T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ (%30 İNGİLİZCE)
BİLİM DALI

TOPLUM GÜVENLİĞİ VE KİŞİSEL MAHREMİYET İÇİN İNSANSIZ HAVA ARACI ANOMALİ TESPİTİ

YÜKSEK LİSANS TEZİ

Tansel ÖZTÜRK

İstanbul
Ocak 2024

T.C.
İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDSİLİĞİ ANABİLİM DALI
BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ (%30 İNGİLİZCE)
BİLİM DALI

TOPLUM GÜVENLİĞİ VE KİŞİSEL MAHREMİYET İÇİN İNSANSIZ HAVA ARACI ANOMALİ TESPİTİ

YÜKSEK LİSANS TEZİ

Tansel ÖZTÜRK

Tez Danışmanı
Dr. Öğr. Üyesi Şengül BAYRAK HAYTA

İstanbul
Ocak 2024

TEZ ONAY

Lisansüstü Eğitim Enstitüsü Müdürlüğüne,

Bu çalışma, jürimiz tarafından Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Bilimleri ve Mühendisliği (%30 İngilizce) Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman Dr. Öğr. Üyesi Şengül BAYRAK HAYTA

Üye Dr. Öğr. Üyesi Kevser Nur ÇOĞALMIŞ

Üye Doç. Dr. Eylem YÜCEL DEMİREL

Onay

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Erhan İÇENER

Enstitü Müdürü

BİLİMSEL ETİK BİLDİRİMİ

Yüksek lisans tezi olarak hazırladığım **"Toplum Güvenliği ve Kişisel Mahremiyet için İnsansız Hava Aracı Anomali Tespiti"** adlı çalışmanın öneri aşamasından sonuçlandığı aşamaya kadar geçen süreçte bilimsel etiğe ve akademik kurallara özenle uyduğumu, tez içindeki tüm bilgileri bilimsel ahlak ve gelenek çerçevesinde elde ettiğimi, tez yazım kurallarına uygun olarak hazırladığımı, bu çalışmamda doğrudan veya dolaylı olarak yaptığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu beyan ederim.

Tansel ÖZTÜRK

ÖN SÖZ

Araştırmamdaki her aşamada bana yardımcı olan değerli tez danışmanım Dr. Öğr. Üyesi Şengül BAYRAK HAYTA'ya, yüksek lisans eğitimim boyunca benden desteklerini esirgemeyen eşim Nermin Öztürk'e ve aileme teşekkürlerimi sunarım.

Tansel ÖZTÜRK
İstanbul – 2023

ÖZET

TOPLUM GÜVENLİĞİ VE KİŞİSEL MAHREMİYET İÇİN İNSANSIZ HAVA ARACI ANOMALİ TESPİTİ

Tansel ÖZTÜRK

Yüksek Lisans, Bilgisayar Bilimleri ve Mühendisliği (%30 İngilizce)

Tez danışmanı: Dr. Öğr. Üyesi Şengül BAYRAK HAYTA

Aralık – 2023, 48 + XI Sayfa

Sivil İnsansız Hava Aracı (İHA) pazarı son birkaç yılda önemli ölçüde büyümüştür. Sivil İHA'ların yaygın kullanımı, yeni istihdam yaratma ve ekonomiye olumlu katkı sağlama potansiyelinin yanı sıra, kamu güvenliği ve kişisel mahremiyet açısından da çeşitli riskler taşımaktadır. Bu riskleri azaltmak için sivil İHA'ların istilacı ve kötüye kullanımını etkili bir şekilde tespit edip tanımlamaya ihtiyaç vardır. Sivil İHA'ların kamusal ortamlarda sıklıkla kullanıldığı dikkate alındığında, fiziksel tespit yöntemleri (radar, görüş ve ses gibi) birçok durumda verimsiz hale gelebilmektedir. Bu tezde şifrelenmiş WiFi trafik veri kayıtlarından makine öğrenmesi yöntemi ile İHA tespiti yöntemleri karşılaştırmalı olarak analiz edilmiştir. Bu çalışmada, Parrot Bebop I, DBPower UDI, DJI Spark İHA'larından elde edilen çift yönlü şifrelenmiş WiFi verileri ve tek yönlü şifrelenmiş WiFi verileri analiz edilmiştir. Minimum – maksimum normalizasyon yöntem ile normaliz edilmiş veri setine minimum Artıklık Maksimum Alaka (minimum Redundancy Maximum Releavance – mRMR), RelieF, ANOVA (Analysis of Variance) özellik seçim yöntemlerine uygulanmıştır. Üç farklı özellik yönteminden ayrı ayrı anlamlı özellikler hesaplanmıştır. Anlamlı özellikler ile elde edilen veri setlerine Karar Ağacı, Destek Vektör Makinesi (DVM), k–en yakın komşu (K–Nearest Neighbour – KNN) yöntemleri uygulanarak Normal İHA ve Anormal İHA sınıflandırması yapılmıştır. Sınıflandırma başarımları 5–kat çaprazlama yöntemine göre test edilmiştir. Deneysel çalışmalar sonucunda, mRmR, RelieF ve ANOVA özellik seçimlerine uygulanan çift yönlü şifrelenmiş WiFi verisi için Karar Ağacı %100 doğrulukla en başarılı yöntem olurken tek yönlü şifrelenmiş WiFi verisi için Karar Ağacı ve DVM yöntemleri başarılı olmuştur.

Anahtar Kelimeler: İHA, kamu güvenliği, kişisel mahremiyet, özellik seçimi, sınıflandırma

ABSTRACT

UNMANNED AERIAL VEHICLE ANOMALY DETECTION FOR PUBLIC SAFETY AND PERSONAL PRIVACY

Tansel ÖZTÜRK

Master of Science, Computer Science and Engineering (30% English)

Supervisor: Asst. Prof. Dr. Şengül BAYRAK HAYTA

December – 2023, 48 + XI Pages

The civilian Unmanned Aerial Vehicle (UAV) market has grown significantly in the last few years. Civilian drones have the potential to create new jobs and contribute positively to the economy, but also pose risks to public safety and personal privacy. It is essential to effectively detect and identify the invasive and abusive use of civilian drones to mitigate these risks. The physical detection methods (such as radar, vision and sound) can become inefficient in many cases, given that civilian UAVs are frequently used in public environments. In this thesis, a comparative analysis of UAV detection methods using machine learning from encrypted WiFi traffic data records are presented. In this study, bidirectional encrypted WiFi data and unidirectional encrypted WiFi data obtained from Parrot Bebop I, DBPower UDI, DJI Spark UAVs are analyzed. Minimum Redundancy Maximum Relevance (mRMR), RelieF, ANOVA (Analysis of Variance) feature selection methods were applied to the data set normalized with the minimum–maximum normalization method. Significant features were calculated separately from three different feature methods. Decision Tree, Support Vector Machine (SVM), K–Nearest Neighbor (KNN) methods were applied to the datasets obtained with significant features to classify Normal UAVs and Abnormal UAVs. Classification performance was tested according to the 5–fold cross–validation method. As a result of the experimental studies, Decision Tree was the most successful method with 100% accuracy for bidirectional encrypted WiFi data applied to mRmR, RelieF and ANOVA feature selections, while Decision Tree and SVM methods were successful for unidirectional encrypted WiFi data.

