in recent years. In this context, automatic classification of mobile application network traffic is performed using machine learning, deep learning and feature selection methods. In this thesis, two current datasets, MIRAGE-2019 and MIRAGE-COVID-CCMA-2022, published in 2019 and 2022, respectively, were used. The MIRAGE-2019 dataset contains 317 731 mobile network traffic data from 20 mobile applications including AccuWeather, Comics, Dropbox, Duolingo, Facebook, FourSquare, Groupon, Messenger, OneFootball, Pinterest, Slitter.io, Spotify, Subito, Trello, TripAdvisor, Twitter, Viber, Waze, Wish and Youtube. The MIRAGE-COVID-CCMA-2022 dataset contains 263 619 mobile network traffic data for 9 communication and collaboration mobile applications, namely Discord, GotoMeeting, Meet, Messenger, Skype, Slack, Teams, Webex and Zoom. In this context, twenty-class automatic classification was performed on the MIRAGE-2019 dataset in the first classification application of the thesis and nine-class automatic classification was performed on the MIRAGE-COVID-CCMA-2022 dataset in the second classification application. In the experiments of the classification applications, Support Vector Machine, k-Nearest Neighbor, Adaptive Boosting and Random Forest classifiers were used as machine learning; Convolutional Neural Networks, Long Short Term Memory, Bidirectional Long Short Term Memory and Gated Recurrent Unit classifiers were used as deep learning. Feature selection was performed using three different methods: Minimum Redundancy Maximum Relevance, k-Nearest Neighbor and Random Forest. The classification application results indicate that automatic classification can be performed with an overall accuracy and weighted F1-Score success of over 90% for the MIRAGE-2019 dataset and over 95% for the MIRAGE-COVID-CCMA-2022 dataset. Another important conclusion of this thesis is that the inclusion of the feature selection stage in the classification process positively affects the results for almost all classifiers. Also, it is concluded that the most appropriate classification structure in terms of both working time and classification success is the classification of the features selected with Minimum Redundancy Maximum Relevance with Random Forest classifier.

Science Code : 92403

Key Words : Flow features, deep learning, machine learning, mobile network traffic classification, feature selection

Page Number : 52

Supervisor : Prof. Dr. Aysun COŞKUN

TEŞEKKÜR

Yüksek lisans tez çalışmalarında büyük emeği geçen, çok kıymetli yardım ve yönlendirmeleri ile yolumu aydınlatan, danışmanım “Gazi Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği” öğretim üyesi Prof. Dr. Aysun COŞKUN’a katkılarından dolayı teşekkürlerimi sunarım. Yüksek lisans eğitimim esnasında akademik olarak yetişmemde katkıları olan “Gazi Üniversitesi Fen Bilimleri Enstitüsü Bilgi Güvenliği Mühendisliği Ana Bilim Dalı” öğretim üyelerine şükranlarımı iletirim. Son olarak, hayatımın her adımında desteklerini hep hissettiğim değerli aileme, eşim Dr. Hüseyin YAŞAR’a ve en büyük motivasyon kaynağım olan oğullarım Metehan YAŞAR ve Oğuzhan YAŞAR’a en içten minnet duygularıyla teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	xii
SİMGELER VE KISALTMALAR.....	xiii
1. GİRİŞ.....	1
2. LİTERATÜR ÇALIŞMALARI.....	5
3. MATERYAL VE METOT.....	7
3.1. Veri Setleri.....	7
3.1.1. MIRAGE-2019 veri seti	7
3.1.2. MIRAGE-COVID-CCMA-2022 veri seti.....	9
3.2. Makine Öğrenmesi Sınıflandırıcılar	10
3.2.1. Destek vektör makinesi	10
3.2.2. k-En yakın komşu.....	12
3.2.3. Adaptif güçlendirme	12
3.2.4. Rastgele orman	13
3.3. Derin Öğrenme Sınıflandırıcılar	14
3.3.1. Evrişimli sinir ağları	14
3.3.2. Uzun kısa süreli bellek.....	15
3.3.3. Çift yönlü uzun kısa süreli bellek	16
3.3.4. Kapılı tekrarlayan birim.....	17

	Sayfa
3.4. Özellik Seçme Yöntemleri.....	18
3.5. Değerlendirme Kriterleri	19
4. DENEYLER VE SONUÇLAR.....	23
4.1. Deneyler.....	23
4.2. MIRAGE-2019 Veri Seti Sonuçları	24
4.3. MIRAGE-COVID-CCMA-2022 Veri Seti Sonuçları.....	29
5. TARTIŞMA	35
5.1. MIRAGE-2019 Veri Seti Sonuçlarının Değerlendirilmesi	35
5.2. MIRAGE-COVID-CCMA-2022 Veri Seti Sonuçlarının Değerlendirilmesi.....	41
6. SONUÇ VE ÖNERİLER.....	47
KAYNAKLAR	49
ÖZGEÇMİŞ	52

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. MIRAGE-2019 veri seti akış özellikleri	8
Çizelge 3.2. MIRAGE-2019 veri setinin eğitim (%80) ve test (%20) olarak ikiye bölünmesi.....	9
Çizelge 3.3. MIRAGE-COVID-CCMA-2022 veri setinin eğitim (%80) ve test (%20) olarak ikiye bölünmesi	10
Çizelge 3.4. Tez çalışmasında kullanılan DVM yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler.....	11
Çizelge 3.5. Tez çalışmasında kullanılan k-EYK yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler.....	12
Çizelge 3.6. Tez çalışmasında kullanılan AG yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler.....	13
Çizelge 3.7. Tez çalışmasında kullanılan RO yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler.....	14
Çizelge 3.8. Tez çalışmasında kullanılan ESA mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler	15
Çizelge 3.9. Tez çalışmasında kullanılan UKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler	16
Çizelge 3.10. Tez çalışmasında kullanılan ÇYUKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler.....	17
Çizelge 3.11. Tez çalışmasında kullanılan KTB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler.....	18
Çizelge 3.12. Tez çalışmasında kullanılan özellik seçme yöntemin çalıştırılması için kullanılan fonksiyon ve parametreler.....	19
Çizelge 3.13. Çok sınıflı bir sınıflandırma için örnek bir karışıklık matrisi.....	20
Çizelge 3.14. Çok sınıflı bir sınıflandırma için karışıklık matrisinden üzerinden hesaplanan karşılaştırma parametreleri.....	20
Çizelge 4.1. MIRAGE-2019 veri seti için özellik seçme yöntemleri kullanılarak seçilen özellikler ve çalışma süreleri	24
Çizelge 4.2. MIRAGE-2019 veri seti için DVM sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	25

Çizelge	Sayfa
Çizelge 4.3. MIRAGE-2019 veri seti için k-EYK sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	25
Çizelge 4.4. MIRAGE-2019 veri seti için AG sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	26
Çizelge 4.5. MIRAGE-2019 veri seti için RO sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	26
Çizelge 4.6. MIRAGE-2019 veri seti için ESA sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	27
Çizelge 4.7. MIRAGE-2019 veri seti için UKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	27
Çizelge 4.8. MIRAGE-2019 veri seti için ÇYUKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	28
Çizelge 4.9. MIRAGE-2019 veri seti için KTB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	28
Çizelge 4.10. MIRAGE-COVID-CCMA-2022 veri seti için özellik seçme yöntemleri kullanılarak seçilen özellikler ve çalışma süreleri	29
Çizelge 4.11. MIRAGE-COVID-CCMA-2022 veri seti için DVM sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	30
Çizelge 4.12. MIRAGE-COVID-CCMA-2022 veri seti için k-EYK sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	30
Çizelge 4.13. MIRAGE-COVID-CCMA-2022 veri seti için AG sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	31
Çizelge 4.14. MIRAGE-COVID-CCMA-2022 veri seti için RO sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	31
Çizelge 4.15. MIRAGE-COVID-CCMA-2022 veri seti için ESA sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	32
Çizelge 4.16. MIRAGE-COVID-CCMA-2022 veri seti için UKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	32
Çizelge 4.17. MIRAGE-COVID-CCMA-2022 veri seti için ÇYUKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	33
Çizelge 4.18. MIRAGE-COVID-CCMA-2022 veri seti için KTB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri	33

Çizelge	Sayfa
Çizelge 5.1. MIRAGE-2019 veri seti için elde edilen en iyi sonuçların özeti.....	36
Çizelge 5.2. Tez çalışmasında MIRAGE-2019 veri seti için elde edilen sonuçlar ile geçmiş çalışmalarda elde edilen sonuçların karşılaştırılması	40
Çizelge 5.3. MIRAGE-COVID-CCMA-2022 veri seti için elde edilen en iyi sonuçların özeti	42



ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 3.1. MIRAGE-2019 veri seti JSON dosya ağaç yapısı.....	8
Şekil 3.2. UKSB iç yapısı	16
Şekil 3.3. ÇYUKSB genel çalışma yapısı.....	17
Şekil 3.4. KTB iç yapısı.....	18
Şekil 5.1. Tez çalışmada MIRAGE-2019 veri seti için a) özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, b) özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri	38
Şekil 5.2. Tez çalışmada MIRAGE-2019 veri seti için elde edilen en iyi AİKE grafikleri (yeşil çizgi: özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, kırmızı çizgi: özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak).....	39
Şekil 5.3. Tez çalışmada MIRAGE-COVID-CCMA-2022 veri seti için a) özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, b) özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak, c) özellik seçimi yapıldıktan sonra (RO) ve RO sınıflandırıcı kullanılarak elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri	43
Şekil 5.4. Tez çalışmada MIRAGE-COVID-CCMA-2022 veri seti için elde edilen en iyi AİKE grafikleri (yeşil çizgi: özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, kırmızı çizgi: özellik seçimi yapıldıktan sonra (RO) ve RO sınıflandırıcı kullanılarak)	44

SİMGELER VE KISALTMALAR

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

Simgeler

Açıklamalar

a	Kaydırma Parametresi
b	Polinom Derecesi
c	Hedef Faaliyet
i	Parametreleri Hesaplanan Sınıf
k	k-EYK Komşu Sayısı
m	Toplam Sınıf Sayısı
n	Toplam Eleman Sayısı
n_i	Sınıf-(i) Eleman Sayısı
MI	Bilgi Ölçüm Endeksi
s	Özellik Sayısı
S	Özellik Alt Kümesi
σ	Varyans

Kısaltmalar

Açıklamalar

AG	Adaptif Güçlendirme
AİKE	Alıcı İşlem Karakteristiği Eğrisi
AİKE-AKA	Alıcı İşlem Karakteristiği Eğrisi Altında Kalan Alan
COVID-19	Coronavirus Disease 2019
CPU	Central Processing Unit
ÇYUKSB	Çift Yönlü Uzun Kısa Süreli Bellek
DN	Doğru Negatif
DP	Doğru Pozitif
DVM	Destek Vektör Makinesi
ESA	Evrişimli Sinir Ağları
JSON	JavaScript Object Notation

Kısaltmalar**Açıklamalar****k-EYK**

k-En Yakın Komşu

KTB

Kapılı Tekrarlayan Birim

MFMU

Minimum Fazlalık Maksimum Uygunluk

RO

Rastgele Orman

UKSB

Uzun Kısa Süreli Bellek

YP

Yanlış Pozitif

YN

Yanlış Negatif



1. GİRİŞ

Dünya çapında internet kullanımı her geçen yıl daha da yaygınlaşmaktadır. Bu yaygınlaşma, bireylerin ve ticari şirketlerin günlük operasyonlarına da yansımıştır. Günümüzde internet, günlük ve ticari yaşamın önemli ve ayrılmaz bir parçası haline gelmiştir [1]. Bu durumda 2019 yılının son aylarında ortaya çıkan COVID-19 (Coronavirus Disease 2019) pandemisinin de etkisi vardır. Kemp tarafından hazırlanan “Dijital 2021: Küresel Genel Bakış Raporu” [2] sadece Ocak 2020 ile Ocak 2021 dönemini kapsayan bir yıllık süreçte internet kullanıcı sayısında 317 milyon (%7,3) artış yaşandığını göstermektedir. Aynı rapora göre bu dönemde sosyal medya kullanıcı sayısı 490 milyon (%13,2) ve mobil kullanıcı sayısı 93 milyon (%1,8) artmıştır. “Dijital 2023: Küresel Genel Bakış Raporu” ise 2023 yılı Ocak ayı itibariyle küresel internet kullanıcı sayısının 5,16 milyara, sosyal medya kullanıcı sayısının ise 4,76 milyara yükseldiğini bildirmektedir [3].

Teknoloji alanında yaşanan gelişmelere paralel olarak akıllı telefonlar hayatımıza girmiş ve mobil internet kullanımı hızlı şekilde gelişmiştir. “Ericsson Hareketlilik Raporu-2022” [4] verilerine göre 2022 yılı itibariyle dünya genelinde mobil abonelik sayısının 8,4 milyar olduğu bildirilmektedir. Ayrıca, 2028 yılı itibariyle bu rakamın 9,2 milyara ulaşması beklendiği ifade edilmektedir. 5G internet alt yapısının dünya genelinde yaygınlaşmasıyla aboneler daha hızlı ve kaliteli internet kullanıcıları haline gelmektedir. Söz konusu rapora [4] göre 2021 yılında aylık 68 Eksabayt olan mobil internet kullanımı 2022 yılında aylık 90 Eksabayt seviyesine erişmiştir. Ayrıca, önümüzdeki beş yılın sonunda yani 2028 yılında aylık 324 Eksabayt seviyesine ulaşmasının beklendiği bildirilmektedir [4].

Ağ trafiği sınıflandırması dinamik zamanlama, hizmet kalitesi güvencesi, içerik tabanlı faturalandırma ve anormallik algılama gibi çok çeşitli hizmetler için önemli veriler sunmaktadır [5]. Ağ trafiği sınıflandırması üzerine yirmi yılı aşkın süredir önemli çalışmalar gerçekleştirilmiştir. Ancak, süreç içinde artan paket şifreleme ve mobil veri trafiği geleneksel yaklaşımları işlevsiz bırakmıştır [6]. Derin paket incelemesine veya paket bağlantı noktası analizine dayanan geleneksel yaklaşımlar günümüz ağ trafiğinin hemen hemen %50’inde kullanılamaz hale gelmiştir. Ağ trafiğinin sınıflandırılmasında geleneksel yöntemlerin yerini makine öğrenmesi ve son yıllarda derin öğrenme almaya

başlamıştır. Sınıflandırmada kullanılacak sınıf sayısının ihtiyaca göre belirlenebilmesi bu yöntemleri esnek ve kullanışlı hale getirmektedir.

Ağ trafiği sınıflandırma çalışmalarının kolaylaştırılması için son yıllarda birçok açık erişimli veri seti oluşturulmuş ve araştırmacıların kullanımına sunulmuştur. Ağ trafiği sınıflandırma çalışmalarında kullanılmak üzere yayınlanan başlıca veri setleri ISCX2016 [7], USTC-TFC2016 [8], IMTD17 [9], Unicauca18 [10], MTD18 [11], MIRAGE-2019 [12], UTMobileNetTraffic2021 [13] ve MIRAGE-COVID-CCMA-2022 [14] olarak sıralanabilir. Söz konusu veri setleri genel olarak uygulama katmanı [9] veya taşıma katmanı [7-8, 10-13] üzerinden elde edilmiştir. Ayrıca, ham verinin doğrudan kendisini [7, 8, 11], verinin akış özelliklerini [9, 10, 12, 14] veya paket başlık bilgilerini [13] kapsamaktadırlar. Paket sayısı açısından incelendiğinde ise 2 milyon [8] ile 27,8 milyon [12] arasında veri paketi içerdikleri görülmektedir. Veri setleri genel olarak birçok güncel ve popüler uygulamaya ilişkin veri etiket bilgileri taşımaktadır.

Mobil ağ trafiğini diğer standart ağ trafiğinden ayıran kendine has özellikleri vardır. Mobil ağ trafiğinin genel ağ trafiği içindeki ağırlığının artması mobil ağ trafiği sınıflandırılmasını yeni ve önemli bir çalışma başlığı haline getirmiştir. Ayrıca, süreç içinde doğrudan mobil ağlar üzerine siber güvenlik çalışmaları yapılmaya başlanmıştır. Bu durumda zamanla mobil ağ trafiğini içeren veri setlerinin oluşturulmasının önemli bir katkısı vardır. Bu kapsamda yakın zamanda oluşturulan güncel veri setlerinden bir tanesi MIRAGE-2019 [12] veri setidir. Aceto, Ciuonzo, Montieri, Persico ve Pescapé tarafından hazırlanan veri seti taşıma katmanı üzerinden elde edilen mobil ağ trafiğini içermektedir. Paket ve akış özelliklerinin yer aldığı veri seti 27,8 milyon veri paketi ve 40 adet mobil uygulama etiketi taşımaktadır [12].

COVID-19 pandemisi ile ortaya çıkan sokağa çıkma kısıtlamaları çevrimiçi görüşme yapılmasını sağlayan iletişim ve iş birliği uygulamalarına olan ilgiyi büyük ölçüde artırmıştır. Bu süreçte evden çalışmanın yayınlamasıyla ticari şirketler günlük operasyonlarının çoğunu bu çevrimiçi görüşmeler ile gerçekleştirir hale gelmiştir. Sonuç olarak, iletişim ve iş birliği uygulamalarına ait trafiğinin toplam ağ trafiği içindeki ağırlığı artmıştır. Guarino ve diğerleri tarafından 2022 yılında, popülerliği artan iletişim ve iş birliği uygulamalarına ait mobil ağ trafiğini içeren bir veri seti olan MIRAGE-COVID-CCMA-2022 oluşturulmuş ve araştırmacıların kullanımına sunulmuştur [14]. MIRAGE-

COVID-CCMA-2022 veri seti Discord, GotoMeeting, Meet, Messenger, Skype, Slack, Teams, Webex ve Zoom olmak üzere toplam 9 farklı mobil iletişim ve iş birliği uygulamasına ait trafiğin paket ve akış özelliklerini içermektedir.

Tez çalışmasında, mobil uygulama ağ trafiği sınıflandırmasına ilişkin toplam iki sınıflandırma uygulaması gerçekleştirilmiştir. Birinci sınıflandırma uygulamasında, MIRAGE-2019 veri seti kullanılmış olup veri setinin kullanımda olan ikinci versiyonu içinde yer alan mobil ağ trafiği herhangi bir cihaz veya mobil uygulamaya dışlanmayacak şekilde sınıflandırmaya dahil edilmiştir. Bu kapsamda AccuWeather, Comics, Dropbox, Duolingo, Facebook, FourSquare, Groupon, Messenger, OneFootball, Pinterest, Slitter.io, Spotify, Subito, Trello, TripAdvisor, Twitter, Viber, Waze, Wish ve Youtube olmak üzere toplam 20 adet mobil uygulamaya ait ağ trafiği sınıflandırılmıştır. İkinci sınıflandırma uygulamasında ise MIRAGE-COVID-CCMA-2022 veri seti içinde yer alan mobil iletişim ve iş birliği uygulamalarına ait ağ trafiğinin sınıflandırılması gerçekleştirilmiştir. Birinci sınıflandırma uygulamasına benzer şekilde ikinci sınıflandırma uygulamasında da veri seti içinde yer alan mobil ağ trafiği herhangi bir cihaz veya mobil uygulamaya dışlanmayacak şekilde kullanılmıştır. Bu kapsamda Discord, GotoMeeting, Meet, Messenger, Skype, Slack, Teams, Webex ve Zoom olmak üzere toplam 9 adet mobil uygulamayı kapsayan bir ağ trafiği sınıflandırması gerçekleştirilmiştir.