Keywords: UAV, public security, personal privacy, feature selection, classification

İÇİNDEKİLER

TEZ ONAY ..... **i**
BİLİMSEL ETİK BİLDİRİMİ ..... **ii**
ÖN SÖZ ..... **iii**
ÖZET ..... **iv**
ABSTRACT ..... **v**
İÇİNDEKİLER ..... **vi**
TABLOLAR LİSTESİ ..... **viii**
ŞEKİLLER LİSTESİ ..... **ix**
SEMBOLLER LİSTESİ ..... **x**
KISALTMALAR LİSTESİ ..... **xi**

BİRİNCİ BÖLÜM
1. GİRİŞ **1**

İKİNCİ BÖLÜM
2. MALZEME ve YÖNTEM **4**
2.1. Veri Seti 4
2.2. Min–Maks Normalizasyon 6
2.3. Özellik Seçimi 6
2.3.1. mRMR Yöntemi 7
2.3.2. RelieF Yöntemi 7
2.3.3. ANOVA (Analysis of Variance) Yöntemi 8
2.4. Makine Öğrenmesi Yöntemleri 9
2.4.1. ID3 Yöntemi 9
2.4.2. DVM Yöntemi 10
2.4.3. KNN Yöntemi 11

ÜÇÜNCÜ BÖLÜM
3. MODELLERİN DEĞERLENDİRİLMESİ **13**
3.1. 5–kat Çaprazlama Yöntemi 13
3.2. Karışıklık Matrisi 14
3.3. ROC Grafikleri 15

DÖRDÜNCÜ BÖLÜM

4. DENEYSEL ÇALIŞMALAR 18

4.1. Çift Yönlü Şifrelenmiş WiFi Veri Seti ile Elde Edilen Deneysel Sonuçlar 19

4.1.1. mRmR Yöntem ile Elde Edilen Deneysel Sonuçlar 19

4.1.2. RelieF Yöntemi ile Elde Edilen Deneysel Sonuçlar 22

4.1.3. ANOVA ile Elde Edilen Deneysel Sonuçlar 26

4.2. Tek Yönlü Şifrelenmiş WiFi Verisinden Elde Edilen Sonuçlar 30

4.2.1. mRmR Yöntem ile Elde Edilen Deneysel Sonuçlar 31

4.2.2. RelieF Yöntem ile Elde Edilen Deneysel Sonuçlar 34

4.2.3. ANOVA ile Elde Edilen Deneysel Sonuçlar 38

4.3. Deneysel Sonuçların Yorumlanması 42

BEŞİNCİ BÖLÜM

5. SONUÇ 44

KAYNAKÇA 45

ÖZGEÇMİŞ 48

TABLOLAR LİSTESİ

Tablo 2.1: İHA'lardan elde edilen şifrelenmiş WiFi istatistiksel parametreleri 5
Tablo 3.1: 5–kat çaprazlama yöntemi ile doğrulama 13
Tablo 3.2: Karışıklık matrisi 14
Tablo 4.1: İHA'lardan elde edilmiş şifrelenmiş tek yönlü ve çift yönlü WiFi veri seti 18
Tablo 4.2: Sınıflandırma modellerine ait optimum parametreler 18
Tablo 4.3: Çift yönlü şifrelenmiş WiFi eğitim ve test verisi 19
Tablo 4.4: Çift Yönlü Trafik akış veri setinden mRmR ile seçilen özellikler 19
Tablo 4.5: mRmR yöntemi ile elde edilen veri setine ait eğitim sonuçları 20
Tablo 4.6: mRmR yöntemi ile elde edilen veri setine ait test veri seti başarımları 20
Tablo 4.7: Çift Yönlü Trafik akış veri setinden ReflieF ile seçilen özellikler 23
Tablo 4.8: RelieF yöntemi ile elde edilen veri setine ait eğitim sonuçları 23
Tablo 4.9: ReliefF yöntemi ile elde edilen veri setine ait test başarımları 24
Tablo 4.10: Çift Yönlü Trafik akış veri setinden ANOVA ile seçilen özellikler 27
Tablo 4.11: ANOVA yöntemi ile elde edilen veri setine ait eğitim sonuçları 28
Tablo 4.12: ANOVA yöntemi ile elde edilen veri setine ait test başarımları 28
Tablo 4.13: Tek yönlü şifrelenmiş WiFi eğitim ve test verisi 31
Tablo 4.14: Tek Yönlü Trafik akış veri setinden mRmR ile seçilen özellikler 31
Tablo 4.15: mRmR yöntemi ile elde edilen veri setine ait eğitim sonuçları 32
Tablo 4.16: mRmR yöntemi ile elde edilen veri setine ait test veri seti başarımları 32
Tablo 4.17: Tek Yönlü Trafik akış veri setinden RelieF ile seçilen özellikler 35
Tablo 4.18: RelieF yöntemi ile elde edilen veri setine ait eğitim sonuçları 35
Tablo 4.19: RelieF yöntemi ile elde edilen veri setine ait test başarımları 36
Tablo 4.20: Tek Yönlü Trafik akış veri setinden ANOVA ile seçilen özellikler 39
Tablo 4.21: ANOVA yöntemi ile elde edilen veri setine ait eğitim sonuçları 39
Tablo 4.22: ANOVA yöntemi ile elde edilen veri setine ait test başarımları 40

ŞEKİLLER LİSTESİ

Şekil 2.1: Tez çalışmasında kullanılan İHA'ların temsili görüntüsü 4

Şekil 2.2: ID3 yöntemi akış diyagramı 10

Şekil 2.3: İki sınıflı hiper düzlem 11

Şekil 2.4: DVM yöntemi akış diyagramı 11

Şekil 2.5: KNN yöntemi akış diyagramı 12

Şekil 3.1: Beş ayrık sınıflandırıcıyı gösteren temel bir ROC grafiği 16

Şekil 4.1: mRmR yöntemi ile elde edilen veri setine ait karışıklık matrisi 21

Şekil 4.2: mRmR yöntemi ile elde edilen veri setine ait ROC eğrileri 22

Şekil 4.3: RelieF yöntemi ile elde edilen veri setine ait karışıklık matrisi 25

Şekil 4.4: RelieF yöntemi ile elde edilen veri setine ait ROC eğrileri 26

Şekil 4.5: ANOVA yöntemi ile elde edilen veri setine ait karışıklık matrisi 29

Şekil 4.6: ANOVA yöntemi ile elde edilen veri setine ait ROC eğrileri 30

Şekil 4.7: mRmR yöntemi ile elde edilen veri setine ait karışıklık matrisi 33

Şekil 4.8: mRmR yöntemi ile elde edilen veri setine ait ROC eğrileri 34

Şekil 4.9: RelieF yöntemi ile elde edilen veri setine ait karışıklık matrisi 37

Şekil 4.10: RelieF yöntemi ile elde edilen veri setine ait ROC eğrileri 38

Şekil 4.11: ANOVA yöntemi ile elde edilen veri setine ait karışıklık matrisi 41

Şekil 4.12: ANOVA yöntemi ile elde edilen veri setine ait ROC eğrileri 42

SEMBOLLER LİSTESİ

$\bar{x}$: Ortalama

σ : Standart Sapma

γ : Çarpıklık

β : Basıklık

MS : Ortalama Kare

H : Maksimum

L : Minimum

a' : Normalize değer

I : mRmR Yönteminde Karşılıklı Bilgi

W : Relief Skoru

F : ANOVA F istatistiği

SSB : Grup ortalamaları arasındaki fark varyansı

SSW : Grup içi gözlemler arasındaki fark varyansı

D : Öğrenme

H_0 : Hiper düzlemler orta değeri

d : Öklid uzaklığı

KISALTMALAR LİSTESİ

ANOVA : Analysis of Variance

DVM : Destek Vektör Makinesi

FFT : Fast Fourier Transform

FTP : File Transfer Protocol

HTTP : Hyper Text Transfer Protocol

TCP : Transfer Control Protocol

İHA : İnsansız Hava Aracı

STFT : Short–Time Fourier Transform

mRmR : minimum Relevance Maximum Redundancy

STD : Standart Sapma

KNN : k En yakın Komşu

MAD : Medyan

MAKS : Maksimum

MİN : Minimum

TP : True Positive (Doğru Pozitif)

FP : False Positive (Yanlış Pozitif)

TN : True Negative (Doğru Negatif)

FN : False Negative (Yanlış Negatif)

ROC : Receiver Operating Characteristic

obs : Observation (Gözlem)

sec : Second (Saniye)

WiFi : Wireless Fidelity (Kablosuz Ağ)

BİRİNCİ BÖLÜM
GİRİŞ

İnsansız Hava Aracı (İHA), üzerinde insan pilot olmadan çalışan bir hava aracıdır. İHA bir insan operatör tarafından uzaktan kontrol edilebilir veya önceden programlanmış talimatlara veya Küresel Konumlama Sistemi (Global Positioning System) ve sensörler gibi yerleşik sistemlere dayalı olarak otonom olarak uçabilmektedir. İHA'lar, eğlence amaçlı ve hava fotoğrafçılığı için kullanılan askeri keşif, gözetleme, tarım, paket teslimatı, çevresel izleme, arama ve kurtarma görevleri ve daha fazlası için kullanılan daha büyük İHA'lara kadar çeşitli şekil, boyut ve işlevlere sahiptir.

Sivil İHA pazarı son birkaç yılda dikkat çekici ölçüde büyümüştür. Bu araçlar çok yönlülükleri, uzak veya tehlikeli alanlara erişim kabiliyetleri, maliyet etkinlikleri ve sektörler arasında hizmet ettikleri geniş uygulama yelpazesi nedeniyle popülerlik kazanmıştır. Yeni iş alanları oluşturma ve ekonomiye olumlu katkı potansiyeli yanında sivil İHA kullanımının yaygınlaşması, toplum güvenliği ve kişisel mahremiyet açısından da çeşitli riskler oluşturmaktadır (R. Altawy and A. M. Youssef, 2016), (Amir Alipour-Fanid, Ning Wang, Liang Zhao, 2020). Bu riskleri azaltmak için sivil İHA'ların işgalci amaçla ve kötüye kullanımının etkin bir şekilde tespit edilmesi ve tanımlanması bir ihtiyaçtır.

Literatürde, son yıllarda İHA tespitinde yapılmış çok sayıda çalışmalar mevcuttur. Messina ve arkadaşları, radardan alınan yankıyı (taranmış frekans) örneklemiş ve yüksek geçiren (high–pass) filtre ve Hızlı Fourier Dönüşümü (Fast Fourier Transform – FFT) kullanarak özellik çıkarımı ile İHA'ları tespit ederek sınıflandırmıştır (M. Messina and G. Pinelli, 2019). Zhang ve arkadaşları, radar verilerinin K–bant ve X–bant radar sensörleri tarafından ayrı ayrı toplandığı, ardından bir Kısa–Zamanlı Fourier Dönüşümünün (Short–Time Fourier Transform – STFT) yapıldığı ve son olarak İHA sınıflandırması için Destek Vektör Makinesinin (DVM) kullanıldığı çift frekanslı bir radar sınıflandırma şeması önermektedir (P. Zhang, L. Yang, G. Chen, and G. Li, 2017). Bununla birlikte, doğrudan görüş hattı gerektirdiği için metropollerde ve şehirlerde radar tabanlı tespit verimliliği yetersiz olabilmektedir (G. J. Mendis, T. Randeny, J. Wei, and A. Madanayake, 2016). Video kameralara dayalı görüntü tabanlı

İHA tespiti de, kamera ile İHA arasında doğrudan görüş hattı gerektirdiği için radar tabanlı tekniklerle aynı zayıflığı taşımaktadır (F. Gökçe, G. Üçoluk, E. ¸ Sahin, and S. Kalkan, 2015). Akustik sinyal tabanlı İHA tespiti, görüş alanı dışına çıkma problemini çözebilecek bir yöntemdir (G. J. Mendis, T. Randeny, J. Wei, and A. Madanayake, 2016). Ancak bu yöntemin de kendine has dezavantajları vardır. Birincisi, İHA'dan gelen akustik sinyal, motorlarının ürettiği ses nedeniyle oldukça gürültülü olabilmektedir (Marmaroli, Falourd, & Lissek, Apr. 2012). İkincisi, elektrikli ot biçme makineleri gibi cihazlar, İHA'lara oldukça benzer ses sinyalleri üretebilmektedir. Tekniklerin dezavantajlarının üstesinden gelmek için akustik sensör ve video kameranın birleştirilmesiyle hibrit çözümler önerilmiştir (Busset, ve diğerleri, 2015). Kablolu ve kablosuz ağlarda, genellikle istatistiksel ve makine öğrenme yaklaşımlarının bir kombinasyonunun kullanıldığı, protokol verilerinin parmak izine dayalı olarak şifrelenmiş veri akışını tanımlamak için çeşitli çalışmalar vardır. Jing ve arkadaşlarının çalışmasında üç tür ağ trafiği (HTTP, FTP ve E–posta) tanımlamak için yeni bir DVM tabanlı yöntem önerilmiştir (Jing, Yang, Cheng, Dong, & Xiong, Nov. 2011). Bu alandaki öncü çalışmalardan birini gerçekleştiren McGregor ve arkadaşları, kablolu bir ağdaki trafiği "toplu aktarım (bulk transfer)", "küçük işlemler (small transactions)" ve "çoklu işlemler (multiple transactions)" şeklinde sınıflandırma teknikleri uygulamaktadır (McGregor, Hall, Lorier, & Brunskill, 2004). Bar Yanai ve arkadaşları, bir uygulamanın davranışını, Transfer Kontrol Protokolü (Transfer Control Protocol – TCP) bağlantısının ilk birkaç paketinin boyutunun ve yönünün gözlemlenmesinden ayırt etmenin mümkün olduğunu göstermektedir. Ancak bu yöntem hem paket başlığı izleme analizini hem de ilk TCP bağlantı paketlerini yakalamayı gerektirmektedir (Bar-Yanai, Langberg, Peleg, & Roditty, 2010). Sciancalepore ve arkadaşları şifreli ağ trafiği tanımlamasını kullanarak IHA'ların uçma veya yerde durma durumunu tespit etmeyi önermektedir. İHA durumunu belirlemek için üç farklı standart ikili (binary) sınıflandırma algoritması, Trees–J48 (J48), Rastgele Orman (Random Forest) ve Sinir Ağları (Neural Networks), 3DR SOLO İHA trafik veri setine uygulanmıştır (Sciancalepore, Ibrahim, Oligeri, & Di Pietro, 2019). Literatürdeki çalışmalara göre kamu güvenliği ve kişisel mahremiyetin korunması için anormal ve normal İHA tespitinde kapsamlı çalışmaların yapılmasına ihtiyaç vardır.