Tez çalışmasında MIRAGE-2019 ve MIRAGE-COVID-CCMA-2022 veri seti içinde yer alan mobil ağ trafiğinin akış özellikleri üzerinden makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak sınıflandırılması üzerine yoğunlaşmıştır. Makine öğrenmesi sınıflandırıcısı olarak Destek Vektör Makinesi (DVM), k-En Yakın Komşu (k-EYK), Adaptif Güçlendirme (AG) ve Rastgele Orman (RO) yöntemleri kullanılırken derin öğrenme sınıflandırıcısı olarak Evrişimli Sinir Ağları (ESA), Uzun Kısa Süreli Bellek (UKSB), Çift Yönlü Uzun Kısa Süreli Bellek (ÇYUKSB) ve Kapılı Tekrarlayan Birim (KTB) yöntemlerinden yararlanılmıştır. Ayrıca, Minimum Fazlalık Maksimum Uygunluk (MFMU), k-EYK, RO olmak üzere toplam üç adet özellik seçme yöntemi kullanılarak daha anlamlı özelliklerin belirlenmesi ve sınıflandırma başarısının artırılması sağlanmıştır. Bu kapsamda tez çalışmasının literatüre başlıca katkıları aşağıdaki şekilde sıralanabilir.

- Tez çalışmasında mobil uygulamalara ait trafiğin sınıflandırılmasına yönelik deneyler yapılmıştır. Söz konusu mobil uygulamaların önemli bir kısmı yoğun kullanıma sahip

güncel ve popüler uygulamalardır. Tez çalışmasında, mobil uygulamalara ait trafiğin her geçen gün artması nedeniyle karşılaştırmalı sonuçlara dayanılarak yeni bir sınıflandırma modeli geliştirilmiş ve büyük veri ortamında sonuçları doğrulanmıştır.

- Deneylerde sekiz farklı akış özellik grubu ile dört farklı makine öğrenmesi ve dört farklı derin öğrenme sınıflandırıcısı kullanılmıştır. Bu sayede hem akış özellik gruplarındaki hem de makine öğrenmesi ve derin öğrenme yöntemlerindeki farklılaşmanın sınıflandırma sonuçları üzerine etkileri kapsamlı bir şekilde analiz edilmiştir.
- Üç farklı özellik seçme yöntemi kullanılarak seçilen akış özellikleri yine aynı sınıflandırıcılar kullanılarak sınıflandırılmıştır. Bu sayede hem özellik seçme yöntemlerinin kendi içinde başarıları karşılaştırılmış hem de daha anlamlı özelliklerin sınıflandırma için kullanılmasının sonuçlara etkisi ortaya çıkarılmıştır.
- Makine öğrenmesi, derin öğrenme ve özellik seçme yöntemlerinin sınıflandırma başarılarının yanı sıra çalışma süreleri açısından karşılaştırması gerçekleştirilmiştir.
- Söz konusu deneyler iki farklı veri seti için ayrı ayrı gerçekleştirilmiş ve mobil uygulama kümesinde değişiklik olmasının sınıflandırma sonuçlarına etkileri gözlemlenmiştir.
- Kapsamlı deneyler sonucunda en yüksek sınıflandırma başarısını sağlayan ve en hızlı şekilde çalışan sınıflandırma modeli ortaya çıkarılmıştır.

Tez çalışması toplam altı ana bölümden oluşmaktadır. Tez çalışmasının ikinci bölümünde aynı veri setlerinin kullanıldığı geçmiş çalışmalara ilişkin bilgilere yer verilmiştir. Söz konusu literatür çalışmalarında kullanılan yöntemler ve elde edilen sonuçlar detaylı şekilde açıklanmıştır. Tez çalışmasının üçüncü bölümünde kullanılan veri setleri, makine öğrenmesi, derin öğrenme ve özellik seçme yöntemleri detaylı şekilde ele alınmıştır. Bu bölümde ayrıca, karşılaştırma kriterlerinin hesaplanma yöntemleri üzerinde durulmuştur. Dördüncü bölümde ise deneylerin gerçekleştirilme prosedürleri açıklanmış ve deneyler neticesinde elde edilen sınıflandırma sonuçlarına yer verilmiştir. Tez çalışmasının beşinci bölümünde, deneyler sonucunda elde edilen sonuçlar özetlenmiş, birbirleriyle karşılaştırılmış ve tartışılmıştır. Son bölüm olan altıncı bölümde ise tez çalışmasında elde edilen temel çıkarımlara yer verilmiş, önerilerde bulunulmuş ve gelecek çalışmalara ilişkin bir perspektif ortaya konulmuştur.

2. LİTERATÜR ÇALIŞMALARI

2021 yılında Aceto ve diğerleri [15] tarafından mobil ağ trafiğinin tahmini ve karakterize edilmesi için Markov zincirleri ve gizli Markov modelleri kullanılması önerilmiştir. MIRAGE-2019 veri seti kullanılarak gerçekleştirilen çalışma sonuçları önerilen modelin mobil ağ trafiğinin tahmini ve karakterize edilmesi için k-EYK regresyonu, doğrusal regresyon ve RO regresyonu yöntemlerine göre daha başarılı olduğunu göstermektedir. Aynı yıl Montieri ve diğerleri [16] tarafından mobil ağ trafiğinin paket bazında tahmin edilmesi için Markov modeli, makine öğrenmesi ve derin öğrenme yöntemlerini içeren bir çalışma gerçekleştirilmiştir. Çalışmada MIRAGE-2019 veri setinde yer alan bazı uygulamalara ait ağ trafiğinin yanı sıra MIRAGE-VIDEO veri setinde yer alan uygulamalara ait mobil ağ trafiği kullanılmıştır. Varış süresi, paket yönü ve yük uzunluğu parametreleri kullanılarak tahmin yapılan çalışmada karşılaştırma parametresi olarak G-Ortalama ve ortalama karesel hata kriterlerinden yararlanılmıştır. Çalışma sonucunda derin öğrenmenin ağ trafiği tahmini için diğer yöntemlerden daha başarılı sonuçlar elde edilmesini sağladığı görülmüştür.

2021 yılında Nascita ve diğerleri [17] tarafından açıklanabilir yapay zeka yöntemleri kullanılarak mobil ağ trafiği sınıflandırması gerçekleştirilmiştir. MIRAGE-2019 veri seti içinde yer alan 41 adet mobil uygulamaya ait ağ trafiği üzerinden gerçekleştirilen sınıflandırma çalışmasında, deneylerin kolaylaştırılması için veri seti içindeki sadece Xiaomi Mi5 cihaz ile toplanan veriler ile eğitim ve test işlemleri gerçekleştirilmiştir. Çalışma sonuçları, her mobil ağ sınıfı için elde edilen sonuçlarının basit ortalaması alınarak hesaplanan ortalama doğruluk, F1-Skoru ve G-Ortalama parametreleri kullanılarak karşılaştırılmıştır. On kat çapraz doğrulama kullanılarak gerçekleştirilen deneyler sonucunda en yüksek ortalama doğruluk 0,9214 olarak elde edilirken ortalama F1-Skoru 0,9147 olarak hesaplanmıştır. Ayrıca, G-Ortalama parametrelerinin ortalaması 0,9536 olarak elde edilmiştir.

2022 yılında Hussey, Stone ve Camp [18] tarafından gerçekleştirilen mobil uygulama ağ trafiğinin sınıflandırma çalışmasında MIRAGE-2019 veri seti ve bu veri seti ile aynı prosedürler kullanılarak elde edilen ikinci bir özel veri seti kullanılmıştır. Veri seti içinde yer alan paket uzunluklarına ait toplam 51 adet çift yönlü akış, yukarı yönlü akış ve aşağı

yönlü akış özelliğe ilave olarak paket sayısı, IP paketlerindeki toplam bayt, L4 yüklerinin toplam baytı, saniye cinsinden akış süresi ve çift akış etiketi olmak üzere toplam 13 adet ilave özellik kullanılmıştır. Yani, sınıflandırma deneyleri toplam 64 özellik kullanılarak gerçekleştirilmiştir. Deneylerde, herhangi bir mobil uygulama için örnek sayısı 1 000'in altında olması durumunda yeterince temsil edilmediği için sınıflandırma işlemlerine dahil edilmemiştir. Sonuç olarak çalışmada 18 adet mobil uygulama için ağ trafiği sınıflandırması yapılmış olup çalışmanın deneyleri %70 eğitim ve %30 test ayrışımına göre gerçekleştirilmiştir. Aşırı gradyan güçlendirme sınıflandırıcı kullanılarak gerçekleştirilen çalışma sonucunda ortalama doğruluk 0,8013; ortalama F1-Skoru 0,7791 ve ortalama Alıcı İşlem Karakteristiği Eğrisi (AİKE) Altında Kalan Alan (AİKE-AKA) 0,9787 olarak hesaplanmıştır.

MIRAGE-COVID-CCMA-2022 veri seti yakın zamanda yayınlandığı için literatürde hali hazırda bu veri setini kullanan sınıflandırma çalışma sayısı oldukça kısıtlıdır. Veri setini oluşturan çalışma ekibi (Guarino ve diğerleri) aynı zamanda ilk sınıflandırma uygulamasını [14] da gerçekleştirmiştir. Bu kapsamda bir boyutlu ESA, hibrit ve taklitçi-geliştirilmiş olmak üzere üç farklı yöntem kullanılarak mobil ağ trafiği ve etkinlik sınıflandırması yapılmıştır. Bu çalışmada veri paketlerinin doğrudan kendisinden yararlanılmış olup çalışma sonucunda mobil ağ trafiği sınıflandırması için 0,94 ile 0,99 arasında değişen doğruluk ve 0,95 ile 0,99 arasında F1-Skoru sonucu elde edilmiştir.

3. MATERİYAL VE METOT

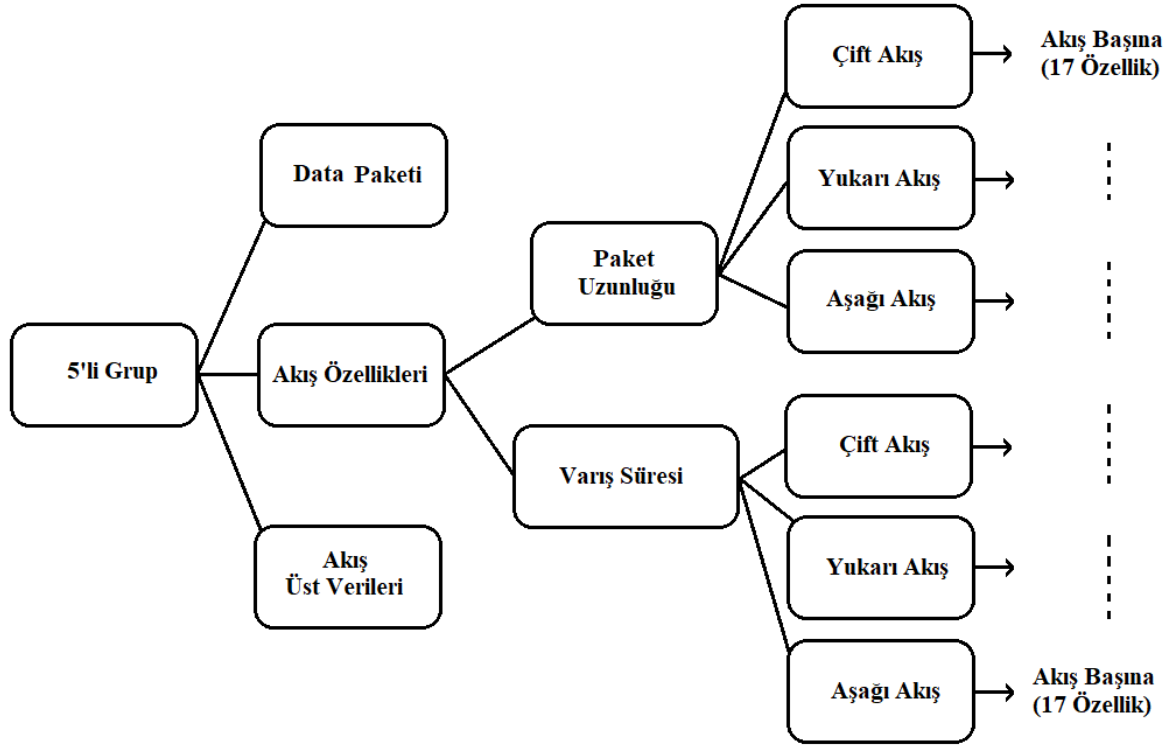
Tez çalışmasının bu bölümünde deneylerde kullanılan veri setleri, makine öğrenmesi, derin öğrenme ve özellik seçme yöntemleri ele alınmıştır. Ayrıca, tez çalışmasında kullanılan karşılaştırma kriterlerinin hesaplanma yöntemleri açıklanmıştır.

3.1. Veri Setleri

Tez çalışmasında gerçekleştirilen deneylerde MIRAGE-2019 ve MIRAGE-COVID-CCMA-2022 olmak üzere iki farklı veri seti kullanılmıştır.

3.1.1. MIRAGE-2019 veri seti

Tez çalışmasının birinci sınıflandırma uygulamasında Aceto, Ciuonzo, Montieri, Persico ve Pescapé [12] tarafından 2019 yılında oluşturulan MIRAGE-2019 veri seti içinde yer alan mobil ağ trafiği kullanılmıştır. Söz konusu veri setinin tez çalışmasında kullanılan son versiyonuna (ikinci versiyonu) [19] üzerinden erişilmesi mümkündür. Bu veri seti AccuWeather, Comics, Dropbox, Duolingo, Facebook, FourSquare, Groupon, Messenger, OneFootball, Pinterest, Slitter.io, Spotify, Subito, Trello, TripAdvisor, Twitter, Viber, Waze, Wish ve Youtube olmak üzere toplam 20 adet mobil uygulamaya ait ağ trafiği verisi içermektedir. Söz konusu ağ trafiği Xiaomi Mi5 ve Google Nexus 7 akıllı cihazlarından toplanmıştır. Veri seti JSON (JavaScript Object Notation) uzantılı dosyalar halinde arşivlenmiştir. Her bir JSON dosyası içinde her bit akış paketine ait data paketleri, akış özellikleri ve akış üst verileri olmak üzere toplam üç ana klasör bulunmaktadır. Akış özellik dosyaları ise kendi içinde paket uzunluğu ve varış süresi dosyalarını içermektedir. Bu dosyalar ise çift yönlü akış, yukarı yönlü akış ve aşağı yönlü akış olmak üzere üç özellik vektörünü bünyesinde barındırmaktadır. Her özellik vektörü toplam 17 adet özellikten oluşmaktadır. Söz konusu JSON dosyalarının bahsedilen bu yapısı Şekil 3.1’de özetlenmiştir. Çizelge 3.1 ise özellik vektörlerinin içeriğini açıklamaktadır.



Şekil 3.1. MIRAGE-2019 veri seti JSON dosya ağaç yapısı [12]

Çizelge 3.1. MIRAGE-2019 veri seti akış özellikleri [12]

No	Özellikler	Açıklama
1	min	Minimum
2	max	Maksimum
3	mean	Aritmetik Ortalama
4	std	Standart Sapma
5	var	Varyans
6	mad	Ortalama Mutlak Sapma
7	skew	Çarpıklık
8	kurtosis	Basıklık
9-17	q_percentile	q. Yüzdelik ($q \in [10 : 10 : 90]$)

Tez çalışmasında gerçekleştirilen sınıflandırma deneyleri veri setinin eğitim kümesi ve test kümesi olmak üzere ikiye ayrıştırılmasıyla gerçekleştirilmiştir. Veri setinde yer alan her bir mobil ağ trafiğinin %80'i eğitimlerde kullanılmıştır. Eğitim kümesi haricinde geriye kalan %20'lik kısımlar kullanılarak test işlemleri gerçekleştirilmiştir. Eğitim ve test kümesinin mobil uygulama bazında ayrışımına ilişkin bilgiler Çizelge 3.2'de özetlenmiştir.

Çizelge 3.2. MIRAGE-2019 veri setinin eğitim (%80) ve test (%20) olarak ikiye bölünmesi

Uygulama	Eğitim	Test	Toplam
AccuWeather	17 000	4 250	21 250
Comics	9 098	2 275	11 373
Dropbox	9 099	2 275	11 374
Duolingo	10 420	2 605	13 025
Facebook	30 025	7 507	37 532
FourSquare	13 039	3 260	16 299
Groupon	4 873	1 219	6 092
Messenger	6 124	1 531	7 655
OneFootball	18 572	4 643	23 215
Pinterest	11 489	2 873	14 362
Slitter.io	5 022	1 256	6 278
Spotify	12 208	3 052	15 260
Subito	12 486	3 122	15 608
Trello	5 152	1 289	6 441
TripAdvisor	6 346	1 587	7 933
Twitter	17 551	4 388	21 939
Viber	7 004	1 752	8 756
Waze	28 560	7 140	35 700
Wish	19 099	4 775	23 874
Youtube	11 012	2 753	13 765
Toplam	254 179	63 552	317 731

3.1.2. MIRAGE-COVID-CCMA-2022 veri seti

Tez çalışmasının ikinci sınıflandırma uygulamasında Guarino ve diğerleri [14] tarafından 2022 yılının Ekim ayında yayınlanan açık erişimli MIRAGE-COVID-CCMA-2022 veri seti içinde yer alan mobil ağ trafiği kullanılmıştır. Söz konusu veri setinin tez çalışmasında kullanılan son versiyonuna [20] üzerinden erişilmesi mümkündür. Veri seti Discord, GotoMeeting, Meet, Messenger, Skype, Slack, Teams, Webex ve Zoom olmak üzere toplam 9 adet mobil iletişim ve iş birliği uygulamasına ait ağ trafik verisini içermektedir. Veri seti Nisan 2021 ile Aralık 2021 tarihleri arasında Napoli Üniversitesinde toplanmıştır. Söz konusu ağ trafiği bir adet Google Nexus 6 ve iki adet Samsung Galaxy A5 mobil cihaz üzerinden elde edilmiştir. Ağ trafiğinin toplanması için kullanıcılar tarafından 15 ile 80 dakika arasında değişen etkinlikler düzenlenmiştir. Bu etkinlikler sohbet, sesli arama ve görüntülü arama aktivitelerini içermektedir. Veri seti JSON uzantılı dosyalar halinde arşivlenmiştir. MIRAGE-COVID-CCMA-2022 veri setinin dosya ağaç yapısı MIRAGE-

2019 veri seti ile bire bir aynıdır. Söz konusu ağaç yapısı ve özellik vektörlerinin açıklaması Şekil 3.1 ve Çizelge 3.1 üzerinden incelenebilir.