Bu tezin temel amacı, İHA'ların toplum güvenliğini olumsuz etkileyecek şekilde kullanımını ve aynı zamanda özel hayatın gizliliği ihlallerini önlemek için etkili bir İHA tespiti yöntemi geliştirmektir. Tezde, İHA'ların toplum güvenliğine etkisi, Şifrelenmiş WiFi trafik verilerinin nasıl elde edildiği, İHA'ların tespiti için mevcut makine öğrenmesi yöntemlerinin incelenmesi ve eksikliklerin belirlenmesi, özgün bir İHA tespit yönteminin geliştirilmesi ve bu yöntemin ayrıntılı açıklaması, geliştirilen yöntemin gerçek dünya verileri üzerindeki performansının değerlendirilmesi, bulguların ve yöntemin önemini vurgulayan bir değerlendirmesi, gelecekte yapılabilecek yeni çalışmalara ve yeni önerilere odaklanılacaktır.

Bu amaçlar doğrultusunda, şifrelenmiş WiFi trafik verilerinin analizi ve İHA tespiti için yeni ve özgün bir yaklaşımın geliştirilmesi hedeflenmektedir. Bu çalışmada Parrot Bebop I, DBPower UDI, DJI Spark İHA'larından elde edilen çift yönlü şifrelenmiş WiFi verileri ile tek yönlü şifrelenmiş WiFi verileri analiz edilmiştir (Zhao, Unmanned Aerial Vehicle (UAV) Intrusion Detection Datasets, 2015). Ham veri olarak paket boyutu ve paketlerin varış süreleri kullanılmış ve bu verilerden çeşitli istatistiksel ölçüm hesaplamalarıyla farklı özellikler elde edilmiştir. Veri setinde İHA trafik verisinden hesaplanan özellikler olduğu gibi İHA olmayan trafik verisinden hesaplanan özellikler de mevcuttur. Tezin uygulama adımları; (i) şifrelenmiş WiFi verilerinin normalizasyonu, (ii) mRmR, RelieF, ANOVA özellik seçim yöntemleri ile anormal İHA ve normal İHA'ların ayrımında önemli özelliklerin hesaplanması, (iii) üç farklı özellik seçim yönteminden elde edilen veri setlerinin Karar Ağacı, DVM, KNN makine öğrenmesi yöntemleri ile 5–kat çaprazlama yöntemleri ile sınıflandırılması, (iv) sınıflandırma yöntemlerinin test başarımları doğruluk, özgüllük ve duyarlılık bakımından karşılaştırılması yapılmıştır. (v) Model tahminleme başarımı alıcı eğrisi (Receiver Operating Characteristic – ROC) ile değerlendirilmiştir.

İKİNCİ BÖLÜM
MALZEME ve YÖNTEM

2.1. Veri Seti

Bu çalışmada, Şekil 2.1 (a)'da Parrot Bebop I, Şekil 2.1 (b) 'de DBPower UDI, Şekil 2.1 (c)'de DJI Spark İHA'larından elde edilen veriler ile çalışılmıştır (Zhao, Unmanned Aerial Vehicle (UAV) Intrusion Detection Datasets, 2015).

a) Parrot Bebop I b) DBPower UDI c) DJI Spark

Şekil 2.1: Tez çalışmasında kullanılan İHA'ların temsili görüntüsü

Her bir İHA'dan WiFi trafik kayıtları için hem çift yönlü veri akış modunda i) yukarı bağlantı akışı, ii) aşağı bağlantı akışı ve iii) toplam trafik akışı hem de tek yönlü veri akış modunda toplam trafik akışı elde edilmiştir. Paket boyutu ve paketlerin varış süresi ham veri kaynaklarıdır. Her kaynak için

Tablo *2.1*'de verilen 9 istatistiksel ölçüm yapılmıştır (Liang Zhao, Amir Alipour-Fanid, Martin Slawski and Kai Zeng, 2018).

Tablo 2.1: İHA'lardan elde edilen şifrelenmiş WiFi istatistiksel parametreleri

Özellik ID: Adı	Tanımı
V_1: ortalama	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x(i)$
V_2: medyan	Sıralı veri örneğinin tam ortasındaki değer veya tam ortadaki iki değerin ortalaması
V_3: MAD	$MAD = medyan(\|x(i) - medyan(x)\|)$
V_4: STD	$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x(i) - ortalama(x))^2}$
V_5: Çarpıklık	$\gamma = \frac{1}{N}\sum_{i=1}^{N}(x(i) - ortalama(x)/\sigma)^3$
V_6: Basıklık	$\beta = \frac{1}{N}\sum_{i=1}^{N}(x(i) - ortalama(x)/\sigma)^4$
V_7: MAKS	$H = (Maks(x(i))\|i = 1 \dots N)$
V_8: MİN	$L = (Min(x(i))\|i = 1 \dots N)$
V_9: Ortalama Kare	$MS = \frac{1}{N}\sum_{i=1}^{N}(x(i))^2$

Çift Yönlü WiFi veri seti 9 özellik × 2 kaynak × 3 yönlü akış = 54 özellik içermektedir. Tek yönlü WiFi veri seti 9 özellik × 2 kaynak = 18 özellik içermektedir.

Çift yönlü ve tek yönlü şifrelenmiş WiFi veri setine min–maks normalizasyon yöntemi uygulanmıştır.

2.2. Min–Maks Normalizasyon

Normalizayon, veri setinden seçilen verinin 0 – 1 gibi aralıklara indirgenmesi işlemidir. Bu tezde, şifrelenmiş WiFi verisinin doğrusal şekilde normalize eden min–maks normalizasyon yöntemi kullanılmıştır (Bayrak, Yucel, & Takci, Epilepsy Radiology Reports Classification Using Deep Learning Networks, 2022). min, bir verinin alabileceği en düşük değeri ve $maks$, bir verinin alabileceği en büyük değeri temsil etmektedir. Bir verinin 0 – 1 aralığında normalizasyonu Denklem 2.1'de verilmiştir.

$$a' = \frac{a-min}{maks-min} \tag{2.1}$$

Burada:

- a veri setindeki bir örnek değerini temsil etmektedir.
- a' ölçeklendirilmiş a değerini ifade etmektedir.
- min değeri veri setindeki minimum değeri temsil etmektedir.
- $maks$ değeri veri setindeki maksimum değeri temsil etmektedir.

2.3. Özellik Seçimi

Özellik seçimi, veri madenciliği ve makine öğrenmesinde verimliliği ve doğruluğu artırmada önemlidir. Bu tezde, üç özellik seçimi yöntemi olan mRMR, RelieF ve ANOVA özellik seçim yöntemleri kullanılmıştır. mRMR yöntemi, genellikle yüksek boyutlu veriler için tercih edilmektedir. Özellikler arası bağımlılıkları ve yinelemeyi dikkate alarak hesaplama yapmaktadır. RelieF yöntemi, özellikle etkileşimli ve bağımlı özelliklerin olduğu veri setlerinde kullanılmaktadır. ANOVA yöntemi veri setindeki sınıflar arası verilerin ilişkisini istatistiksel olarak hesaplamaktadır.