Tez çalışmasının deneyleri veri setinin eğitim kümesi ve test kümesi olmak üzere ikiye ayrıştırılmasıyla gerçekleştirilmiştir. Veri setinde yer alan her bir mobil ağ trafiğinin %80'i eğitimlerde kullanılmıştır. Eğitim kümesi haricinde geriye kalan %20'lik kısımlar kullanılarak test işlemleri gerçekleştirilmiştir. Eğitim ve test kümesinin mobil uygulama bazındaki eleman sayısına ilişkin bilgiler Çizelge 3.3'de özetlenmiştir.

Çizelge 3.3. MIRAGE-COVID-CCMA-2022 veri setinin eğitim (%80) ve test (%20) olarak ikiye bölünmesi

Uygulama	Eğitim	Test	Toplam
Discord	14 542	3 636	18 178
GotoMeeting	7 497	1 875	9 372
Meet	7 710	1 928	9 638
Messenger	12 632	3 158	15 790
Skype	106 099	26 525	132 624
Slack	5 081	1 271	6 352
Teams	29 006	7 252	36 258
Webex	10 050	2 513	12 563
Zoom	18 275	4 569	22 844
Toplam	210 892	52 727	263 619

3.2. Makine Öğrenmesi Sınıflandırıcılar

Tez çalışmasında gerçekleştirilen sınıflandırma uygulamalarında Destek Vektör Makinesi (DVM), k-En Yakın Komşu (k-EYK), Adaptif Güçlendirme (AG) ve Rastgele Orman (RO) olmak üzere dört farklı makine öğrenmesi yönteminden yararlanılmıştır.

3.2.1. Destek vektör makinesi

Destek Vektör Makinesi (DVM) fikrinin ilk temelleri AT&T Bell laboratuvarlarında atılmıştır. Günümüzde kullanılan halini alması Cortes ve Vapnik tarafından 1995 yılına kadar gerçekleştirilen bir dizi çalışma sonucunda olmuştur [21]. DVM, kısa zamanda popülerlik kazanmış ve yaygın olarak kullanılmaya başlayan bir makine öğrenmesi yöntemi haline gelmiştir. Özellikle, sınıflandırma ve regresyon işlemi içeren uygulamalarda çalıştırılmasının kolay olması ve yüksek performans göstermesi sebebiyle

tercih edilmiştir. DVM genel olarak ilişkili öğrenme algoritmasını esas alan bir denetimli öğrenme yöntemidir [21].

DVM, sınıflandırma yaparken modele giriş olarak sokulan verileri doğrudan değil dolaylı olarak kullanmaktadır. Bu kapsamda, giriş verilerinden elde edilen özellikler bir matris haline getirilmekte ve bu şekilde işlemlere devam edilmektedir. DVM eğitiminde giriş matrislerine karşılık gelen çıkış etiketleri arasındaki uzaklıklar olabildiğince yüksek olacak şekilde karar sınırları belirlenmektedir. Bu sınırlar doğrusal olabileceği gibi kullanılan çekirdek fonksiyonuna göre doğrusal olmayan şekle de sahip olabilir. Sınıflandırma uygulamalarında homojen polinomik, homojen olmayan polinomik veya radyal fonksiyonlar en sık tercih edilen çekirdek fonksiyonlarıdır.

Eşitlik 3.1, Eşitlik 3.2 ve Eşitlik 3.3’de sırasıyla dot, polinom ve radyal temelli çekirdek fonksiyonları matematiksel olarak ifade edilmiştir. Eşitliklerde yer alan çarpma işlemleri noktasaldır. Ayrıca, a kaydırma parametresini, b polinomun derecesini ve σ varyansı temsil etmektedir.

$$K(x, \bar{x}) = x \cdot \bar{x} \quad (3.1)$$

$$K(x, \bar{x}) = (x \cdot \bar{x} + a)^b \quad (3.2)$$

$$K(x, x') = \exp \left(-\frac{(\|x - x'\|)^2}{2\sigma^2} \right) \quad (3.3)$$

Tez çalışmasında gerçekleştirilen sınıflandırma uygulamalarına ait programsal kodlamalar MATLAB 2023(a) yazılım platformu kullanılarak gerçekleştirilmiştir. Bu kapsamda DVM yönteminin çalıştırılması için kullanılan fonksiyonlar ve parametreler Çizelge 3.4’de yer almaktadır.

Çizelge 3.4. Tez çalışmasında kullanılan DVM yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler

Sınıflandırıcı	Fonksiyonlar	Fonksiyon Parametreleri
DVM	<i>templateSVM</i> <i>fitcecoc</i>	Çekirdek Fonksiyonu: Radyal Temel Fonksiyonu Standartlaştırma: "Var" Çekirdek Ölçeği: "Otomatik" Diğer Parametreler: "Varsayılan"

3.2.2. k-En yakın komşu

k-En Yakın Komşu (k-EYK) kökleri 1950'li yıllara kadar uzanan temel olarak nitelendirilebilecek bir makine öğrenmesi yöntemidir. İlk olarak 1951 yılında Fix ve Hodges [22] tarafından fikri temelleri inşa edilen k-EYK'nin günümüzde kullanılan son halini alması 1992 yılında Altman [23] tarafından gerçekleştirilen çalışmaların sonucunda olmuştur. Hızlı çalışması ve yüksek performans göstermesi regresyon ve sınıflandırma çalışmalarında yaygın olarak kullanılmaya başlanmasını sağlamıştır.

k-EYK yönteminde giriş vektörlerinin uzunluğu kadar boyutu olan bir uzay tanımlanmaktadır. Yani, giriş olarak kullanılan her bir özellik vektörü bu uzayda bir nokta ile temsil edilmektedir. k-EYK ile sınıflandırma yapılırken öncelikle kendisi hakkında karar verilecek örneğe en yakın komşular tespit edilmektedir. Bu aşamada dikkate alınacak komşu sayısı genel olarak k şeklinde nitelendirilmektedir. Daha sonra söz konusu komşuların gerçekteki etiketlerinin ağırlığına göre örnek hakkında sınıflandırma sonucu oluşturulmaktadır. Bu aşamada komşu sayısının çok yüksek seçilmesi işlem yükünü artırırken çok düşük seçilmesi ise sınıflandırma başarısını olumsuz etkilemektedir. Tez çalışmasında kullanılan k-EYK yönteminin çalıştırılması için kullanılan fonksiyonlar ve parametreler Çizelge 3.5'de yer almaktadır.

Çizelge 3.5. Tez çalışmasında kullanılan k-EYK yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler

Sınıflandırıcı	Fonksiyonlar	Fonksiyon Parametreleri
k-EYK	<i>fitcknn</i>	Komşu Sayısı: "21" Standartlaştırma: "Var" Diğer Parametreler: "Varsayılan"

3.2.3. Adaptif güçlendirme

Adaptif Güçlendirme (AG), karar ağacı temelli bir yaklaşım olarak Freund ve Schapire [24] tarafından önerilmiştir. Karar ağacı yapı olarak fiziki bir ağaca benzetilebilir. Genel olarak kök ve düğümler karar ölçütlerini, yapraklar karar durumlarını temsil etmektedir. Ağacın dalları ise kök, düğümler ve yaprakları birbirine bağlayan yapı bileşenleridir [24]. 1997 yılında ortaya atılan AG, daha yüksek performanslı bir sınıflandırıcı oluşturmak için birçok zayıf performanslı sınıflandırıcının birleştirilmesini esas almaktadır. Bu aşamada

kullanılan zayıf performanslı sınıflandırıcılar tek parça karar ağaçlarıdır. Bu sebeple AG'nin inatçı olarak nitelendirilebilecek sınıflandırılması zor örneklerin sınıflandırılmasının kolaylaştırılmasına odaklandığı söylenilebilir. Denetimli öğrenmenin güçlendirilmesi olarak nitelendirilebilecek bu yaklaşım sayesinde inatçı örnekler daha kolay sınıflandırılan örneklere dönüştürülmektedir. Yani, eğitim işlemlerinde her iterasyon sonrasında eğitim kümesi daha kolay sınıflandırılabilir örnekler ile yeniden inşa edilmektedir. Tez çalışmasında kullanılan AG yönteminin çalıştırılması için kullanılan fonksiyonlar ve parametreler Çizelge 3.6'de yer almaktadır.

Çizelge 3.6. Tez çalışmasında kullanılan AG yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler

Sınıflandırıcı	Fonksiyonlar	Fonksiyon Parametreleri
AG	<i>fitcensemble</i>	Tüm Parametreler: "Varsayılan"

3.2.4. Rastgele orman

Rastgele Orman (RO), AG sınıflandırıcıya benzer şekilde karar ağacı yapılarını temel alan bir makine öğrenmesi yöntemi olarak 2000'li yılların başında ortaya atılmıştır. Breiman tarafından [25] tanımlanan bu yöntem, temel olarak karar ağaçlarının oluşturulması esnasında optimum bir ağaç yapısının elde edilememesi sorununun çözümüne odaklanmaktadır. RO, bu sorunun çözümü ve sınıflandırma sonuçlarına etkisinin azaltılması için farklı yapılara sahip karar ağaçlarını birleştirmektedir. Bu aşamada Breiman tarafından [26] 1994 yılında tanımlanan Bagging yöntemini ve 1998 yılında Ho tarafından [27] tanımlanan Rastgele Alt Uzay yöntemini kullanmaktadır. Bagging yöntemi karar ağacının inşa edilmesinde rastgele vektörler kullanılmasını temel alırken Rastgele Alt Uzay karar ağacı dallanmalarının optimize edilmesi üzerine yoğunlaşmaktadır. Sonuç olarak söz konusu bu yöntemler kullanılarak oluşturulan karar ağaçlarının her biri kendi sınıflandırma sonucu üretmekte ve bütün sonuçların birlikte ele alınmasıyla RO sınıflandırıcısının nihai sınıflandırma sonucu oluşturulmaktadır [25]. RO yönteminin başarısının arkasındaki temel motivasyon birçok sınıflandırma sonucunun aynı anda dikkate alınmasıdır. Tez çalışmasında kullanılan RO yönteminin çalıştırılması için kullanılan fonksiyonlar ve parametreler Çizelge 3.7'de yer almaktadır.

Çizelge 3.7. Tez çalışmasında kullanılan RO yönteminin çalıştırılması için kullanılan fonksiyon ve parametreler

Sınıflandırıcı	Fonksiyonlar	Fonksiyon Parametreleri
RO	<i>fitcensemble</i>	Tüm Parametreler: "Varsayılan"

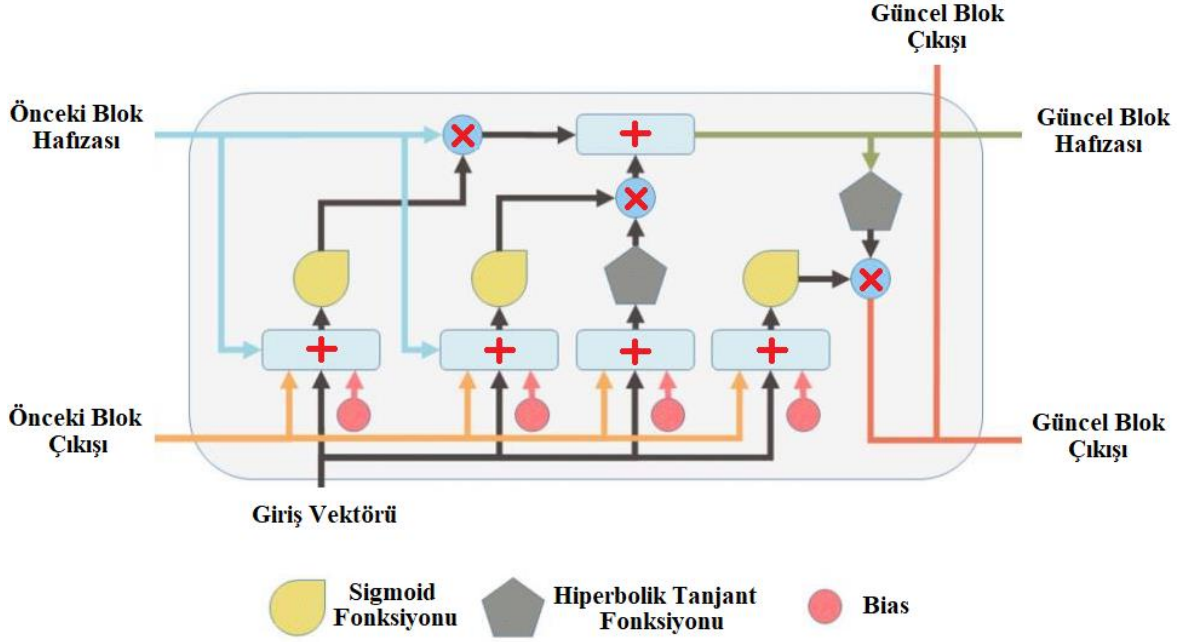
3.3. Derin Öğrenme Sınıflandırıcılar

Tez çalışmasında gerçekleştirilen sınıflandırma uygulamalarında Evrişimli Sinir Ağları (ESA), Uzun Kısa Süreli Bellek (UKSB), Çift Yönlü Uzun Kısa Süreli Bellek (ÇYUKSB) ve Kapılı Tekrarlayan Birim (KTB) olmak üzere dört farklı derin öğrenme yönteminden yararlanılmıştır.

3.3.1. Evrişimli sinir ağları

Evrişimli Sinir Ağları (ESA) 1987 yılında Atlas, Homma ve Marks tarafından [28] ortaya atılmış bir derin öğrenme yöntemidir. ESA'nın temellerinin daha eskiye dayanmasına karşın popülerlik kazanması 2010'lu yılların ikinci yarında olmuştur. Bunun temel sebebi ilk ortaya atıldığında bilgisayar kapasitelerinin yöntemin çalıştırılması için yetersiz olmasıdır [29]. Son yıllarda ESA, bir ve çok boyutlu verilerin sınıflandırılması ve regresyonunda sıklıkla kullanılan bir yöntem haline gelmiştir. ESA gibi derin öğrenme yöntemlerinde kullanılacak mimarilerinin uygulama ihtiyacına göre tasarlanabilmesi makine öğrenmesi yöntemlerine göre önemli bir avantaj sağlamaktadır. ESA mimarileri genel olarak evrişim, aktivasyon fonksiyonu, havuzlama, normalizasyon, düzleştirme, unutturma ve tam bağlı katmanlardan meydana gelmektedir. Evrişim katmanında giriş verileri evrişime tabi tutulduktan sonra aktivasyon fonksiyonlarından geçirilmektedir. Elde edilen çıkışlar havuzlama katmanına giriş olarak verilmekte ve boyutlarının azaltılması sağlanmaktadır. Bu aşamada istenirse ara işlem olarak normalizasyon yapılabilmektedir. Giriş verisinin iki veya çok boyutlu olması durumunda düzleştirme yapılarak özellik matrisleri özellik vektörlerine dönüştürülmektedir. Tam bağlı katman ise özellik vektörlerinin sınıflandırıldığı bölümdür. Bu aşamada kullanılan veri miktarının düşük olmasına bağlı olarak aşırı uyum sorununun önüne geçmek amacıyla unutturma işlemi yapılabilmektedir. Tam bağlı katman olarak genellikle yapay sinir ağları kullanılmakla birlikte DVM veya k-EYK gibi bir makine öğrenmesi sınıflandırıcısından da

kullanılan UKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler ise Çizelge 3.9’da yer almaktadır.



Şekil 3.2. UKSB iç yapısı [32]

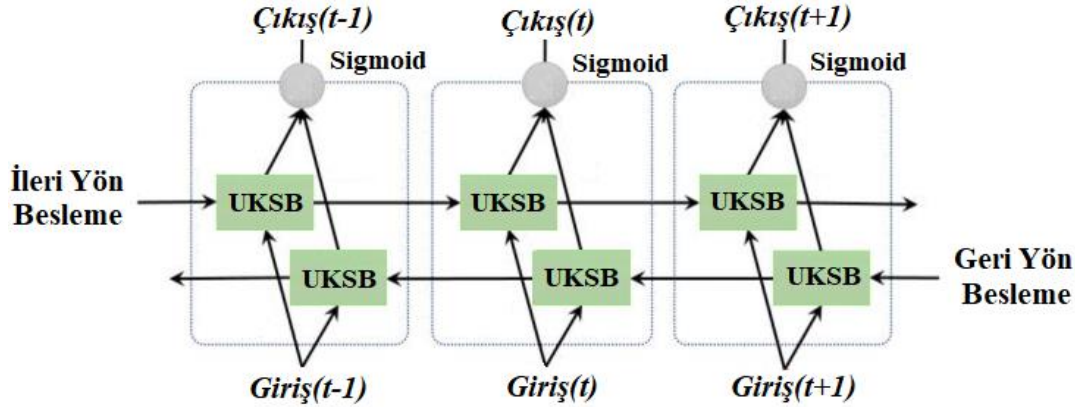
Çizelge 3.9. Tez çalışmasında kullanılan UKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler

Sınıflandırıcı	Katman Yapısı	Fonksiyon Parametreleri	Eğitim Hiperparametreleri
UKSB	<i>sequenceInputLayer</i> <i>lstmLayer</i> <i>fullyConnectedLayer</i> <i>softmaxLayer</i> <i>classificationLayer</i>	(1) (100,'OutputMode','last') (9 ve 20) Default Default	Eğitim Ağı Çözücüsü: Momentumlu Stokastik Gradyan Maksimum İterasyon Sayısı: "50" Başlangıç Öğrenme Oranı: "0,01" Diğer Parametreler: "Varsayılan"

3.3.3. Çift yönlü uzun kısa süreli bellek

Çift Yönlü Uzun Kısa Süreli Bellek (ÇYUKSB), tekrarlayan sinir ağlarına dayanan bir derin öğrenme yöntemidir. UKSB’nin birçok avantajı olmakla birlikte önemli dezavantajlarından bir tanesi sadece mevcut giriş ile gelecek girişleri karşılaştırmasıdır. Yani, tek ve ileri yönlü bir hareket halindedir. Ancak, metin analizi gibi birçok uygulamada önceki ve sonraki verilerin bir bütün halinde dikkate alınması zorunluluğu vardır [33]. Bu dezavantajın ortadan kaldırılması için bir tanesi ileri yönlü bir tanesi ise

geri yönlü olmak üzere iki adet UKSB yapısının birlikte kullanılması önerilmiştir. Şekil 3.3'de ÇYUKSB'nin genel çalışma yapısı görülmektedir. Tez çalışmasında kullanılan ÇYUKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler ise Çizelge 3.10'da yer almaktadır.



Şekil 3.3. ÇYUKSB genel çalışma yapısı [32]

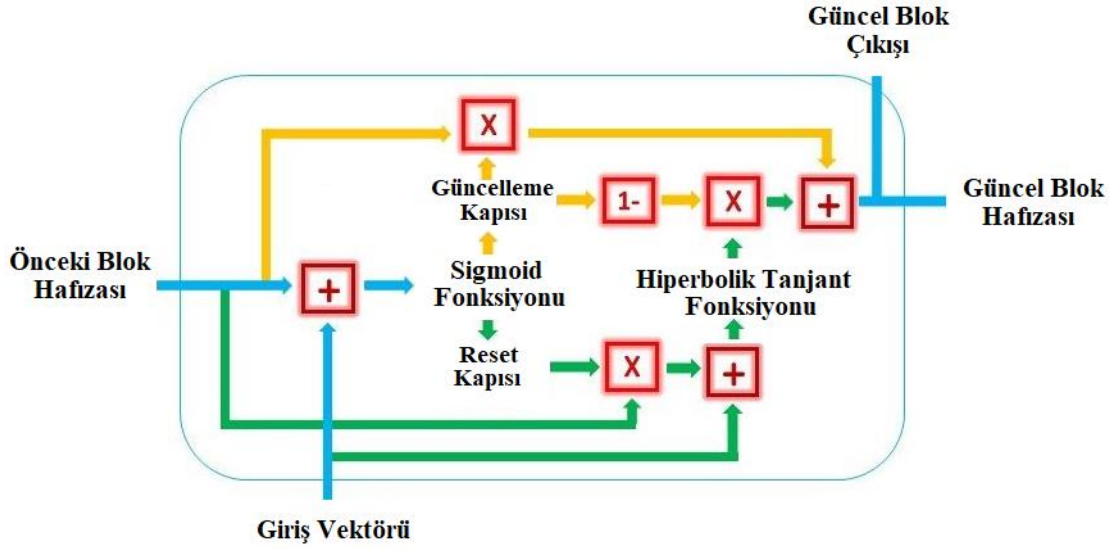
Çizelge 3.10. Tez çalışmasında kullanılan ÇYUKSB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler

Sınıflandırıcı	Katman Yapısı	Fonksiyon Parametreleri	Eğitim Hiperparametreleri
ÇYUKSB	<i>sequenceInputLayer</i> <i>biLstmLayer</i> <i>fullyConnectedLayer</i> <i>softmaxLayer</i> <i>classificationLayer</i>	(1) (100,'OutputMode','last') (9 ve 20) Default Default	Eğitim Ağı Çözücüsü: Momentumlu Stokastik Gradyan Maksimum İterasyon Sayısı: "50" Başlangıç Öğrenme Oranı: "0,01" Diğer Parametreler: "Varsayılan"

3.3.4. Kapılı tekrarlayan birim

Kapılı Tekrarlayan Birim (KTB) UKSB'ye benzer şekilde tekrarlayan sinir ağlarını temel olarak kullanmaktadır. UKSB'de olduğu gibi tekrarlayan sinir ağlarındaki ağ mimarisinin uzunluğuna bağlı olarak katlanarak artan gradyan sebebiyle ağ eğitiminde yaşanan bozulmaya çözüm bulmak amacıyla önerilmiştir. Bu aşamada UKSB'den farklı olarak unutma kapısı yerine unutma anahtarı kullanılmaktadır [34]. UKSB'ye göre kullanılan kapı sayısının azaltılması hızını ve performansını artırmaktadır. Ayrıca, aynı mimariye sahip bir UKSB ile kıyaslandığında KTB daha hızlı eğitilebilmektedir. Şekil 3.4'de KTB'nin iç yapısı görülmektedir. Tez çalışmasında kullanılan KTB mimarisi, fonksiyon parametreleri

ve eğitimlerinde kullanılan hiperparametreler ise Çizelge 3.11’de yer almaktadır.



Şekil 3.4. KTB iç yapısı [35]

Çizelge 3.11. Tez çalışmasında kullanılan KTB mimarisi, fonksiyon parametreleri ve eğitimlerinde kullanılan hiperparametreler

Sınıflandırıcı	Katman Yapısı	Fonksiyon Parametreleri	Eğitim Hiperparametreleri
KTB	<i>sequenceInputLayer</i> <i>gruLayer</i> <i>fullyConnectedLayer</i> <i>softmaxLayer</i> <i>classificationLayer</i>	(1) (100,'OutputMode','last') (9 ve 20) Default Default	Eğitim Ağı Çözücüsü: Momentumlu Stokastik Gradyan Maksimum İterasyon Sayısı: "50" Başlangıç Öğrenme Oranı: "0,01" Diğer Parametreler: "Varsayılan"

3.4. Özellik Seçme Yöntemleri

Tez çalışmasında Minimum Fazlalık Maksimum Uygunluk (MFMU), k-EYK ve RO olmak üzere toplam üç farklı özellik seçme yöntemi kullanılmıştır.

Geleneksel özellik seçme yöntemlerinin birçoğu özellikler ve sınıflandırma etiketleri arasındaki ilişkiyi dikkate almaktadırlar. Ancak, özelliklerin kendi içindeki etkileşim de özellikler ile sınıf etiketleri arasındaki etkileşim kadar önemlidir. MFMU, geleneksel özellik seçme yöntemlerinden farklı olarak bu olguların her ikisi üzerine aynı anda yoğunlaşmaktadır [36]. *MI*, herhangi iki özellik arasındaki bilgi miktarını ölçen bir endeks olarak kabul edilirse özellikler arasındaki minimum fazlalık ve maksimum uygunluğun

elde edilmesi için sırasıyla Eşitlik 3.4 ve Eşitlik 3.5'in sağlanması hedeflenir [37].

$$\min R(S), R = \frac{1}{|S|^2} \sum_{xi, xj \in S} I(xi; xj) \quad (3.4)$$

$$\max D(S, c), D = \frac{1}{|S|} \sum_{xi \in S} I(xi; c) \quad (3.5)$$

Eşitlik 3.4'de $|S|$ S özellik alt kümesindeki özelliklerin sayısı ve $I(xi; xj)$ ise i ve j özellikleri arasındaki MI değeridir. Eşitlik 3.5'de c hedef faaliyet, $I(xi; c)$ ise i özelliği ve hedef faaliyet arasındaki MI değeridir.

k-EYK ve RO sınıflandırmanın yanı sıra özellik seçme içinde kullanılabilir. Bu aşamada daha önce ifade edilen çalışma yapıları özellik kümesi ile sınıflandırma etiketleri arasındaki maksimum bağlantıyı sağlayacak özellik alt kümelerinin belirlenmesi için kullanılır. Tez çalışmasında kullanılan özellik seçme yöntemlerinin çalıştırılması için kullanılan fonksiyonlar ve parametreler Çizelge 3.12'de yer almaktadır.

Çizelge 3.12. Tez çalışmasında kullanılan özellik seçme yöntemin çalıştırılması için kullanılan fonksiyon ve parametreler

Özellik Seçme Yöntemi	Fonksiyonlar	Fonksiyon Parametreleri
MFMU	<i>fscmrnr</i>	Tüm Parametreler: "Varsayılan"
k-EYK	<i>relieff</i>	Komşu Sayısı: "21" Diğer Parametreler: "Varsayılan"
RO	<i>oobPermutedPredictorImportance</i>	Tüm Parametreler: "Varsayılan"

3.5. Değerlendirme Kriterleri

Tez çalışmasındaki sınıflandırma uygulamalarında çok sınıflı sınıflandırmalar gerçekleştirilmiş olup çok sınıflı bir sınıflandırma için örnek karışıklık matrisi Çizelge 3.13'de yer almaktadır [29]. Çizelge 3.13'de satırlar tahmin edilen sınıf, sütunlar ise gerçek sınıf etiketlerini temsil etmektedir.

Çizelge 3.13. Çok sınıflı bir sınıflandırma için örnek bir karışıklık matrisi

		Gerçek Sınıf			
		Sınıf-1	Sınıf-2	Sınıf-3	Sınıf-m
Tahmin Edilen Sınıf	Sınıf-1	X_{11}	X_{12}	X_{13}	X_{1m}
	Sınıf-2	X_{21}	X_{22}	X_{23}	X_{2m}
	Sınıf-3	X_{31}	X_{32}	X_{33}	X_{3m}
	Sınıf-m	X_{m1}	X_{m2}	X_{m3}	X_{mm}

Tez çalışmasında elde edilen sınıflandırma sonuçlarının değerlendirilmesinde karışıklık matrisi üzerinden hesaplanan Duyarlılık, Özgüllük, Doğruluk ve F1-Skoru parametreleri kullanılmıştır. Söz konusu parametrelerin sınıf bazında ve sınıflandırmanın geneli için ağırlıklı ve ortalama olarak hesaplanmasına ilişkin matematiksel Eşitlikler Çizelge 3.14’de yer almaktadır.

Çizelge 3.14. Çok sınıflı bir sınıflandırma için karışıklık matrisinden üzerinden hesaplanan karşılaştırma parametreleri

Parametre	Sınıf Bazlı Hesaplama	Sınıflandırma Geneli	
		Ağırlıklı Hesaplama	Ortalama Hesaplama
DP	X_{ii}	X	
YN	$\sum_{l=1}^m X_{li} - DP_i$		
YP	$\sum_{l=1}^m X_{il} - DP_i$		
DN	$\sum_{l=1}^m \sum_{k=1}^m X_{lk} - DP_i - YN_i - YP_i$		
Duyarlılık	$DP_i / (DP_i + YN_i)$	$\sum_{i=1}^m (n_i / n) \times Duyarlılık_i$	$(\sum_{i=1}^m Duyarlılık_i) / m$
Özgüllük	$DN_i / (DN_i + YP_i)$	$\sum_{i=1}^m (n_i / n) \times Özgüllük_i$	$(\sum_{i=1}^m Özgüllük_i) / m$
Doğruluk	$\frac{DP_i + DN_i}{DP_i + DN_i + YP_i + YN_i}$	$\sum_{i=1}^m (n_i / n) \times Doğruluk_i$	$(\sum_{i=1}^m Doğruluk_i) / m$
F1-Skoru	$\frac{2 \times DP_i}{2 \times DP_i + YP_i + YN_i}$	$\sum_{i=1}^m (n_i / n) \times F1-Skoru_i$	$(\sum_{i=1}^m F1-Skoru_i) / m$
AİKE-AKA	AİKE Üzerinden	$\sum_{i=1}^m (n_i / n) \times AİKE - AKA_i$	$(\sum_{i=1}^m AİKE - AKA_i) / m$
Genel Doğruluk	$(\sum_{i=1}^m X_{ii}) / (\sum_{l=1}^m \sum_{k=1}^m X_{lk})$		

Çizelge 3.14'ün son satırında, sınıflandırmanın geneli için Genel Doğruluk parametresinin matematiksel olarak hesaplanmasına ilişkin bilgiler verilmiştir. Çizelge 3.14'de i parametreleri hesaplanan sınıfı, m toplam sınıf sayısını, n_i i sınıfına ait eleman sayısını, n ise toplam eleman sayısını temsil etmektedir. Ayrıca, Çizelge 3.14'de Doğru Pozitif (DP) gerçekte pozitif olan örneklerin sınıflandırma sonucunda pozitif olarak sınıflandırılma sayısını, Doğru Negatif (DN) gerçekte negatif olan örneklerin sınıflandırma sonucunda negatif olarak sınıflandırılma sayısını, Yanlış Pozitif (YP) gerçekte negatif olan örneklerin sınıflandırma sonucunda pozitif olarak sınıflandırılma sayısını ve Yanlış Negatif (YN) ise gerçekte pozitif olan örneklerin sınıflandırma sonucunda negatif olarak sınıflandırılma sayısını ifade etmektedir. Tez çalışmasında kullanılan bir diğer karşılaştırma kriteri Alıcı İşlem Karakteristiği Eğrisi (AİKE) Altında Kalan Alan (AİKE-AKA)'dır. AİKE, bir sınıflandırmada gerçek sonuçların etiketlenilmesinde kullanılan eşik değer minimum değerden maksimum değere doğru hareket ederken Duyarlılık parametresi ile Özgüllük parametresinin bire tümleyeni (1-Özgüllük) arasındaki değişimi göstermektedir. AİKE-AKA değeri ile sınıflandırma performansı doğru orantılı olarak değişmektedir. Çok sınıflı bir sınıflandırmada AİKE-AKA büyüklükleri her sınıf için çizdirilen eğriler üzerinden ayrı ayrı hesaplanır. Çizelge 3.14'de AİKE-AKA değerlerinin sınıflandırmanın geneli için ağırlıklı ve ortalama olarak hesaplanmasına ilişkin bilgiler yer almaktadır.



4. DENEYLER VE SONUÇLAR

Tez çalışmasının bu bölümünde sınıflandırma deneylerinin gerçekleştirilme prosedürlerine ilişkin genel bilgiler yer almaktadır. Ayrıca, iki farklı veri seti kullanılarak gerçekleştirilen sınıflandırma deneylerinde elde edilen sayısal sonuçlar verilmiştir.

4.1. Deneyler

Tez çalışmasının sınıflandırma deneylerinde ilk adım olarak veri setlerinden dosyaların okutulması ve düzenlenmesi gerçekleştirilmiştir. Deneylerin programsal kodlamaları MATLAB2023(a) platformu kullanılarak oluşturulmuştur. Bu kapsamda veri setleri içinde yer alan JSON uzantılı dosyalar okutulmuş ve daha önce Şekil 3.1’de yapısı açıklanan paket uzunluğu ve varış süresi özellik vektörleri program platformuna yüklenmiştir. Bu kapsamda, MIRAGE-2019 veri seti için daha önce Çizelge 3.2’de açıklandığı üzere toplam 20 mobil ağ trafik sınıfı için toplam 317 731 özellik vektörü elde edilmiştir. MIRAGE-COVID-CCMA-2022 veri seti içinse daha önce Çizelge 3.3’de açıklandığı üzere toplam 9 mobil ağ trafik sınıfı için toplam 263 619 özellik vektörü elde edilmiştir. Bu özellik vektörlerinin her biri 102 (17×6) adet özellik bileşeni içermektedir. Bu aşamada, söz konusu özellik vektörlerine karşılık gelen sınıf etiketleri de ayrı vektörler olarak oluşturulmuştur. Daha sonra her veri seti kendi içinde olmak üzere bu özellik ve sınıf etiketi vektörleri homojenliğin sağlanması için rastgele oluşturulmuş bir diziye göre karıştırılmıştır. Karıştırma işleminden sonra, veri setlerinde bulunan bazı eksik vektör elemanları *fillmissing* fonksiyonu yardımıyla hareketli bir pencere (pencere uzunluğu=30) kullanılarak medyan değerler ile doldurulmuştur.

Sınıflandırma deneylerinin ikinci aşamasında, ilk aşamada elde edilen özellik vektörleri eğitim ve test olmak üzere iki gruba ayrılmıştır. Söz konusu ayrışma ilişkin detaylar MIRAGE-2019 veri seti ve MIRAGE-COVID-CCMA-2022 veri seti için sırasıyla Çizelge 3.2 ve Çizelge 3.3’de daha önce verilmiştir. Sonuç olarak MIRAGE-2019 veri seti için eğitim işlemlerinde 254 179 özellik vektörü kullanılırken kalan 63 552 özellik vektörü ise test işlemlerinde kullanılmıştır. Benzer şekilde MIRAGE-COVID-CCMA-2022 veri seti için eğitim işlemlerinde 210 892 özellik vektörü kullanılırken kalan 52 727 özellik vektörü ise test işlemlerinde kullanılmıştır. Tez çalışmasında, paket uzunluğu ve varış süresi özellik gruplarının her biri için çift yönlü akış, yukarı yönlü akış ve aşağı yönlü akış özellik

grupları için ayrı ayrı (her biri 17 özellik) ve bu özellik grupları için bir bütün olarak (her biri 51 özellik) deneyler gerçekleştirilmiştir. Ayrıca, tüm bu özelliklerin birleştirilmesiyle elde edilen toplam 102 özellikten 51 tanesinin üç farklı özellik seçme yöntemi kullanılarak seçilmesi ve sınıflandırma işlemlerinde kullanılması durumu için söz konusu deneyler tekrarlanmıştır. Tez çalışmasındaki sınıflandırma deneyleri, iki farklı veri seti, dört farklı makine öğrenmesi sınıflandırıcı, dört farklı derin öğrenme sınıflandırıcısı, özellik gruplarının tek başlarına kullanılması, grup olarak kullanılması ve üç farklı özellik seçme yöntemi ile seçilerek kullanılması durumu için ayrı ayrı gerçekleştirilmiştir. Tez çalışmasında özellik seçimi ve sınıflandırma aşamasındaki çalışma süreleri CPU (Central Processing Unit) üzerinden saniye cinsinden ölçülmüştür. Tez çalışmasının deneyleri 32 Gigabayt RAM'e sahip bir iş istasyonu (Intel(R) Core(TM) i7-9700 CPU @ 3,00 Gigahertz) kullanılarak gerçekleştirilmiştir.

4.2. MIRAGE-2019 Veri Seti Sonuçları

MIRAGE-2019 veri seti için paket uzunluğu ve varış süresi özelliklerinin birleştirilmesiyle elde edilen toplam 102 özellikten 51 tanesi seçilmesine ilişkin sonuçlar Çizelge 4.1'de verilmiştir. Özellik seçme yöntemlerinin CPU çalışma sürelerinin Çizelge 4.1'in son satırı üzerinden incelenmesi mümkündür. MIRAGE-2019 veri seti için makine öğrenmesi sınıflandırıcılar kullanılarak elde edilen sınıflandırma sonuçları Çizelge 4.2 ile Çizelge 4.5 arasında, derin öğrenme sınıflandırıcıları kullanılarak elde edilen sonuçlar ise Çizelge 4.6 ile Çizelge 4.9 arasında yer almaktadır. CPU çalışma süresi bilgilerinin bu Çizelgeler üzerinden incelenmesi mümkündür.