2.3.1. mRMR Yöntemi

mRMR (minimum Redundancy Maximum Relevance) yöntemi, özellik seçimi veya özellik sıralama için kullanılan bir tekniktir. Bu yöntem sayesinde, bir veri setindeki özelliklerin önem sırası hesaplanmaktadır. mRMR, veri setindeki özelliklerin hedef değişkeniyle (maximum relevance) ilişkili olmasını ve birbirleriyle düşük derecede ilişkili olmasını (minimum redundancy) sağlamaktadır. Bu sayede, modelin genelleştirilmesini ve daha iyi performans göstermesini amaçlamaktadır.

Özelliklerin sınıflandırma performansına katkısını maksimize ederken, özellikler arasındaki yinelemeyi minimize etmektedir. Bu yöntem, her bir özelliğin sınıf etiketleriyle olan bağımlılığını ve diğer özelliklerle olan karşılıklı bilgisini hesaplamaktadır. İki değişken arasındaki karşılıklı bilgi bir değişkenin belirsizliğinin diğer değişkenin bilinmesiyle ne kadar azaltılabileceğini göstermektedir (H. Peng, F. Long and C. Ding, Aug. 2005).

$$I(x;y) = \iint p(x,\ y) \log \frac{p(x,y)}{p(x)p(y)}\, dx\, dy \qquad (2.2)$$

Denklem 2.2'de I, gereksiz ve yineleyen veriyi temsil etmektedir. Bu veriyi filtrelediği için özellikle yüksek boyutlu veri kümelerinde etkilidir (H. Peng, F. Long and C. Ding, Aug. 2005).

2.3.2. RelieF Yöntemi

RelieF algoritması, veri setindeki özelliklerin bir hedef değişkenle ilişkisini ölçmekte ve bu ilişkiyi temel alarak önemli özellikleri belirlemektedir. Özellik etkileşimine dayalı, özellik seçiminde filtreleme yaklaşımını benimseyen bir algoritmadır. RelieF, her özellik için bir puan hesaplamaktadır ve bu puan özellikleri sıralamak ve seçmek için kullanılmaktadır. RelieF özellik puanlaması, en yakın komşu örnek çiftleri arasındaki farkların tanımlanmasına dayanmaktadır. Özellik değerleri farkı gözlemlenen iki komşu örnek çifti aynı sınıfa ait ise isabet ("hit") denir ve özellik puanı düşer, farklı sınıfa ait ise kaçırılmış ("miss") denir ve özellik puanı artar (Contributors, 2023).

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \quad (2.3)$$

RelieF algoritması her bir özelliğin RelieF skorlarını hesapladıktan sonra bu skorları kullanarak özelliklerin sınıflandırma performansına olan katkısını değerlendirmektedir. Daha sonra, belirli bir eşik değeri kullanılarak önemli özellikleri seçim veya gereksiz özellikleri eleme işlemi gerçekleştirilmektedir.

2.3.3. ANOVA (Analysis of Variance) Yöntemi

ANOVA, grupların ortalamaları arasındaki farklılıkları değerlendirmek için kullanılmaktadır. ANOVA'nın özellik seçimi için kullanılması, genellikle bir öğrenme modeline girdi değişkenleri olarak kullanılan özellikler arasındaki önemli farklılıkları belirlemek amacıyla gerçekleşmektedir.

ANOVA'nın temel prensibi, gruplar arası varyansın, gruplar içindeki varyansa oranının test edilmesidir. Bu oran, F istatistiği olarak adlandırılır ve şu şekilde hesaplanmaktadır (Akyıldız, 2009), (Gajawada, 2019).

$$F = \frac{SSB/(k-1)}{SSW/(N-k)} \quad (2.4)$$

Denklem 2.4'te k grup sayısını, N toplam gözlem sayısını, SSB grup ortalamaları arasındaki farkın neden olduğu varyansı, SSW her bir grup içindeki gözlemler arasındaki farkın neden olduğu varyansı temsil etmektedir.

$$SSB = \sum_{j=1}^{k} n_j (\bar{y}_j - \bar{y})^2 \quad (2.5)$$

Denklem 2.5'te k grup sayısını, n_j her bir grup içindeki gözlem sayısını, $\bar{y}_j$ ise her bir grup ortalamasını temsil etmektedir.

$$SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2 \quad (2.6)$$

Denklem 2.6'da y_{ij} her bir gözlem değerini temsil etmektedir.

Özellik seçimi için ANOVA kullanılırken, genellikle her bir özellik için bir *F* istatistiği hesaplanır. *F* istatistiği büyük olduğunda, o özellik modelin açıklanmasında önemli olabilir ve bu nedenle seçilebilir. *F* istatistiği, özelliklerin gruplar arasında nasıl değiştiğini ve modelin başarısını ne kadar etkilediğini ölçer. *F* istatistiği, p–değeri ile birlikte değerlendirilir ve eğer p–değeri belirli bir anlamlılık düzeyinden küçükse (örneğin, 0,05), o özellik modelde kullanılabilir olarak kabul edilmektedir (Gajawada, 2019), (AJPAS, 2016).

2.4. Makine Öğrenmesi Yöntemleri

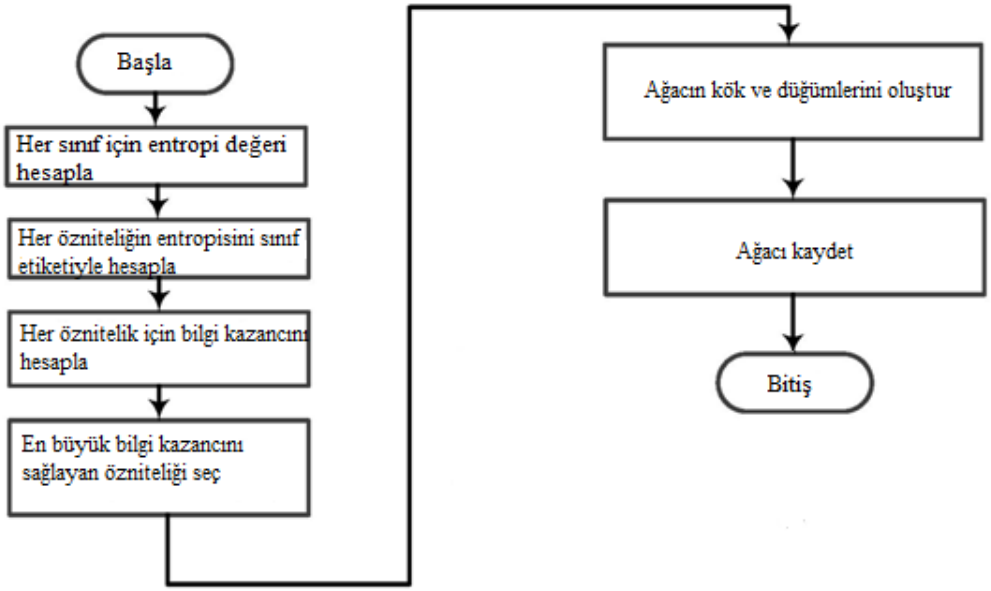
Öğreticili öğrenme, eğitim ve test veri setleri ile oluşturulmaktadır ve öğrenme işleminin matematiksel formülü Denklem 2.7'deki gibidir.