Çizelge 4.1. MIRAGE-2019 veri seti için özellik seçme yöntemleri kullanılarak seçilen özellikler ve çalışma süreleri

Özellik Vektörü		MF MU	k-EYK	RO
Paket Uzunluğu	Çift Yönlü Akış	1, 2, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	2, 3, 4, 5, 6, 7, 11, 12, 13, 14, 15, 16, 17	3, 4, 5, 7, 8, 15, 16
	Yukarı Yönlü Akış	1, 2, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	2, 3, 4, 5, 6, 10, 11, 13, 14, 15, 16, 17	2, 4, 7, 8, 14, 15, 16
	Aşağı Yönlü Akış	1, 2, 3, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	4, 5, 7, 8, 9, 13, 14
Varış Süresi	Çift Yönlü Akış	1, 2, 4, 9	2, 3, 4, 7	9, 10, 11, 12, 13, 14, 15, 16, 17
	Yukarı Yönlü Akış	1, 2, 13	2, 3, 4, 7	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
	Aşağı Yönlü Akış	1, 2, 3, 10, 11	2, 3, 4, 7	1, 6, 9, 10, 11, 12, 14, 16, 17
CPU Çalışma Süresi (Saniye)		25,45	626,99	718,62

Çizelge 4.2. MIRAGE-2019 veri seti için DVM sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
DVM	Paket Uzunluğu	Çift Yönlü Akış	17	0,7144	0,9794	0,9661	0,7101	0,9339	0,7144	5 099,0
		Yukarı Yönlü Akış	17	0,7099	0,9790	0,9655	0,7066	0,9313	0,7099	5 282,9
		Aşağı Yönlü Akış	17	0,7061	0,9800	0,9662	0,7028	0,9345	0,7061	5 498,3
		Tüm Özellikler	51	0,7822	0,9859	0,9758	0,7810	0,9518	0,7822	7 001,0
	Varış Süresi	Çift Yönlü Akış	17	0,5584	0,9610	0,9406	0,5513	0,8723	0,5584	11 589,0
		Yukarı Yönlü Akış	17	0,5540	0,9609	0,9399	0,5487	0,8722	0,5540	11 847,5
		Aşağı Yönlü Akış	17	0,5515	0,9628	0,9415	0,5432	0,8733	0,5515	11 932,6
		Tüm Özellikler	51	0,5994	0,9662	0,9474	0,5930	0,8933	0,5994	14 218,1
	Özellik Seçme	MFMU	51	0,7847	0,9861	0,9761	0,7836	0,9520	0,7847	6 748,2
		k-EYK	51	0,7929	0,9865	0,9768	0,7920	0,9536	0,7929	6 628,3
		RO	51	0,7704	0,9842	0,9736	0,7689	0,9502	0,7704	6 840,4

Çizelge 4.3. MIRAGE-2019 veri seti için k-EYK sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
k-EYK	Paket Uzunluğu	Çift Yönlü Akış	17	0,6609	0,9771	0,9604	0,6600	0,9051	0,6609	192,9
		Yukarı Yönlü Akış	17	0,6615	0,9782	0,9614	0,6671	0,8949	0,6615	197,2
		Aşağı Yönlü Akış	17	0,6722	0,9780	0,9619	0,6730	0,9127	0,6722	198,3
		Tüm Özellikler	51	0,7274	0,9821	0,9689	0,7260	0,9409	0,7274	562,0
	Varış Süresi	Çift Yönlü Akış	17	0,6308	0,9743	0,9556	0,6268	0,9237	0,6308	197,1
		Yukarı Yönlü Akış	17	0,6277	0,9738	0,9550	0,6238	0,9215	0,6277	196,2
		Aşağı Yönlü Akış	17	0,6201	0,9736	0,9545	0,6151	0,9221	0,6201	197,6
		Tüm Özellikler	51	0,6496	0,9757	0,9582	0,6455	0,9319	0,6496	559,4
	Özellik Seçme	MFMU	51	0,7564	0,9845	0,9727	0,7546	0,9659	0,7564	784,3
		k-EYK	51	0,7512	0,9839	0,9719	0,7489	0,9643	0,7512	690,2
		RO	51	0,7499	0,9837	0,9716	0,7477	0,9629	0,7499	871,5

Çizelge 4.4. MIRAGE-2019 veri seti için AG sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
AG	Paket Uzunluğu	Çift Yönlü Akış	17	0,3510	0,9323	0,9055	0,2854	0,7892	0,3510	218,4
		Yukarı Yönlü Akış	17	0,3105	0,9166	0,8892	0,2517	0,7696	0,3105	215,7
		Aşağı Yönlü Akış	17	0,3877	0,9418	0,9153	0,3293	0,7988	0,3877	214,0
		Tüm Özellikler	51	0,4002	0,9405	0,9152	0,3455	0,8302	0,4002	547,5
	Varış Süresi	Çift Yönlü Akış	17	0,3012	0,9189	0,8905	0,2339	0,7562	0,3012	348,3
		Yukarı Yönlü Akış	17	0,2910	0,9191	0,8900	0,2241	0,7485	0,2910	353,5
		Aşağı Yönlü Akış	17	0,3215	0,9288	0,9001	0,2506	0,7680	0,3215	333,2
		Tüm Özellikler	51	0,3409	0,9301	0,9023	0,2779	0,7788	0,3409	960,7
	Özellik Seçme	MFMU	51	0,4199	0,9418	0,9180	0,3664	0,8375	0,4199	732,4
		k-EYK	51	0,4067	0,9417	0,9172	0,3487	0,8303	0,4067	665,7
		RO	51	0,3667	0,9266	0,9020	0,3209	0,8141	0,3667	844,6

Çizelge 4.5. MIRAGE-2019 veri seti için RO sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
RO	Paket Uzunluğu	Çift Yönlü Akış	17	0,8427	0,9896	0,9821	0,8429	0,9816	0,8427	279,0
		Yukarı Yönlü Akış	17	0,8294	0,9895	0,9812	0,8319	0,9814	0,8294	258,5
		Aşağı Yönlü Akış	17	0,8384	0,9902	0,9825	0,8390	0,9838	0,8384	262,1
		Tüm Özellikler	51	0,8883	0,9933	0,9880	0,8884	0,9903	0,8883	443,4
	Varış Süresi	Çift Yönlü Akış	17	0,8448	0,9896	0,9821	0,8444	0,9809	0,8448	398,1
		Yukarı Yönlü Akış	17	0,8353	0,9888	0,9808	0,8348	0,9772	0,8353	407,9
		Aşağı Yönlü Akış	17	0,8419	0,9895	0,9820	0,8412	0,9795	0,8419	386,4
		Tüm Özellikler	51	0,8634	0,9909	0,9844	0,8630	0,9858	0,8634	636,6
	Özellik Seçme	MFMU	51	0,9191	0,9952	0,9914	0,9190	0,9937	0,9191	537,4
		k-EYK	51	0,9150	0,9949	0,9910	0,9148	0,9930	0,9150	473,7
		RO	51	0,9090	0,9943	0,9901	0,9088	0,9922	0,9090	616,4

Çizelge 4.6. MIRAGE-2019 veri seti için ESA sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
ESA	Paket Uzunluğu	Çift Yönlü Akış	17	0,5276	0,9642	0,9410	0,5136	0,8922	0,5276	2 151,5
		Yukarı Yönlü Akış	17	0,5847	0,9678	0,9487	0,5744	0,9169	0,5847	2 131,0
		Aşağı Yönlü Akış	17	0,5067	0,9573	0,9346	0,4861	0,8852	0,5067	2 063,5
		Tüm Özellikler	51	0,6639	0,9770	0,9614	0,6592	0,9451	0,6639	2 162,1
	Varış Süresi	Çift Yönlü Akış	17	0,5143	0,9630	0,9390	0,5033	0,8837	0,5143	2 040,3
		Yukarı Yönlü Akış	17	0,4618	0,9573	0,9309	0,4452	0,8606	0,4618	2 040,6
		Aşağı Yönlü Akış	17	0,4727	0,9567	0,9314	0,4551	0,8638	0,4727	2 033,5
		Tüm Özellikler	51	0,5101	0,9619	0,9383	0,4944	0,8823	0,5101	2 163,0
	Özellik Seçme	MFMU	51	0,5679	0,9669	0,9471	0,5506	0,9110	0,5679	2 216,1
		k-EYK	51	0,6801	0,9800	0,9646	0,6768	0,9501	0,6801	2 161,3
		RO	51	0,6072	0,9704	0,9522	0,5994	0,9212	0,6072	2 207,0

Çizelge 4.7. MIRAGE-2019 veri seti için UKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
UKSB	Paket Uzunluğu	Çift Yönlü Akış	17	0,3242	0,9328	0,9017	0,2832	0,7746	0,3242	758,0
		Yukarı Yönlü Akış	17	0,3446	0,9341	0,9041	0,3098	0,7916	0,3446	749,9
		Aşağı Yönlü Akış	17	0,3415	0,9445	0,9121	0,3091	0,7876	0,3415	747,6
		Tüm Özellikler	51	0,4945	0,9606	0,9370	0,4699	0,8704	0,4945	842,5
	Varış Süresi	Çift Yönlü Akış	17	0,4962	0,9587	0,9348	0,4791	0,8789	0,4962	750,9
		Yukarı Yönlü Akış	17	0,5048	0,9598	0,9361	0,4899	0,8817	0,5048	751,3
		Aşağı Yönlü Akış	17	0,4902	0,9596	0,9350	0,4726	0,8804	0,4902	753,0
		Tüm Özellikler	51	0,5542	0,9673	0,9456	0,5451	0,9034	0,5542	832,0
	Özellik Seçme	MFMU	51	0,5567	0,9694	0,9470	0,5485	0,9108	0,5567	861,8
		k-EYK	51	0,5431	0,9623	0,9416	0,5313	0,8987	0,5431	856,4
		RO	51	0,6116	0,9725	0,9539	0,6042	0,9245	0,6116	859,2

Çizelge 4.8. MIRAGE-2019 veri seti için ÇYUKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
ÇYUKSB	Paket Uzunluğu	Çift Yönlü Akış	17	0,3218	0,9350	0,9031	0,2823	0,7689	0,3218	1 060,4
		Yukarı Yönlü Akış	17	0,3831	0,9470	0,9169	0,3545	0,8208	0,3831	1 051,7
		Aşağı Yönlü Akış	17	0,3883	0,9478	0,9183	0,3648	0,8142	0,3883	1 047,0
		Tüm Özellikler	51	0,5362	0,9674	0,9452	0,5173	0,8918	0,5362	1 197,8
	Varış Süresi	Çift Yönlü Akış	17	0,5374	0,9651	0,9427	0,5251	0,8959	0,5374	1 047,3
		Yukarı Yönlü Akış	17	0,5410	0,9638	0,9419	0,5295	0,8965	0,5410	1 055,8
		Aşağı Yönlü Akış	17	0,5359	0,9646	0,9420	0,5227	0,8977	0,5359	1 042,5
		Tüm Özellikler	51	0,6000	0,9700	0,9510	0,5913	0,9221	0,6000	1 188,6
	Özellik Seçme	MFMU	51	0,5479	0,9634	0,9426	0,5380	0,9042	0,5479	1 260,7
		k-EYK	51	0,5756	0,9692	0,9486	0,5684	0,9142	0,5756	1 228,9
		RO	51	0,6785	0,9788	0,9634	0,6754	0,9472	0,6785	1 271,6

Çizelge 4.9. MIRAGE-2019 veri seti için KTB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
KTB	Paket Uzunluğu	Çift Yönlü Akış	17	0,3623	0,9393	0,9098	0,3263	0,8046	0,3623	731,1
		Yukarı Yönlü Akış	17	0,3701	0,9438	0,9132	0,3507	0,8141	0,3701	727,0
		Aşağı Yönlü Akış	17	0,3881	0,9473	0,9177	0,3597	0,8209	0,3881	725,3
		Tüm Özellikler	51	0,6179	0,9737	0,9555	0,6075	0,9276	0,6179	797,1
	Varış Süresi	Çift Yönlü Akış	17	0,5171	0,9617	0,9388	0,5020	0,8856	0,5171	723,6
		Yukarı Yönlü Akış	17	0,5189	0,9627	0,9392	0,5078	0,8845	0,5189	727,6
		Aşağı Yönlü Akış	17	0,5229	0,9635	0,9404	0,5092	0,8907	0,5229	727,9
		Tüm Özellikler	51	0,5929	0,9706	0,9510	0,5847	0,9174	0,5929	792,2
	Özellik Seçme	MFMU	51	0,6082	0,9709	0,9527	0,5988	0,9304	0,6082	845,5
		k-EYK	51	0,6170	0,9736	0,9550	0,6103	0,9275	0,6170	832,8
		RO	51	0,6279	0,9745	0,9565	0,6219	0,9303	0,6279	827,0

Çizelge 4.2 ile Çizelge 4.9 arasında yer alan Çizelgelerde her sınıflandırıcı için özellik seçimi yapılmadan önce ve özellik seçimi yapıktan sonra elde edilen en yüksek sonuçlar takibin kolay yapılabilmesi amacıyla kırmızı renk ile belirtilmiştir.

4.3. MIRAGE-COVID-CCMA-2022 Veri Seti Sonuçları

MIRAGE-COVID-CCMA-2022 veri seti için paket uzunluğu ve varış süresi özelliklerinin birleştirilmesiyle elde edilen toplam 102 özellikten 51 tanesinin seçilmesine ilişkin sonuçlar Çizelge 4.10’da verilmiştir. Özellik seçme yöntemlerinin CPU çalışma sürelerinin Çizelge 4.10’un son satırı üzerinden incelenmesi mümkündür.

Çizelge 4.10. MIRAGE-COVID-CCMA-2022 veri seti için özellik seçme yöntemleri kullanılarak seçilen özellikler ve çalışma süreleri

Özellik Vektörü		MF MU	k-EYK	RO
Paket Uzunluğu	Çift Yönlü Akış	1, 2, 9, 10, 11, 12, 13, 14, 15, 17	2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	2, 8, 14, 16, 17
	Yukarı Yönlü Akış	1, 2, 4, 6, 9, 10, 11, 12, 13, 15, 16, 17	1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	1, 2, 3, 4, 5, 7, 8, 17
	Aşağı Yönlü Akış	1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16,	2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17	1, 2, 3, 5, 6, 12, 13, 14, 15, 17
Varış Süresi	Çift Akış	1, 2	2, 4	2, 7, 8, 9, 10, 11, 12, 17
	Yukarı Yönlü Akış	1, 2, 8	2, 4, 7	6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17
	Aşağı Yönlü Akış	1, 2, 3, 4, 5, 8, 16, 17	2, 4, 17	1, 7, 8, 9, 10, 11, 12, 16
CPU Çalışma Süresi (Saniye)		19,41	411,11	502,98

MIRAGE-COVID-CCMA-2022 veri seti için makine öğrenmesi sınıflandırıcılar kullanılarak elde edilen sınıflandırma sonuçları Çizelge 4.11 ile Çizelge 4.14 arasında, derin öğrenme sınıflandırıcıları kullanılarak elde edilen sonuçlar ise Çizelge 4.15 ile Çizelge 4.18 arasında yer almaktadır. CPU çalışma süresi bilgilerinin bu Çizelgeler üzerinden incelenmesi mümkündür. Çizelge 4.11 ile Çizelge 4.18 arasında yer alan Çizelgelerde her sınıflandırıcı için özellik seçimi yapılmadan önce ve özellik seçimi yapıktan sonra elde edilen en yüksek sonuçlar takibin kolay yapılabilmesi amacıyla kırmızı renk ile belirtilmiştir.

Çizelge 4.11. MIRAGE-COVID-CCMA-2022 veri seti için DVM sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
DVM	Paket Uzunluğu	Çift Yönlü Akış	17	0,8526	0,9676	0,9610	0,8520	0,9517	0,8526	1 861,5
		Yukarı Yönlü Akış	17	0,8408	0,9609	0,9568	0,8405	0,9414	0,8408	2 142,5
		Aşağı Yönlü Akış	17	0,8473	0,9646	0,9585	0,8467	0,9470	0,8473	1 884,3
		Tüm Özellikler	51	0,8695	0,9760	0,9678	0,8704	0,9549	0,8695	3 168,7
	Varış Süresi	Çift Yönlü Akış	17	0,7576	0,8533	0,8970	0,7404	0,9280	0,7576	3 954,0
		Yukarı Yönlü Akış	17	0,7602	0,8701	0,9030	0,7465	0,9234	0,7602	4 123,6
		Aşağı Yönlü Akış	17	0,7746	0,8820	0,9116	0,7622	0,9257	0,7746	3 845,1
		Tüm Özellikler	51	0,7972	0,9057	0,9242	0,7881	0,9387	0,7972	6 367,1
	Özellik Seçme	MFMU	51	0,8750	0,9750	0,9686	0,8750	0,9570	0,8750	3 356,1
		k-EYK	51	0,8712	0,9714	0,9667	0,8710	0,9570	0,8712	3 588,5
		RO	51	0,8742	0,9781	0,9691	0,8751	0,9582	0,8742	3 324,6

Çizelge 4.12. MIRAGE-COVID-CCMA-2022 veri seti için k-EYK sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
k-EYK	Paket Uzunluğu	Çift Yönlü Akış	17	0,7712	0,9642	0,9322	0,7916	0,8983	0,7712	135,6
		Yukarı Yönlü Akış	17	0,6977	0,9578	0,9000	0,7363	0,8481	0,6977	144,8
		Aşağı Yönlü Akış	17	0,7161	0,9641	0,9081	0,7491	0,8670	0,7161	145,3
		Tüm Özellikler	51	0,7934	0,9738	0,9422	0,8103	0,9126	0,7934	596,8
	Varış Süresi	Çift Yönlü Akış	17	0,8447	0,9589	0,9561	0,8426	0,9726	0,8447	143,8
		Yukarı Yönlü Akış	17	0,8424	0,9555	0,9540	0,8399	0,9714	0,8424	146,3
		Aşağı Yönlü Akış	17	0,8480	0,9603	0,9565	0,8463	0,9734	0,8480	160,6
		Tüm Özellikler	51	0,8579	0,9633	0,9601	0,8564	0,9765	0,8579	596,8
	Özellik Seçme	MFMU	51	0,8990	0,9788	0,9742	0,8989	0,9861	0,8990	370,1
		k-EYK	51	0,8989	0,9787	0,9739	0,8987	0,9861	0,8989	388,9
		RO	51	0,8978	0,9781	0,9735	0,8975	0,9861	0,8978	380,8

Çizelge 4.13. MIRAGE-COVID-CCMA-2022 veri seti için AG sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
AG	Paket Uzunluğu	Çift Yönlü Akış	17	0,6615	0,7698	0,8459	0,6005	0,8944	0,6615	122,5
		Yukarı Yönlü Akış	17	0,6475	0,7621	0,8421	0,5792	0,8796	0,6475	116,5
		Aşağı Yönlü Akış	17	0,6613	0,7734	0,8480	0,6014	0,8846	0,6613	125,9
		Tüm Özellikler	51	0,7065	0,8198	0,8746	0,6685	0,9074	0,7065	319,0
	Varış Süresi	Çift Yönlü Akış	17	0,6182	0,7075	0,8135	0,5382	0,8633	0,6182	171,6
		Yukarı Yönlü Akış	17	0,5957	0,6797	0,7999	0,5069	0,8596	0,5957	174,2
		Aşağı Yönlü Akış	17	0,6223	0,7286	0,8250	0,5331	0,8783	0,6223	170,1
		Tüm Özellikler	51	0,6412	0,7338	0,8297	0,5719	0,8913	0,6412	513,9
	Özellik Seçme	MFMU	51	0,7160	0,8358	0,8837	0,6855	0,9121	0,7160	334,0
		k-EYK	51	0,7161	0,8311	0,8826	0,6824	0,9096	0,7161	335,0
		RO	51	0,7084	0,8199	0,8766	0,6735	0,9200	0,7084	421,9

Çizelge 4.14. MIRAGE-COVID-CCMA-2022 veri seti için RO sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA		
RO	Paket Uzunluğu	Çift Yönlü Akış	17	0,8850	0,9768	0,9705	0,8850	0,9890	0,8850	160,8
		Yukarı Yönlü Akış	17	0,8612	0,9722	0,9649	0,8624	0,9837	0,8612	149,5
		Aşağı Yönlü Akış	17	0,8800	0,9732	0,9685	0,8794	0,9870	0,8800	162,1
		Tüm Özellikler	51	0,8987	0,9834	0,9764	0,8990	0,9910	0,8987	304,9
	Varış Süresi	Çift Yönlü Akış	17	0,9279	0,9841	0,9803	0,9276	0,9950	0,9279	222,5
		Yukarı Yönlü Akış	17	0,9240	0,9824	0,9790	0,9236	0,9941	0,9240	223,8
		Aşağı Yönlü Akış	17	0,9272	0,9828	0,9799	0,9269	0,9944	0,9272	221,6
		Tüm Özellikler	51	0,9379	0,9875	0,9838	0,9377	0,9968	0,9379	383,3
	Özellik Seçme	MFMU	51	0,9509	0,9925	0,9888	0,9509	0,9976	0,9509	309,9
		k-EYK	51	0,9486	0,9920	0,9881	0,9486	0,9970	0,9486	298,4
		RO	51	0,9516	0,9925	0,9887	0,9517	0,9978	0,9516	361,6

Çizelge 4.15. MIRAGE-COVID-CCMA-2022 veri seti için ESA sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
ESA	Paket Uzunluğu	Çift Yönlü Akış	17	0,8099	0,9492	0,9452	0,8063	0,9703	0,8099	1 719,4
		Yukarı Yönlü Akış	17	0,7895	0,9271	0,9341	0,7821	0,9607	0,7895	1 698,5
		Aşağı Yönlü Akış	17	0,7571	0,9037	0,9168	0,7409	0,9425	0,7571	1 706,7
		Tüm Özellikler	51	0,8460	0,9692	0,9604	0,8448	0,9794	0,8460	1 813,0
	Varış Süresi	Çift Yönlü Akış	17	0,5031	0,4969	0,7098	0,3367	0,5041	0,5031	1 697,9
		Yukarı Yönlü Akış	17	0,7904	0,9328	0,9354	0,7846	0,9599	0,7904	1 702,6
		Aşağı Yönlü Akış	17	0,7948	0,9394	0,9390	0,7893	0,9602	0,7948	1 708,8
		Tüm Özellikler	51	0,7764	0,9390	0,9326	0,7685	0,9556	0,7764	1 815,3
	Özellik Seçme	MFMU	51	0,8464	0,9502	0,9543	0,8430	0,9813	0,8464	1 805,1
		k-EYK	51	0,5031	0,4969	0,7098	0,3367	0,5760	0,5031	1 789,6
		RO	51	0,7613	0,9368	0,9259	0,7554	0,9502	0,7613	1 795,3

Çizelge 4.16. MIRAGE-COVID-CCMA-2022 veri seti için UKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
UKSB	Paket Uzunluğu	Çift Yönlü Akış	17	0,6802	0,8347	0,8716	0,6507	0,9035	0,6802	628,7
		Yukarı Yönlü Akış	17	0,6049	0,7551	0,8213	0,5615	0,8508	0,6049	619,5
		Aşağı Yönlü Akış	17	0,6667	0,8753	0,8824	0,6423	0,8973	0,6667	619,0
		Tüm Özellikler	51	0,7664	0,9398	0,9347	0,7583	0,9488	0,7664	691,7
	Varış Süresi	Çift Yönlü Akış	17	0,7694	0,9294	0,9281	0,7634	0,9524	0,7694	618,7
		Yukarı Yönlü Akış	17	0,7756	0,9329	0,9299	0,7693	0,9551	0,7756	621,9
		Aşağı Yönlü Akış	17	0,7986	0,9451	0,9412	0,7934	0,9637	0,7986	621,2
		Tüm Özellikler	51	0,8205	0,9491	0,9474	0,8156	0,9723	0,8205	691,0
	Özellik Seçme	MFMU	51	0,8302	0,9581	0,9529	0,8273	0,9750	0,8302	692,7
		k-EYK	51	0,7859	0,9419	0,9375	0,7806	0,9592	0,7859	684,6
		RO	51	0,8300	0,9658	0,9563	0,8280	0,9738	0,8300	685,9

Çizelge 4.17. MIRAGE-COVID-CCMA-2022 veri seti için ÇYUKSB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
ÇYUKSB	Paket Uzunluğu	Çift Yönlü Akış	17	0,6955	0,8587	0,8825	0,6731	0,9085	0,6955	867,6
		Yukarı Yönlü Akış	17	0,6710	0,8467	0,8757	0,6447	0,9061	0,6710	864,4
		Aşağı Yönlü Akış	17	0,7024	0,8865	0,8972	0,6802	0,9167	0,7024	865,3
		Tüm Özellikler	51	0,7934	0,9443	0,9418	0,7879	0,9629	0,7934	1 001,9
	Varış Süresi	Çift Yönlü Akış	17	0,7866	0,9246	0,9316	0,7776	0,9604	0,7866	869,3
		Yukarı Yönlü Akış	17	0,7925	0,9384	0,9373	0,7883	0,9605	0,7925	864,3
		Aşağı Yönlü Akış	17	0,8047	0,9473	0,9441	0,8012	0,9660	0,8047	868,9
		Tüm Özellikler	51	0,8290	0,9598	0,9532	0,8254	0,9748	0,8290	999,3
	Özellik Seçme	MFMU	51	0,8233	0,9581	0,9517	0,8215	0,9723	0,8233	1 003,9
		k-EYK	51	0,7945	0,9491	0,9421	0,7910	0,9628	0,7945	1 003,5
		RO	51	0,8457	0,9608	0,9574	0,8438	0,9796	0,8457	1 000,4

Çizelge 4.18. MIRAGE-COVID-CCMA-2022 veri seti için KTB sınıflandırıcı kullanılarak elde edilen sınıflandırma sonuçları ve çalışma süreleri

Sınıflandırıcı	Özellik Vektörü		Özellik Sayısı	Ağırlıklı					Genel Doğruluk	CPU Çalışma Süresi (Saniye)
				Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AIKE-AKA		
KTB	Paket Uzunluğu	Çift Yönlü Akış	17	0,7021	0,8703	0,8900	0,6795	0,9209	0,7021	616,3
		Yukarı Yönlü Akış	17	0,6720	0,8369	0,8732	0,6402	0,8990	0,6720	610,6
		Aşağı Yönlü Akış	17	0,7192	0,9041	0,9072	0,7053	0,9254	0,7192	608,9
		Tüm Özellikler	51	0,8316	0,9608	0,9552	0,8293	0,9758	0,8316	678,3
	Varış Süresi	Çift Yönlü Akış	17	0,7918	0,9372	0,9374	0,7868	0,9599	0,7918	610,5
		Yukarı Yönlü Akış	17	0,8114	0,9441	0,9445	0,8072	0,9683	0,8114	608,4
		Aşağı Yönlü Akış	17	0,8026	0,9490	0,9444	0,8001	0,9664	0,8026	605,7
		Tüm Özellikler	51	0,8303	0,9668	0,9558	0,8292	0,9757	0,8303	676,5
	Özellik Seçme	MFMU	51	0,8385	0,9420	0,9486	0,8332	0,9783	0,8385	680,5
		k-EYK	51	0,8209	0,9615	0,9522	0,8192	0,9724	0,8209	675,8
		RO	51	0,8561	0,9687	0,9628	0,8541	0,9820	0,8561	678,7



5. TARTIŞMA

Tez çalışmasının bu bölümünde deneylerde elde edilen sonuçlar özetlenmiş, kendi içinde tartışılmış ve geçmiş çalışmalarda elde edilen sonuçlarla karşılaştırılmıştır. Ayrıca, elde edilen en iyi sonuçlara ilişkin karışıklık matrisi ve AİKE grafikleri gibi daha detaylı bilgiler verilmiştir.

5.1. MIRAGE-2019 Veri Seti Sonuçlarının Değerlendirilmesi

MIRAGE-2019 veri seti için makine öğrenmesi ve derin öğrenme sınıflandırıcılar kullanılarak elde edilen mobil ağ trafiği sınıflandırma sonuçları daha önce sırasıyla Çizelge 4.2 ile Çizelge 4.5 arasında ve Çizelge 4.6 ile Çizelge 4.9 arasında verilmiştir. Bu bölümde en başarılı sınıflandırma sonuçları özetlenmiş ve tartışılmıştır. Bu kapsamda özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonrası elde edilen en yüksek sınıflandırma sonuçları ve karşılaştırma parametrelerin değişim miktarları Çizelge 5.1’de verilmiştir. Çizelge 5.1’de takibin kolay yapılabilmesi amacıyla her sınıflandırıcı için parametrelerdeki değişim miktarları yeşil renk ile belirtilmiştir. Ayrıca, sınıflandırmanın geneli için özellik seçimi yapılmadan önce ve özellik seçimi yapıktan sonra elde edilen en yüksek sonuçlar kırmızı renk ile verilmiştir.

Çizelge 5.1 incelendiğinde, özellik seçimi yapılmasıyla birlikte tüm sınıflandırıcılar için hemen hemen tüm karşılaştırma parametrelerinde iyileşme sağlandığı görülmektedir. Bu kapsamda ağırlıklı duyarlılık parametresi %0,9992 ile %7,8471 arasında değişen oranlarda iyileşme göstermiştir. Ağırlıklı özgüllük parametresi -%0,0001 ile %0,875 arasında değişen oranlarda bir değişim gösterirken, ağırlıklı doğruluk parametresi %0,0981 ile %1,2416 arasında ve ağırlıklı F1-Skoru parametresi %1,1 ile %8,4093 arasında değişen oranlarda iyileşmiştir. Benzer şekilde, ağırlıklı AİKE-AKA parametresi %0,1787 ile %2,5089 arasında ve genel doğruluk parametresi %0,9992 ile %7,8471 arasında artış göstermiştir.

Çizelge 5.1. MIRAGE-2019 veri seti için elde edilen en iyi sonuçların özeti

Sınıflandırıcı	Aşama/Değişim Miktarı	Özellik Sayısı	Ağırlıklı					Genel Doğruluk
			Duyarlılık	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA	
DVM	Özellik Seçiminden Önce	51	0,7822	0,9859	0,9758	0,7810	0,9518	0,7822
	Özellik Seçiminden Sonra	51	0,7929	0,9865	0,9768	0,7920	0,9536	0,7929
	Değişim Miktarı (%)		1,0637	0,0536	0,0981	1,1000	0,1787	1,0637
k-EYK	Özellik Seçiminden Önce	51	0,7274	0,9821	0,9689	0,7260	0,9409	0,7274
	Özellik Seçiminden Sonra	51	0,7564	0,9845	0,9727	0,7546	0,9659	0,7564
	Değişim Miktarı (%)		2,8984	0,2400	0,3818	2,8614	2,5089	2,8984
AG	Özellik Seçiminden Önce	51	0,4002	0,9418	0,9153	0,3455	0,8302	0,4002
	Özellik Seçiminden Sonra	51	0,4199	0,9418	0,9180	0,3664	0,8375	0,4199
	Değişim Miktarı (%)		1,9716	-0,0001	0,2680	2,0937	0,7266	1,9716
RO	Özellik Seçiminden Önce	51	0,8883	0,9933	0,9880	0,8884	0,9903	0,8883
	Özellik Seçiminden Sonra	51	0,9191	0,9952	0,9914	0,9190	0,9937	0,9191
	Değişim Miktarı (%)		3,0762	0,1905	0,3405	3,0547	0,3427	3,0762
ESA	Özellik Seçiminden Önce	51	0,6639	0,9770	0,9614	0,6592	0,9451	0,6639
	Özellik Seçiminden Sonra	51	0,6801	0,9800	0,9646	0,6768	0,9501	0,6801
	Değişim Miktarı (%)		1,6239	0,3035	0,3236	1,7667	0,4942	1,6239
UKSB	Özellik Seçiminden Önce	51	0,5542	0,9673	0,9456	0,5451	0,9034	0,5542
	Özellik Seçiminden Sonra	51	0,6116	0,9725	0,9539	0,6042	0,9245	0,6116
	Değişim Miktarı (%)		5,7402	0,5150	0,8335	5,9059	2,1111	5,7402
ÇYUKSB	Özellik Seçiminden Önce	51	0,6000	0,9700	0,9510	0,5913	0,9221	0,6000
	Özellik Seçiminden Sonra	51	0,6785	0,9788	0,9634	0,6754	0,9472	0,6785
	Değişim Miktarı (%)		7,8471	0,8750	1,2416	8,4093	2,5052	7,8471
KTB	Özellik Seçiminden Önce	51	0,6179	0,9737	0,9555	0,6075	0,9276	0,6179
	Özellik Seçiminden Sonra	51	0,6279	0,9745	0,9565	0,6219	0,9304	0,6279
	Değişim Miktarı (%)		0,9992	0,0784	0,0990	1,4390	0,2833	0,9992

Çizelge 5.1 incelendiğinde, özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra en yüksek sınıflandırma sonuçlarının RO sınıflandırıcısı ile elde edildiği görülmektedir. Sınıflandırma başarısı açısından RO sınıflandırıcısı DVM ve k-EYK sınıflandırıcılar takip etmektedir. En başarısız sınıflandırıcının ise AG olduğu görülmektedir. Sınıflandırma başarısı açısından derin öğrenme sınıflandırıcılarının başarısı, k-EYK ve AG sınıflandırıcılarının arasında sıralanmaktadır.

Makine öğrenmesi ve derin öğrenme sınıflandırıcıların çalışma süreleri karşılaştırıldığında, k-EYK, AG ve RO en hızlı sınıflandırıcılar olarak sıralanırken DVM'nin en yavaş sınıflandırıcı olduğu görülmektedir. Derin öğrenme sınıflandırıcılar ise DVM'den daha hızlı, diğer makine öğrenimi sınıflandırıcılarından daha yavaş çalışmaktadır. Söz konusu

bu çalışma süreleri ağ trafiği başına eğitim ve test dahil sınıflandırma sürelerini yansıtmaktadır. Ancak, bu çalışma sürelerinin önemli bir kısmının eğitim süreçlerini kapsadığı hususunun dikkate alınması gerekmektedir.

Sınıflandırma sonuçlarına etkisi açısından özellik seçim yöntemlerinin başarısı ve çalışma süreleri kendi aralarında karşılaştırıldığında, MFMU yönteminin k-EYK yöntemine göre hem daha hızlı hem de daha başarılı sonuçların elde edilmesini sağladığı görülmektedir. Ayrıca, MFMU yöntemi k-EYK yöntemine benzer şekilde RO yönteminden daha hızlıdır.

Çizelge 5.1 incelendiğinde, özellik seçimi yapılmadan önce sınıflandırma sonucunda elde edilen en yüksek ağırlıklı duyarlılık, özgüllük, doğruluk, F1-Skoru, AİKE-AKA ve genel doğruluk parametresinin sırasıyla 0,8883; 0,9933; 0,9880; 0,8884; 0,9903 ve 0,8883 olduğu görülmektedir. Özellik seçimi yapıldıktan sonra bu parametreler sırasıyla 0,9191; 0,9952; 0,9914; 0,9190; 0,9937 ve 0,9191 seviyesine yükselmiştir. Şekil 5.1’de özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri verilmiştir. Şekil 5.2’de ise, en yüksek sayıda ağ trafiğine sahip ilk sekiz mobil uygulama için özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonraki AİKE grafiklerinin durumu yer almaktadır.