$$D = \{x_1, x_2, \dots, x_n\}; \; Y = \{y_1, y_2, \dots, y_m\} \quad (2.7)$$

2.4.1. ID3 Yöntemi

ID3 yöntemi veri setindeki her bir özellik için entropi hesaplaması yapmaktadır. Veri seti için en optimum karar ağacını bulmak amacıyla veri setinin bölünme öncesindeki ve bölünme sonrasındaki fark değerini kullanmaktadır. Bu fark hangi alt bölüm için daha büyük değere sahipse öncelikli düğüm ve dallanma, hesaplanan bu değerin büyüklüğüne göre belirlenmektedir. Aradaki bu fark kazanımdır (S) ve Denklem 2.8'e göre hesaplanmaktadır (T. Daniya, M. Geetha,& Dr, K. Suresh Kumar, 2020). Şekil 2.2'de ID3 yönteminin akış diyagramı verilmiştir.

$$Kazanım(D; S) = H(D) - \sum_{i=1}^{n} P(D_i) H(D_i) \quad (2.8)$$

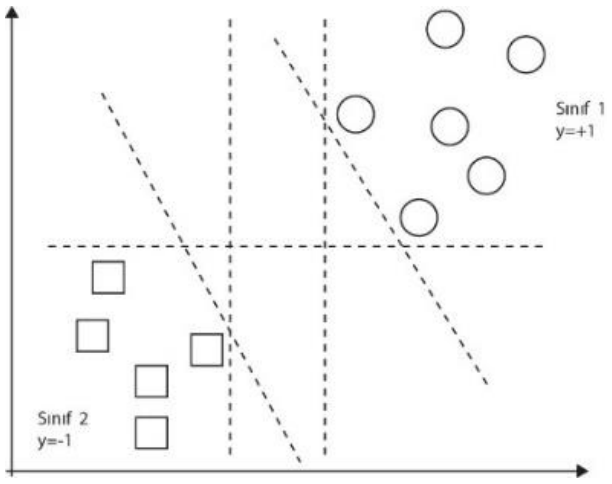


Şekil 2.2: ID3 yöntemi akış diyagramı

Kaynak: *(Bayrak S. , 2021)*

2.4.2. DVM Yöntemi

Verinin doğrusal bir fonksiyon yardımıyla en uygun hiper düzlemi tahmin ederek sınıflandırılmasıdır. Bu çalışma iki sınıflı sınıflandırma problemi olduğu için matematiksel işlemler 2–sınıflı sınıflandırmaya göre verilmiştir. n elemanlı D veri seti $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_n, y_n)$ şeklindedir ve $y \in \{+1, -1\}$ olarak kabul edilmektedir. Şekil 2.3**Hata! Başvuru kaynağı bulunamadı.**'e göre, veri setinin sınıflandırılması için birbirleri arasındaki boşluğu en büyük olan hiper düzlemler seçilmektedir (Küçüksille & Ateş, 2016).

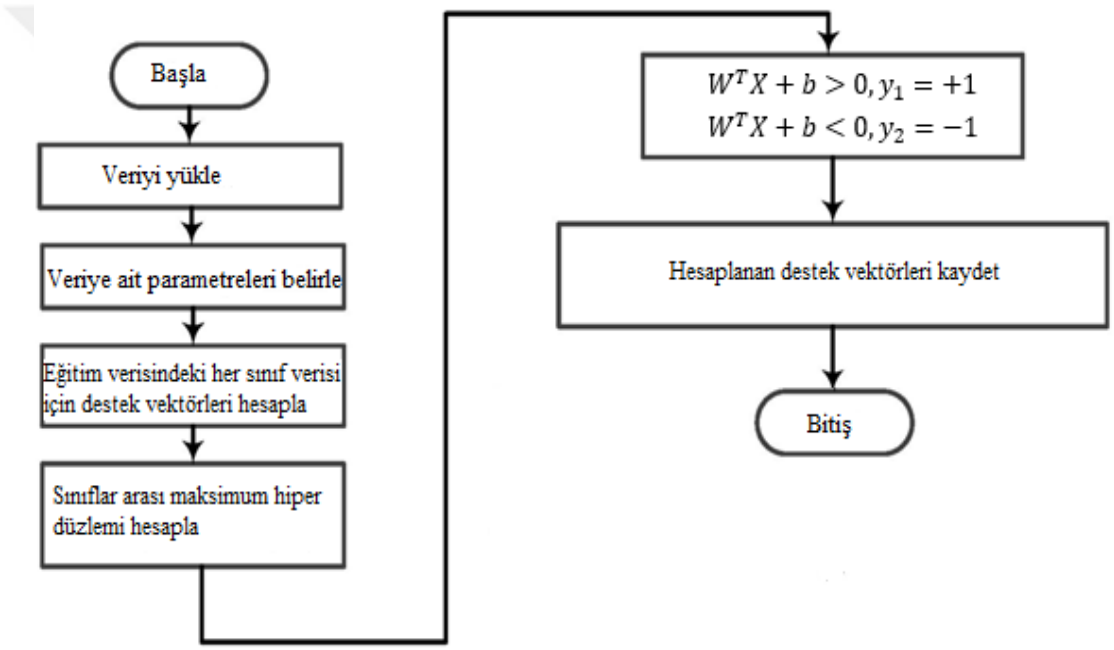


Şekil 2.3: İki sınıflı hiper düzlem

Şekil 2.3'e göre H_1 ve H_2 hiper düzlemlerinin orta değeri H_0 değeridir. H_0, denklem 2.9'daki gibi hesaplanmaktadır (Bayrak & Yucel, Methods for the Recognition of Multisource Data in Intelligent Medicine: A Review and Next-Generation Trends, 2022).

$$H_0 = \sum_{i=1}^{n} w_i x_i + b = 0 \qquad (2.9)$$

Şekil 2.4'te DVM modeline ait akış diyagramı verilmiştir.



Şekil 2.4: DVM yöntemi akış diyagramı

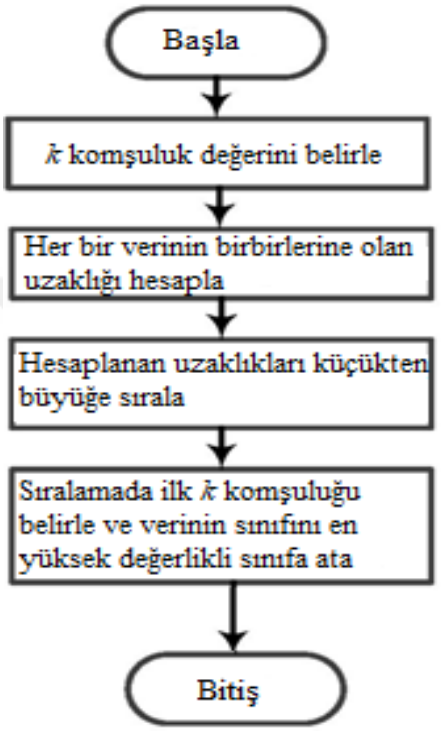
Kaynak: (Bayrak S. , 2021)

2.4.3. KNN Yöntemi

Yeni bir verinin hangi sınıfa ait olduğunu uzaklık hesabına göre belirlemektedir (Guo, Wang, Bell, Bi, & Greer, 2003). Bu çalışmada uzaklık hesabı olarak Öklid uzaklığı seçilmiştir. En küçük Öklid uzaklığına sahip k sayıdaki veri için hesaplama Denklem 2.10'a göre yapılmaktadır (Fiori, 2020).

$$d(i,j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (2.10)$$

Sınıfı bulunmak istenen verinin tüm verilere uzaklığı hesaplanmaktadır ve sıralanmaktadır. En sık tekrarlanan sınıf değeri, yeni verinin sınıfı olarak belirlenmektedir. **Şekil 2.5**'te KNN yönteminin akış diyagramı verilmiştir.