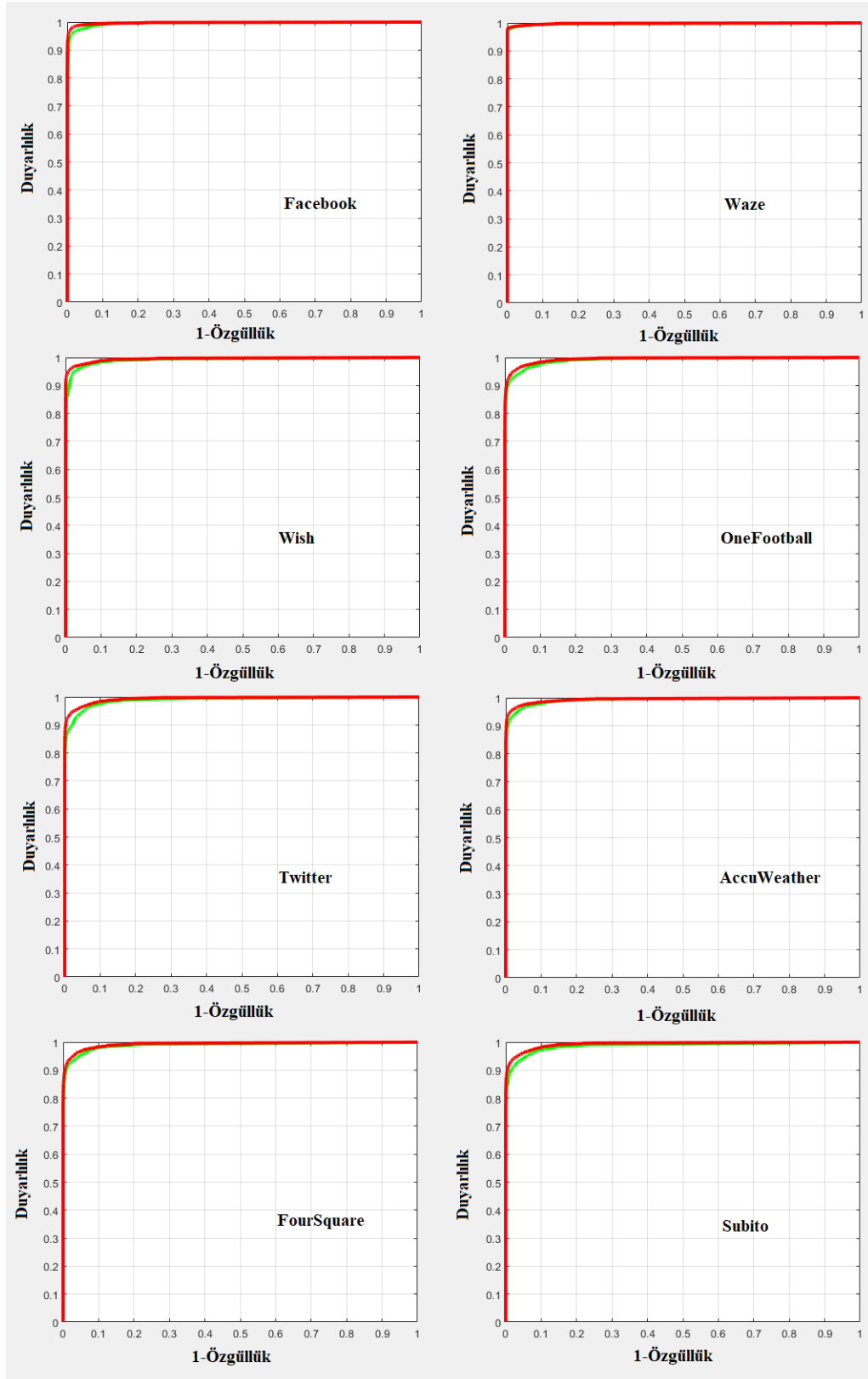
	AccuWeather	Comics	Dropbox	Duolingo	Facebook	FourSquare	Groupon	Messenger	OneFootball	Pinterest	Slitter.io	Spotify	Subito	Trello	TripAdvisor	Twitter	Viber	Waze	Wish	Youtube
AccuWeather	3 796	23	8	12	11	18	9	2	49	22	18	9	52	6	20	17	21	10	25	7
Comics	36	2 040	3	22	13	6	3	6	26	13	3	2	10	3	4	16	19	6	12	4
Dropbox	8	7	1 872	37	8	11	8	2	12	8	8	11	3	13	10	31	11	12	16	12
Duolingo	17	10	52	2 134	11	18	9	7	32	16	31	10	25	21	23	49	24	6	25	36
Facebook	22	5	18	14	7 088	15	6	287	29	14	10	22	13	13	10	22	26	15	17	14
FourSquare	19	5	20	15	2	2 897	16	2	43	26	13	19	26	15	28	26	15	13	25	21
Groupon	18	7	13	6	3	19	1 014	0	1	15	1	6	14	3	13	12	7	7	10	4
Messenger	2	2	2	4	122	1	0	1 103	6	3	5	2	1	3	1	6	1	2	7	2
OneFootball	46	29	19	34	28	30	19	7	4 139	39	7	30	37	8	41	37	38	11	27	17
Pinterest	33	18	28	28	18	33	7	17	40	2 495	16	23	29	39	43	64	31	17	42	37
Slitter.io	24	23	38	12	44	34	14	38	22	12	1 028	26	34	36	30	45	29	18	16	25
Spotify	6	5	14	9	3	20	4	8	21	12	8	2 790	6	4	16	12	8	2	11	8
Subito	50	11	28	48	35	14	17	17	35	26	41	18	2 703	17	32	13	29	6	18	16
Trello	5	13	24	27	4	24	7	3	18	13	6	11	22	1 060	20	22	10	8	15	9
TripAdvisor	16	9	19	25	10	27	17	9	16	23	3	12	9	9	1 209	18	13	6	13	12
Twitter	11	26	51	54	22	22	13	11	19	33	13	17	40	14	50	3 831	28	12	28	23
Viber	4	5	16	14	9	4	3	2	17	2	13	3	7	3	9	10	1 388	5	5	3
Waze	6	11	9	5	2	4	6	1	15	1	8	9	6	2	1	7	8	6 962	7	5
Wish	117	16	27	85	64	43	37	6	87	89	13	15	73	12	22	135	29	12	4 442	34
Youtube	14	10	14	20	10	20	10	3	16	11	11	17	12	8	5	15	17	10	14	2 464

(a)

	AccuWeather	Comics	Dropbox	Duolingo	Facebook	FourSquare	Groupon	Messenger	OneFootball	Pinterest	Slitter.io	Spotify	Subito	Trello	TripAdvisor	Twitter	Viber	Waze	Wish	Youtube
AccuWeather	3 952	16	5	5	11	30	4	4	61	13	12	11	48	7	22	18	16	6	14	12
Comics	36	2 080	4	11	8	5	5	9	28	16	1	3	12	3	9	12	20	0	5	5
Dropbox	10	10	1 967	16	6	22	6	3	14	14	7	15	8	18	18	25	11	13	13	16
Duolingo	12	10	47	2 320	11	20	7	3	35	10	20	11	31	22	37	42	24	8	24	43
Facebook	10	5	8	16	7 265	15	2	139	22	12	6	11	6	3	22	15	18	9	20	13
FourSquare	29	7	24	19	8	2 932	12	4	31	20	7	16	27	12	19	22	13	6	17	16
Groupon	8	2	12	11	2	16	1 069	3	5	11	6	8	16	2	13	12	7	6	9	6
Messenger	4	2	2	3	84	2	1	1 309	6	5	3	6	3	6	5	8	6	3	6	3
OneFootball	48	38	21	39	33	32	18	11	4 252	38	4	29	32	10	36	40	37	12	28	15
Pinterest	13	15	16	12	9	30	16	4	24	2 630	10	13	25	16	27	28	20	10	20	12
Slitter.io	21	5	13	13	3	9	3	4	7	0	1 100	6	24	2	10	5	10	7	9	7
Spotify	5	7	9	11	4	13	2	1	33	10	3	2 823	11	3	13	16	14	1	11	4
Subito	31	15	12	18	4	14	14	2	26	22	14	17	2 801	10	21	13	24	3	11	8
Trello	2	6	18	11	8	14	5	0	6	6	5	5	12	1 118	10	9	5	0	11	3
TripAdvisor	11	12	17	14	7	24	18	6	16	14	10	6	6	8	1 263	19	13	6	15	11
Twitter	15	20	41	38	22	30	16	9	31	16	18	27	35	17	38	4 048	31	22	30	21
Viber	5	5	11	10	8	10	3	3	8	9	8	4	7	2	2	7	1 452	5	8	6
Waze	6	4	8	3	3	9	5	3	10	3	3	8	5	4	4	11	14	7 000	10	2
Wish	18	7	21	15	3	16	8	10	16	14	11	14	8	10	14	24	7	11	4 497	18
Youtube	14	9	19	20	8	17	5	4	12	10	8	19	5	16	4	14	10	12	17	2 532

(b)

Şekil 5.1. Tez çalışmada MIRAGE-2019 veri seti için a) özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, b) özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri



Şekil 5.2. Tez çalışmada MIRAGE-2019 veri seti için elde edilen en iyi AİKE grafikleri (yeşil çizgi: özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, kırmızı çizgi: özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak)

Çizelge 5.2, MIRAGE-2019 veri seti için tez çalışmasında elde edilen mobil ağ trafiği sınıflandırma sonuçları ile aynı veri seti kullanılarak daha önce yapılan çalışmalarda elde edilen sonuçların karşılaştırmasını göstermektedir. Geçmiş çalışmalarda karşılaştırma parametreleri ağırlıklı ortalama yerine genel ortalama alınarak hesaplanmıştır. Bu kapsamda daha anlamlı bir karşılaştırma sağlamak amacıyla genel ortalama sonuçları da hesaplanmış ve Çizelge 5.2'ye eklenmiştir. Tez çalışmasında elde edilen sonuçların Hussey, Stone ve Camp tarafından [18] yapılan çalışmada elde edilen sonuçlardan daha yüksek, Nascita ve diğerleri tarafından [17] yapılan çalışmada elde edilen sonuçlara ise genel olarak denk olduğu görülmektedir. Ancak, Hussey, Stone ve Camp [18] tarafından yapılan çalışmada sınıflandırılan mobil uygulama sayısı tez çalışmasında kullanılan mobil uygulama sayısında daha düşüktür. Nascita ve diğerleri [17] tarafından gerçekleştirilen çalışmada ise veri setindeki ağ trafiğinin sadece belirli bir kısmı kullanılmıştır. Bu sebeple tez çalışmasında elde edilen sonuçların daha geçerli olduğu açıkça ifade edilebilir.

Çizelge 5.2. Tez çalışmasında MIRAGE-2019 veri seti için elde edilen sonuçlar ile geçmiş çalışmalarda elde edilen sonuçların karşılaştırılması

Çalışma	Yöntem	Sonuçlar
Nascita ve diğerleri [17]	Açıklanabilir Yapay Zeka	Ortalama Doğruluk: 0,9214 Ortalama F1-Skoru: 0,9147 Ortalama G-Ortalama: 0,9536
Hussey, Stone ve Camp [18]	Aşırı Gradyan Güçlendirme	Ortalama Doğruluk: 0,8013 Ortalama F1-Skoru: 0,7791 Ortalama AİKE-AKA: 0,9787
Tez çalışması	Makine Öğrenmesi, Derin Öğrenme ve Özellik Seçme	Ağırlıklı Duyarlılık: 0,9191 Ağırlıklı Özgüllük: 0,9952 Ağırlıklı Doğruluk: 0,9914 Ağırlıklı F1-Skoru: 0,9190 Ağırlıklı AİKE-AKA: 0,9937 Genel Doğruluk: 0,9191
Tez çalışması	Makine Öğrenmesi, Derin Öğrenme ve Özellik Seçme	Ortalama Duyarlılık: 0,8992 Ortalama Özgüllük: 0,9957 Ortalama Doğruluk: 0,9919 Ortalama F1-Skoru: 0,9022 Ortalama AİKEA-AKA: 0,9921 Genel Doğruluk: 0,9191

5.2. MIRAGE-COVID-CCMA-2022 Veri Seti Sonuçlarının Değerlendirilmesi

MIRAGE-COVID-CCMA-2022 veri seti için makine öğrenmesi ve derin öğrenme sınıflandırıcılar kullanılarak elde edilen mobil ağ trafiği sınıflandırma sonuçları daha önce sırasıyla Çizelge 4.11 ile Çizelge 4.14 arasında ve Çizelge 4.15 ile Çizelge 4.18 arasında verilmiştir. Bu bölümde en başarılı sınıflandırma sonuçları özetlenmiş ve tartışılmıştır. Bu kapsamda özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonrası elde edilen en yüksek sınıflandırma sonuçları ve karşılaştırma parametrelerin değişim miktarları Çizelge 5.3’de verilmiştir. Çizelge 5.3’de takibin kolay yapılabilmesi amacıyla her sınıflandırıcı için parametrelerdeki değişim miktarları yeşil renk ile belirtilmiştir. Ayrıca, sınıflandırmanın geneli için özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra elde edilen en yüksek sonuçlar kırmızı renk ile verilmiştir.

Çizelge 5.3 incelendiğinde, özellik seçimi yapılmasıyla birlikte hemen hemen tüm sınıflandırıcılar için karşılaştırma parametrelerinde iyileşme sağlandığı görülmektedir. Bu kapsamda ağırlıklı duyarlılık parametresi %0,0398 ile %4,1098 arasında değişen oranlarda iyileşme göstermiştir. ESA sınıflandırıcı için ağırlıklı özgüllük, doğruluk ve F1-Skoru parametresinde düşüş olmakla birlikte diğer sınıflandırıcılar için ağırlıklı özgüllük parametresinde % 0,1025 ile % 1,6689 arasında, ağırlıklı doğruluk parametresinde % 0,1280 ile % 1,4048 arasında ve ağırlıklı F1-Skoru parametresinde % 0,4774 ile % 4,2512 arasında iyileşme sağlanmıştır. Benzer şekilde, ağırlıklı AİKE-AKA parametresi % 0,1012 ile % 1,2680 arasında ve genel doğruluk parametresi %0,0398 ile %4,1098 arasında artış olmuştur.

Çizelge 5.3 incelendiğinde, özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra en yüksek sınıflandırma sonuçlarının RO sınıflandırıcı kullanılarak elde edildiği görülmektedir. Sınıflandırma başarısı açısından RO sınıflandırıcıyı DVM ve k-EYK sınıflandırıcılar takip etmektedir. En başarısız sınıflandırıcının ise AG olduğu görülmektedir. Sınıflandırma başarısı açısından derin öğrenme sınıflandırıcılarının başarısı, k-EYK ve AG sınıflandırıcıların arasında sıralanmaktadır.

Çizelge 5.3. MIRAGE-COVID-CCMA-2022 veri seti için elde edilen en iyi sonuçların özeti

Sınıflandırıcı	Aşama/Değişim Miktarı	Özellik Sayısı	Ağırlıklı					Genel Doğruluk
			Duyarluluk	Özgüllük	Doğruluk	F1-Skoru	AİKE-AKA	
DVM	Özellik Seçiminden Önce	51	0,8695	0,9760	0,9678	0,8704	0,9549	0,8695
	Özellik Seçiminden Sonra	51	0,8750	0,9781	0,9691	0,8751	0,9582	0,8750
	Değişim Miktarı (%)		0,5500	0,2124	0,1280	0,4774	0,3325	0,5500
k-EYK	Özellik Seçiminden Önce	51	0,8579	0,9738	0,9601	0,8564	0,9765	0,8579
	Özellik Seçiminden Sonra	51	0,8990	0,9788	0,9742	0,8989	0,9861	0,8990
	Değişim Miktarı (%)		4,1098	0,5039	1,4048	4,2512	0,9657	4,1098
AG	Özellik Seçiminden Önce	51	0,7065	0,8198	0,8746	0,6685	0,9074	0,7065
	Özellik Seçiminden Sonra	51	0,7161	0,8358	0,8837	0,6855	0,9200	0,7161
	Değişim Miktarı (%)		0,9653	1,6037	0,9179	1,7049	1,2680	0,9653
RO	Özellik Seçiminden Önce	51	0,9379	0,9875	0,9838	0,9377	0,9968	0,9379
	Özellik Seçiminden Sonra	51	0,9516	0,9925	0,9888	0,9517	0,9978	0,9516
	Değişim Miktarı (%)		1,3731	0,5040	0,4991	1,3955	0,1012	1,3731
ESA	Özellik Seçiminden Önce	51	0,8460	0,9692	0,9604	0,8448	0,9794	0,8460
	Özellik Seçiminden Sonra	51	0,8464	0,9502	0,9543	0,8430	0,9813	0,8464
	Değişim Miktarı (%)		0,0398	-1,9036	-0,6035	-0,1870	0,1948	0,0398
UKSB	Özellik Seçiminden Önce	51	0,8205	0,9491	0,9474	0,8156	0,9723	0,8205
	Özellik Seçiminden Sonra	51	0,8302	0,9658	0,9563	0,8280	0,9750	0,8302
	Değişim Miktarı (%)		0,9691	1,6689	0,8889	1,2414	0,2691	0,9691
ÇYUKSB	Özellik Seçiminden Önce	51	0,8290	0,9598	0,9532	0,8254	0,9748	0,8290
	Özellik Seçiminden Sonra	51	0,8457	0,9608	0,9574	0,8438	0,9796	0,8457
	Değişim Miktarı (%)		1,6671	0,1025	0,4156	1,8375	0,4802	1,6671
KTB	Özellik Seçiminden Önce	51	0,8316	0,9668	0,9558	0,8293	0,9758	0,8316
	Özellik Seçiminden Sonra	51	0,8561	0,9687	0,9628	0,8541	0,9820	0,8561
	Değişim Miktarı (%)		2,4560	0,1815	0,7023	2,4775	0,6254	2,4560

Makine öğrenmesi ve derin öğrenme sınıflandırıcıların çalışma süreleri karşılaştırıldığında, k-EYK, AG ve RO en hızlı sınıflandırıcılar olarak sıralanırken DVM ve ESA en yavaş sınıflandırıcılar olarak görülmektedir. Diğer derin öğrenme sınıflandırıcılar ise DVM ve ESA'dan daha hızlı ve diğer makine öğrenimi sınıflandırıcılarından daha yavaş çalışmaktadır. Söz konusu bu çalışma süreleri ağ trafiği başına eğitim ve test dahil sınıflandırma sürelerini yansıtmaktadır. Ancak, bu çalışma sürelerinin önemli bir kısmının eğitim süreçlerini kapsadığı hususunun dikkate alınması gerekmektedir.

Sınıflandırma sonuçlarına etkisi açısından özellik seçim yöntemlerinin başarısı ve çalışma süreleri kendi aralarında karşılaştırıldığında, MFMU yönteminin k-EYK yöntemine göre hem daha hızlı hem de daha başarılı sonuçların elde edilmesini sağladığı görülmektedir.

Ayrıca, MFMU yöntemi k-EYK yöntemine benzer şekilde RO yönteminden daha hızlıdır. Genel olarak MFMU ile RO özellik seçme yöntemlerinin sınıflandırma başarısına etkilerinin birbirlerine yakın olduğu görülmektedir.

Çizelge 5.3 incelendiğinde, özellik seçimi yapılmadan önce sınıflandırma sonucunda elde edilen en yüksek ağırlıklı duyarlılık, özgüllük, doğruluk, F1-Skoru, AİKE-AKA ve genel doğruluk parametresinin sırasıyla 0,9379; 0,9875; 0,9838; 0,9377; 0,9968 ve 0,9379 olduğu görülmektedir. Özellik seçimi yapıldıktan sonra bu parametreler sırasıyla 0,9516; 0,9925; 0,9888; 0,9517; 0,9978 ve 0,9516 seviyesine yükselmiştir. Şekil 5.3’de özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri verilmiştir. Şekil 5.4’de ise, en yüksek sayıda ağ trafiğine sahip ilk sekiz mobil uygulama için özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonraki AİKE grafiklerinin durumu yer almaktadır.

	Discord	GotoMeeting	Meet	Messenger	Skype	Slack	Teams	Webex	Zoom
Discord	3 312	58	46	99	74	96	49	42	76
GotoMeeting	24	1 494	37	23	24	21	29	30	30
Meet	33	24	1 677	43	30	24	28	24	31
Messenger	105	71	43	2 763	73	89	87	48	71
Skype	27	37	30	52	26 029	35	165	34	40
Slack	24	15	10	34	17	922	21	12	17
Teams	41	81	38	65	168	32	6 814	53	73
Webex	10	29	10	12	25	10	11	2 230	19
Zoom	60	66	37	67	85	42	48	40	4 212

(a)

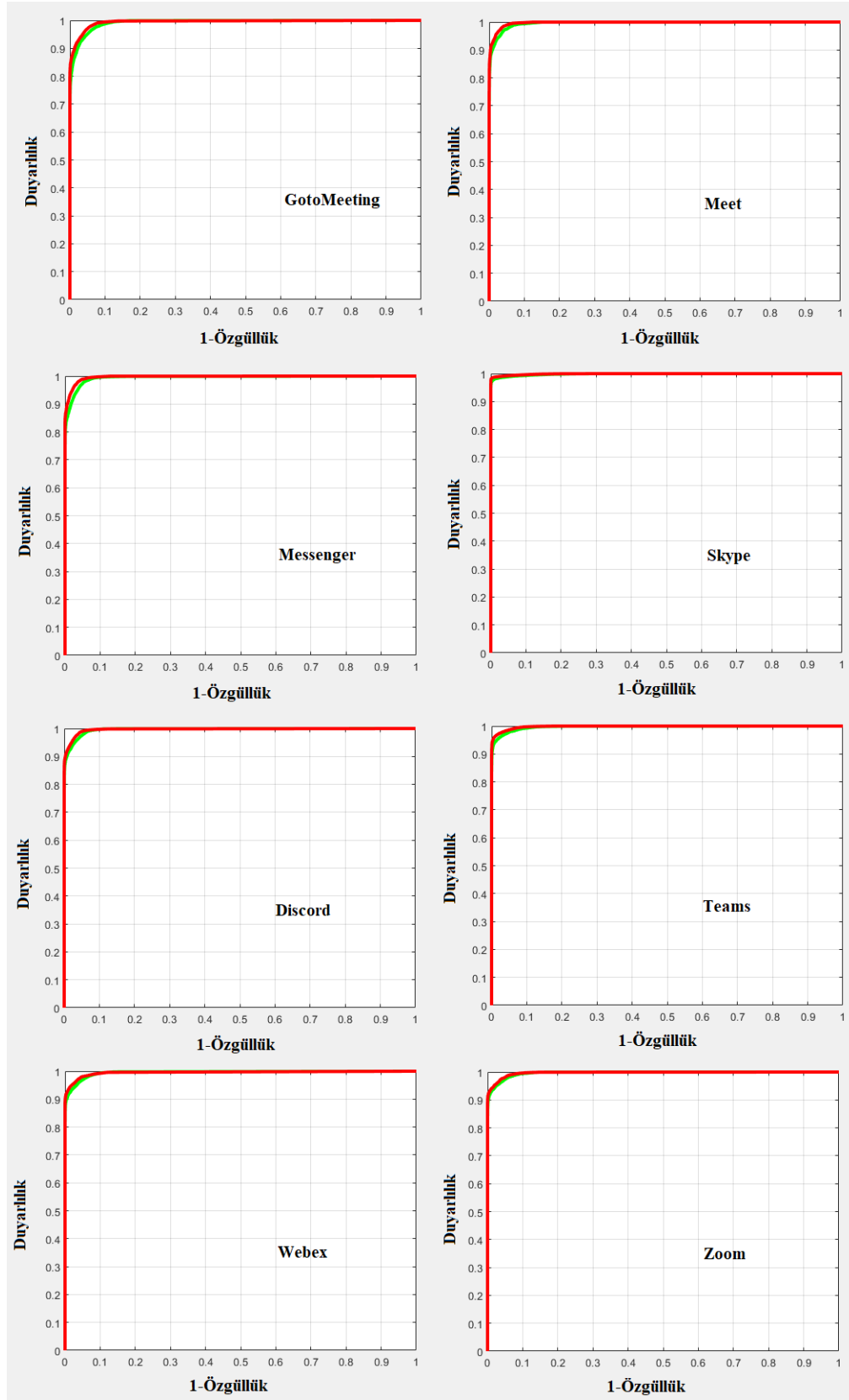
	Discord	GotoMeeting	Meet	Messenger	Skype	Slack	Teams	Webex	Zoom
Discord	3 329	40	27	73	57	89	42	30	68
GotoMeeting	14	1 595	24	20	19	17	19	18	24
Meet	39	34	1 724	39	26	25	41	31	31
Messenger	103	68	32	2 871	69	96	55	54	61
Skype	12	16	21	22	26 177	17	74	17	29
Slack	37	23	12	40	25	978	25	13	19
Teams	38	54	38	32	94	21	6 933	46	66
Webex	10	17	19	8	10	7	10	2 276	18
Zoom	54	28	31	53	48	21	53	28	4 253

(b)

	Discord	GotoMeeting	Meet	Messenger	Skype	Slack	Teams	Webex	Zoom
Discord	3 355	62	35	91	69	91	46	37	83
GotoMeeting	19	1 590	29	13	17	13	17	21	23
Meet	32	26	1 726	29	24	17	31	22	30
Messenger	97	59	30	2 875	69	100	60	38	60
Skype	21	13	29	24	26 164	15	70	15	23
Slack	29	24	12	35	18	984	11	18	14
Teams	27	55	22	30	101	20	6 954	49	72
Webex	8	16	12	7	11	9	11	2 274	9
Zoom	48	30	33	54	52	22	52	39	4 255

(c)

Şekil 5.3. Tez çalışmada MIRAGE-COVID-CCMA-2022 veri seti için a) özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, b) özellik seçimi yapıldıktan sonra (MFMU) ve RO sınıflandırıcı kullanılarak, c) özellik seçimi yapıldıktan sonra (RO) ve RO sınıflandırıcı kullanılarak elde edilen en yüksek sınıflandırma sonuçlarına ilişkin karışıklık matrisleri



Şekil 5.4. Tez çalışmada MIRAGE-COVID-CCMA-2022 veri seti için elde edilen en iyi AİKE grafikleri (yeşil çizgi: özellik seçimi yapılmadan önce ve RO sınıflandırıcı kullanılarak, kırmızı çizgi: özellik seçimi yapıldıktan sonra (RO) ve RO sınıflandırıcı kullanılarak)

MIRAGE-COVID-CCMA-2022 veri seti yakın zamanda yayınlandığı için hali hazırda literatürde bu veri setini kullanan çok az sayıda sınıflandırma çalışması bulunmaktadır. Veri setinin oluşturan çalışma ekibi (Guarino ve diğerleri) ilk sınıflandırma uygulamasını da bu çalışmayla birlikte gerçekleştirmiştir [14]. Bu kapsamda bir boyutlu ESA, hibrit ve taklitçi-geliştirilmiş olmak üzere üç farklı yöntem kullanılarak mobil ağ trafiği ve aktivite sınıflandırması gerçekleştirilmiştir. Bu çalışmada doğrudan veri paketlerinin kendisinden faydalanılmıştır. Çalışma sonucunda mobil ağ trafiği sınıflandırması için 0,94 ile 0,99 arasında değişen doğruluk ve 0,95 ile 0,99 arasında değişen F1-Skoru başarıları elde edilmiştir. Ancak, söz konusu çalışmada akış özellikleri yerine doğrudan veri paketleri kullanılmıştır. Tez çalışmamızda ise akış özelliklerinden yararlanılarak sınıflandırma gerçekleştirilmiştir. Bu nedenle çalışma sonuçlarını sayısal olarak birbirleriyle doğrudan karşılaştırmak mümkün değildir. Ancak, hem bahsi geçen çalışmada doğrudan veri paketlerinin kullanılması hem de daha yavaş (örneğin ESA) sınıflandırıcılar kullanılması sebebiyle tez çalışmasının çalışma süresi açısından daha avantajlı olduğu açık bir şekilde söylenebilir.



6. SONUÇ VE ÖNERİLER

Tez çalışmasında, MIRAGE-2019 ve MIRAGE-COVID-CCMA-2022 veri seti içinde yer alan mobil uygulamaların ağ trafiği akış özelliklerine dayalı olarak makine öğrenmesi ve derin öğrenme sınıflandırıcılar kullanılarak otomatik sınıflandırılması üzerine kapsamlı deneyler gerçekleştirilmiştir. Bu kapsamda makine öğrenmesi olarak DVM, k-EYK, AG ve RO sınıflandırıcılar kullanılırken, derin öğrenme olarak ise ESA, UKSB ÇYUKSB ve KTB sınıflandırıcılar kullanılmıştır. Deneyler, paket uzunluğu ve varış süresine ait alt özellik grupları (çift yönlü akış, yukarı yönlü akış, aşağı yönlü akış) için ayrı ayrı gerçekleştirilmiştir. Ayrıca, paket uzunluğu ve varış süresi alt özellik grupları kendi içinde bir bütün olarak kullanılarak deneyler tekrarlanmıştır. Son olarak, bu özelliklerin tamamının birleştirilmesiyle elde edilen 102 adet özellikten MFMU, k-EYK ve RO olmak üzere üç farklı özellik seçim yöntemi kullanılarak 51 tanesi seçilmiş ve sınıflandırmada kullanılmıştır. Sınıflandırma sonuçlarının hesaplanmasının yanı sıra özellik seçme ve sınıflandırma aşamalarına ilişkin çalışma süreleri de ölçülmüştür. Sonuç olarak, tez çalışmasında çok kapsamlı karşılaştırma ve tartışma yapılmasına olanak sağlayacak seviyede deneyler gerçekleştirilmiştir.

Tez çalışması sonucunda elde edilen önemli çıkarımlardan ilki ara işlem basamağı olarak özellik seçme kullanılmasının sınıflandırma sonuçlarını olumlu yönde etkilediğidir. Bu kapsamda, MIRAGE-2019 veri seti için AG sınıflandırıcı ve MIRAGE-COVID-CCMA-2022 veri seti için ESA sınıflandırıcı özelinde bazı karşılaştırma parametrelerinde istisnalar olmakla birlikte her iki veri seti için de diğer sınıflandırıcılar ve karşılaştırma parametreleri açısından yükselme sağlanmıştır. Ayrıca, her iki veri seti için de elde edilen en iyi sınıflandırma sonuçları özellik seçme yapılarak daha da yükseltilmiştir. Sonuç olarak, tez çalışmasında elde edilen en yüksek sınıflandırma sonuçlarında özellik seçme yöntemleri kullanılmasının önemli katkıları vardır.

Tez çalışmasında kullanılan MFMU, k-EYK ve RO özellik seçme yöntemlerinin sınıflandırma sonuçlarına etkisi açısından başarıları karşılaştırıldığında MFMU ve RO yöntemlerinin başarılarının genel olarak birbirine yakın ve k-EYK'den daha yüksek olduğu görülmektedir. Çalışma süresi açısından karşılaştırıldığında ise MFMU yönteminin RO ve K-EYK'e göre çok daha hızlı olduğu anlaşılmaktadır. Bu sebeple tez çalışması

sonucunda performans ve çalışma hızı açısından kullanımı en optimal özellik seçme yönteminin MFMU olduğu sonucuna varılmıştır.

Tez çalışmasında elde edilen sınıflandırma sonuçlarına göre makine öğrenmesi ve derin öğrenme sınıflandırıcıların performansları kendi aralarında karşılaştırıldığında makine öğrenmesi sınıflandırıcıların derin öğrenme sınıflandırıcılarına göre daha başarılı olduğu görülmektedir. Bunun temel sebebi tez çalışmasında ham veri paketlerinin doğrudan kendisinin kullanılması yerine bu paketlerden elde edilmiş akış özelliklerinin kullanılması olarak görülebilir. Ayrıca, çalışma süresi açısından da makine öğrenmesi sınıflandırıcıları derin öğrenme sınıflandırıcılara göre daha hızlıdır. Sonuç olarak her iki veri seti içinde özellik seçimi yapılmadan önce ve özellik seçimi yapıldıktan sonra en yüksek sınıflandırma sonuçları RO sınıflandırıcı kullanılarak sağlanmıştır. Ayrıca, RO sınıflandırıcı tez çalışmasında kullanılan en hızlı sınıflandırıcılar arasında yer almaktadır.

Tez çalışması kapsamında gerçekleştirilen bütün deneylerin sonuçları dikkate alındığından gerek sınıflandırma başarısı gerekse çalışma süresi açısından en optimal sınıflandırma modelinin MFMU kullanılarak seçilen akış özelliklerinin RO sınıflandırıcı kullanılarak sınıflandırılması olduğu sonucuna varılmıştır.

Tez çalışmasında elde edilen sonuçların iyileştirilmesi için mobil ağ trafiği sınıflandırması ile aktivite sınıflandırmasının bir bütün olarak ele alınmasının yararlı olabileceği değerlendirilmektedir. Ayrıca, çalışma süresinde yavaşlamaya sebep olması muhtemel olmakla birlikte akış özelliklerini doğrudan kullanmak yerine ham veri paketleri üzerinden derin öğrenme yöntemleri kullanılarak elde edilen derin özelliklerin makine öğrenmesi sınıflandırıcılar kullanılarak sınıflandırılması önerilmektedir. Özellik seçme aşamasında yeni yöntemler kullanılmasının sınıflandırma sonuçlarına olumlu katkılar sağlayabileceği değerlendirilmektedir. Bir diğer önemli öneri ise makine öğrenmesi ve derin öğrenme sınıflandırıcılara ilave olarak takviyeli öğrenme sınıflandırıcıların deneylere dahil edilmesidir.

KAYNAKLAR

1. Al Neyadi, E., Al Shehhi, S., Al Shehhi, A., Al Hashimi, N., Mohammad, Q. H. ve Alrabaee, S. (2020). Discovering public Wi-Fi vulnerabilities using raspberry pi and Kali Linux. *12th Annual Undergraduate Research Conference on Applied Computing (URC)*, 1-4
2. Kemp, S. (2021). Digital 2021: Global overview report. URL: <https://datareportal.com/reports/digital-2021-global-overview-report>, Son Erişim Tarihi: 06.06.2023.
3. Kemp, S. (2023). Digital 2023: Global overview report. URL: <https://datareportal.com/reports/digital-2023-global-overview-report>, Son Erişim Tarihi: 06.06.2023.
4. Ericsson S. (2022). *Ericsson mobility report: November 2022*. Stockholm: Ericsson, 5-34. URL: <https://www.ericsson.com/4ae28d/assets/local/reports-papers/mobility-report/documents/2022/ericsson-mobility-report-november-2022.pdf>, Son Erişim Tarihi: 06.06.2023.
5. Wang, W., Zhu, M., Wang, J., Zeng, X. ve Yang, Z. (2017). End-to-end encrypted traffic classification with one-dimensional convolution neural networks. *International Conference on Intelligence and Security Informatics (ISI)*, 43-48.
6. Alam, F., Kashef, R. ve Jaseemuddin, M. (2021). Enhancing the performance of network traffic classification methods using efficient feature selection models. *International Systems Conference (SysCon)*, 1-6.
7. Draper-Gil, G., Lashkari, A. H., Mamun, M. S. I. ve Ghorbani, A. A. (2016). Characterization of encrypted and vpn traffic using time-related. *2nd International Conference on Information Systems Security and Privacy (ICISSP)*, 407-414.
8. Wang, W., Zhu, M., Zeng, X., Ye, X. ve Sheng, Y. (2017). Malware traffic classification using convolutional neural network for representation learning. *International Conference on Information Networking (ICOIN)*, 712-717.
9. Li, D., Zhu, Y. ve Lin, W. (2017). Traffic identification of mobile apps based on variational autoencoder network. *13th International Conference on Computational Intelligence and Security (CIS)*, 287-291.
10. Rojas, J. S., Rendon, A. ve Corrales, J. C. (2019). Consumption behavior analysis of over the top services: Incremental learning or traditional methods?. *IEEE Access*, 7, 136581-136591.
11. Wang, R., Liu, Z., Cai, Y., Tang, D., Yang, J. ve Yang, Z. (2018). Benchmark data for mobile app traffic research. *15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 402-411.

12. Aceto, G., Ciuonzo, D., Montieri, A., Persico, V. ve Pescapé, A. (2019). MIRAGE: Mobile-app traffic capture and ground-truth creation. *4th International Conference on Computing, Communications and Security (ICCCS)*, 1-8.
13. Heng, Y., Chandrasekhar, V. ve Andrews, J. G. (2021). UTMobileNetTraffic2021: A labeled public network traffic dataset. *IEEE Networking Letters*, 3(3), 156-160.
14. Guarino, I., Aceto, G., Ciuonzo, D., Montieri, A., Persico, V. ve Pescapé, A. (2022). Contextual counters and multimodal deep learning for activity-level traffic classification of mobile communication apps during COVID-19 pandemic. *Computer Networks*, 219, 109452.
15. Aceto, G., Bovenzi, G., Ciuonzo, D., Montieri, A., Persico, V. ve Pescapé, A. (2021). Characterization and prediction of mobile-app traffic using Markov modeling. *IEEE Transactions on Network and Service Management*, 18(1), 907-925.
16. Montieri, A., Bovenzi, G., Aceto, G., Ciuonzo, D., Persico, V. ve Pescapé, A. (2021). Packet-level prediction of mobile-app traffic using multitask deep learning. *Computer Networks*, 200, 108529.
17. Nascita, A., Montieri, A., Aceto, G., Ciuonzo, D., Persico, V. ve Pescapé, A. (2021). XAI meets mobile traffic classification: Understanding and improving multimodal deep learning architectures. *IEEE Transactions on Network and Service Management*, 18(4), 4225-4246.
18. Hussey, J., Stone, K. ve Camp, T. (2022). Positive and unlabeled learning for mobile application traffic classification. *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, 25-30.
19. İnternet: MIRAGE-2019 dataset, URL: <https://traffic.comics.unina.it/mirage/mirage-2019.html>, Son Erişim Tarihi: 06.06.2023.
20. İnternet: MIRAGE-COVID-CCMA-2022 dataset, URL: <https://traffic.comics.unina.it/mirage/mirage-covid-ccma-2022.html>, Son Erişim Tarihi: 06.06.2023.
21. Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
22. Fix, E. ve Hodges, J. L. (1951). *Discriminatory analysis, nonparametric discrimination: Consistency properties*, Texas: USAF School of Aviation Medicine, Randolph Field.
23. Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 11.
24. Freund, Y. ve Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139.
25. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

26. Breiman, L. (1994). *Bagging predictors*. California: Department of Statistics University of California Berkeley.
27. Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844.
28. Atlas, L., Homma, T. ve Marks, R. (1987). An artificial neural network for spatio-temporal bipolar patterns: Application to phoneme classification. *Neural Information Processing Systems*.
29. Yaşar, H. (2022). *COVID-19 hastalığının derin öğrenme yöntemleri kullanılarak tespiti*, Doktora Tezi, Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Konya.
30. Hochreiter, S. ve Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
31. Elmaz, F., Eyckerman, R., Casteels, W., Latré, S. ve Hellinckx, P. (2021). CNN-LSTM architecture for predictive indoor temperature modeling. *Building and Environment*, 206, 108327.
32. Cui, Z., Ke, R., Pu, Z. ve Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, 118, 102674.
33. Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *IEEE Access*, 7, 51522-51532.
34. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. ve Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv,1406.1078*.
35. Loye, G. (2023). Gated recurrent unit (GRU) with PyTorch, URL: <https://blog.floydhub.com/gru-with-pytorch/>, Son Erişim Tarihi: 06.06.2023.
36. Radovic, M., Ghalwash, M., Filipovic, N. ve Obradovic, Z. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 1-14.
37. Fang, H., Tang, P. ve Si, H. (2020). Feature selections using minimal redundancy maximal relevance algorithm for human activity recognition in smart home environments. *Journal of Healthcare Engineering*, 1-13.



Gazili olmak ayrıcalıktır