Şekil 2.5: KNN yöntemi akış diyagramı

Kaynak: (Bayrak S. , 2021)

ÜÇÜNCÜ BÖLÜM
MODELLERİN DEĞERLENDİRİLMESİ

Oluşturulan veri seti için en başarılı sınıflandırma modelinin belirlenebilmesi amacıyla 5–kat çaprazlama yöntemine göre karışıklık matrisi, iki sınıflı modelleme için doğruluk, duyarlılık, özgüllük ve kesinlik parametreleri ile karşılaştırma yapılmıştır. Modellerin başarımları ROC eğrisine göre değerlendirilmiştir.

3.1. 5–kat Çaprazlama Yöntemi

Oluşturulan veri setinin n adet eşit alt kümeye bölünüp $n-1$ sayıda kümenin eğitim işlemi için, 1 tanesinin test işlemi için ayrılmasıdır. Bu çalışmada n değeri 5 seçilmiştir. Tablo 3.1'de de gösterildiği gibi, bu işlem 5 kez tekrar etmektedir. Her defasında elde edilen doğruluk değeri ortalaması, ilgili sınıflandırma modelinin doğruluk değeridir (Bayrak, Yucel, & Takci, Classification of extracranial and intracranial EEG signals by using finite impulse response filter through ensemble learning, 2019).

Tablo 3.1: 5–kat çaprazlama yöntemi ile doğrulama

	1. adım	2. adım	3. adım	4. adım	5. adım
1. adım	Test	Eğitim	Eğitim	Eğitim	Eğitim
2. adım	Eğitim	Test	Eğitim	Eğitim	Eğitim
3. adım	Eğitim	Eğitim	Test	Eğitim	Eğitim
4. adım	Eğitim	Eğitim	Eğitim	Test	Eğitim
5. adım	Eğitim	Eğitim	Eğitim	Eğitim	Test

3.2. Karışıklık Matrisi

Sınıflandırma işlemlerinde gerçek verinin sınıfı ile öngörülen sınıf değerlerinin karşılaştırılmasını sağlamaktadır. Çalışmamızda Normal İHA ve Anormal İHA olmak üzere iki sınıflı veri setine ait karışıklık matrisi, Tablo 3.2'de verilmiştir.

Tablo 3.2: Karışıklık matrisi

	Tahmin Değeri: Normal İHA	Tahmin Değeri: Anormal İHA
Gerçek sınıf etiketi: Normal İHA (0 ile etiketli)	TP	FP
Gerçek sınıf etiketi: Anormal İHA (1 ile etiketli)	FN	TN

Tablo 3.2'ye göre Normal İHA ve Anormal İHA olmak üzere iki sınıflı bir modelde;

TP tahmin değeri; Normal İHA sınıfı için başarılı öngörü sayısını ifade etmektedir (Doğru Pozitif).

FP tahmin değeri; Normal İHA sınıfı için başarısız öngörü sayısını ifade etmektedir (Yanlış Pozitif).

FN tahmin değeri; Anormal İHA sınıfı için başarısız öngörü sayısını ifade etmektedir (Yanlış Negatif).

TN tahmin değeri; Anormal İHA sınıfı için başarılı öngörü sayısını ifade etmektedir (Doğru Negatif).

Normal İHA ve Anormal İHA sınıfları için sınıflandırma modellerinin eğitim ve test başarım performansının değerlendirilmesinde, doğruluk (accuracy), özgüllük (specificity) ve duyarlılık (sensitivity) parametrelerinden faydalanılmaktadır (Townsend, 1971). Bu parametreler, Denklem 3.1, Denklem 3.2, Denklem 3.3'e göre hesaplanmaktadır.

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

$$Özgüllük = \frac{TN}{TN+FP} \quad (3.2)$$

$$Duyarlılık = \frac{TP}{TP+FN} \quad (3.3)$$

3.3. ROC Grafikleri

ROC (Receiver Operating Characteristic) eğrisi, sınıflandırma problemlerinde modelin performansını değerlendirmede kullanılmaktadır. Özellikle, ikili sınıflandırma (örneğin, hastalık teşhisi, spam filtresi, vb.) problemlerinde yaygın olarak kullanılmaktadır.

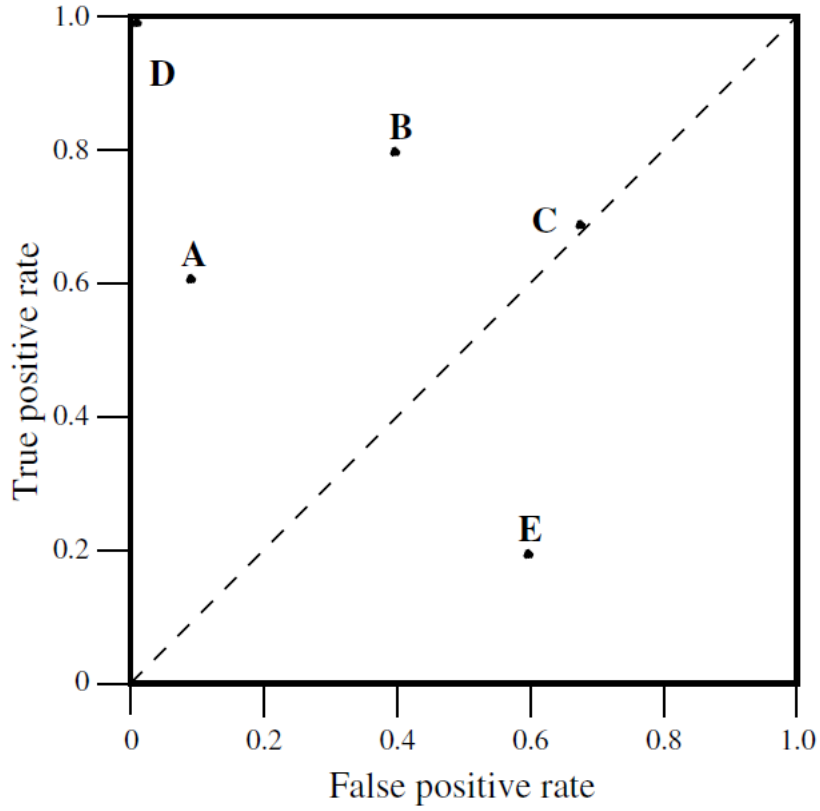
ROC eğrisi, duyarlılık ve (1 – özgüllük) arasındaki dengeyi görselleştirmektedir. Duyarlılık, doğru pozitif oranını (TPR) temsil eder ve pozitif sınıf örnekleri içinde doğru bir şekilde tanımlanan pozitif örneklerin oranını belirtmektedir. (1 – Özgüllük) ise yanlış pozitif oranını (FPR) temsil eder ve negatif sınıf örnekleri içinde yanlış bir şekilde tanımlanan pozitif örneklerin oranını belirtmektedir.

ROC eğrisi, modelin eşik değeri değiştirildiğinde duyarlılık ve özgüllük arasındaki değişimlerden oluşmaktadır. Eğri oluşturulurken, sol alt köşeden sağ üst köşeye doğru bir çizgi oluşturarak, ideal bir sınıflandırıcıyı temsil etmesi hedeflenmektedir. ROC eğrisinin altında kalan alan (AUC – Area Under the Curve), modelin genel performansını ölçen bir metrik olarak kullanılmaktadır. AUC değeri 1'e ne kadar yakınsa, model o kadar iyidir şeklinde yorumlanmaktadır.

ROC eğrisi, modelin doğruluğu dışında bir değerlendirme sağlar ve özellikle dengesiz sınıflandırma problemlerinde (yani, bir sınıf diğerinden çok daha fazla örneğe sahipse), model performansını daha iyi anlamak için kullanışlıdır (Fawcett, 2006).

Sınıflandırma modeli eğitildikten sonra, algoritmanın performansı belirli bir test veri kümesi için değerlendirilmektedir. Yaygın bir yaklaşım, karesel kayıp veya doğruluk gibi genel bir performans ölçütünün, tüm test veri kümesi üzerinden ortalamasının alınarak hesaplanmasıdır. Sınıflandırıcı performansını daha yakından incelemek için

bir ROC eğrisi çizilmekte ve performans metrikleri hesaplanmaktadır. ROC eğrisi üzerindeki her nokta, belirli bir eşik değeri için bir çift TPR ve FPR değerine karşılık gelir. Farklı TPR ve FPR değer çiftlerini bulmak için eşik değeri değiştirilir ve ardından bu çiftler kullanılarak bir ROC eğrisi oluşturulmaktadır.



Şekil 3.1: Beş ayrık sınıflandırıcıyı gösteren temel bir ROC grafiği
Kaynak: (Fawcett, 2006)

Çok sınıflı bir sınıflandırma problemi için, bire–karşı–hepsi kodlama tasarımı kullanılır ve her sınıf için bir ROC eğrisi hesaplanmaktadır. Bire–karşı–hepsi kodlama tasarımı, çok sınıflı bir sınıflandırma problemini bir dizi ikili sınıflandırma problemi olarak ele almaktadır ve her ikili problemde bir sınıfın pozitif, geri kalanın negatif olduğunu varsaymaktadır. İkili sınıflandırıcı tipik olarak bir gözlemi daha yüksek bir puanla sınıflandırmaktadır. Bir sınıflandırıcının eşiği genellikle 0 olarak temsil edilmektedir ve bir gözlemin pozitif veya negatif olduğu belirlenmektedir. Eşiği tüm gözlemlere uygulayarak bir çift TPR ve FPR değeri hesaplanır ve bu çift, ROC eğrisi üzerinde tek bir nokta olarak işaretlenmektedir. Yeni bir eşik değeri olarak 0.25

kullanıldığında, 0.2'lik puana sahip gözlemi negatif sınıfa atamaktadır. Yeni eşiği tüm gözlemlere uygulayarak yeni bir TPR ve FPR değer çifti hesaplar ve ROC eğrisi üzerinde yeni bir nokta olarak belirlenmektedir. Bu işlem, çeşitli eşik değerleri için tekrarlanarak TPR ve FPR değer çiftleri hesaplanmaktadır ve bu çiftler kullanılarak bir ROC eğrisi ortaya çıkmaktadır.

DÖRDÜNCÜ BÖLÜM

DENEYSEL ÇALIŞMALAR

Bu çalışmada kullanılan, İHA'lardan elde edilmiş şifrelenmiş tek yönlü ve çift yönlü WiFi verileri Tablo 4.1'de özetlenmiştir. Tablo 4.1'deki verilere öncelikle mRmR yöntemi uygulanarak anlamlı özellikler seçilmiştir. Anlamlı özellikle ile oluşturulmuş veri setine Karar Ağacı, DVM, KNN yöntemleri uygulanarak normal İHA / Anormal İHA tespiti 5–kat çaprazlama yöntemi ile test edilmiştir.

Tablo 4.1: İHA'lardan elde edilmiş şifrelenmiş tek yönlü ve çift yönlü WiFi veri seti

Veri seti	Toplam Veri Boyutu	Sınıf Etiketi
Çift Yönlü Şifrelenmiş WiFi	42136×54	Normal İHA: 0 Anomaly İHA: 1
Tek Yönlü Şifrelenmiş WiFi	26600×18	

Karar Ağacı, DVM, KNN yöntemlerine uygulanan optimum parametre değerleri Tablo 4.2'deki gibidir.

Tablo 4.2: Sınıflandırma modellerine ait optimum parametreler

Sınıflandırma Yöntemleri	Parametreler ve Değerleri
Karar Ağacı	Gini indeksi
DVM	Lineer çekirdek
KNN	k = 1 için Öklid uzaklığı

4.1. Çift Yönlü Şifrelenmiş WiFi Veri Seti ile Elde Edilen Deneysel Sonuçlar

Çift yönlü WiFi verisine mRmR, RelieF, ANOVA özellik seçimi yöntemleri uygulanarak anlamlı özellikler her bir yöntem için hesaplanmıştır. Özellik seçim yöntemleri ile elde edilen veri setlerine Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. 5–kat çaprazlama yöntemi yanında, veri seti eğitim ve test olarak ayrılmıştır. Oluşturulan eğitim ve test veri seti Tablo 4.3'deki gibidir.

Tablo 4.3: Çift yönlü şifrelenmiş WiFi eğitim ve test verisi

Veri seti	Gözlemler
Eğitim	33709
Test	8427

4.1.1. mRmR Yöntem ile Elde Edilen Deneysel Sonuçlar

Çift yönlü WiFi verisine mRmR özellik seçimi yöntemi uygulanarak anlamlı 9 özellik hesaplanmıştır. Seçilen özellikler Tablo 4.4'de verilmiştir.

Tablo 4.4: Çift Yönlü Trafik akış veri setinden mRmR ile seçilen özellikler

Özellik Adı (*)	**mRMR skoru**
UF_PS_MIN	0.69
TF_PT_MAD	0.67
UF_PS_MeanSquare	0.64
TF_PT_MIN	0.64
DF_PS_MIN	0.64
UF_PT_MAX	0.63
TF_PS_MIN	0.63
UF_PS_MAD	0.61
UF_PT_MIN	0.61

Tablo 4.4'de UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 42136×9 olarak Karar Ağacı, DVM, KNN yöntemlerine uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.5'te verilmiştir.

Tablo 4.5: mRmR yöntemi ile elde edilen veri setine ait eğitim sonuçları

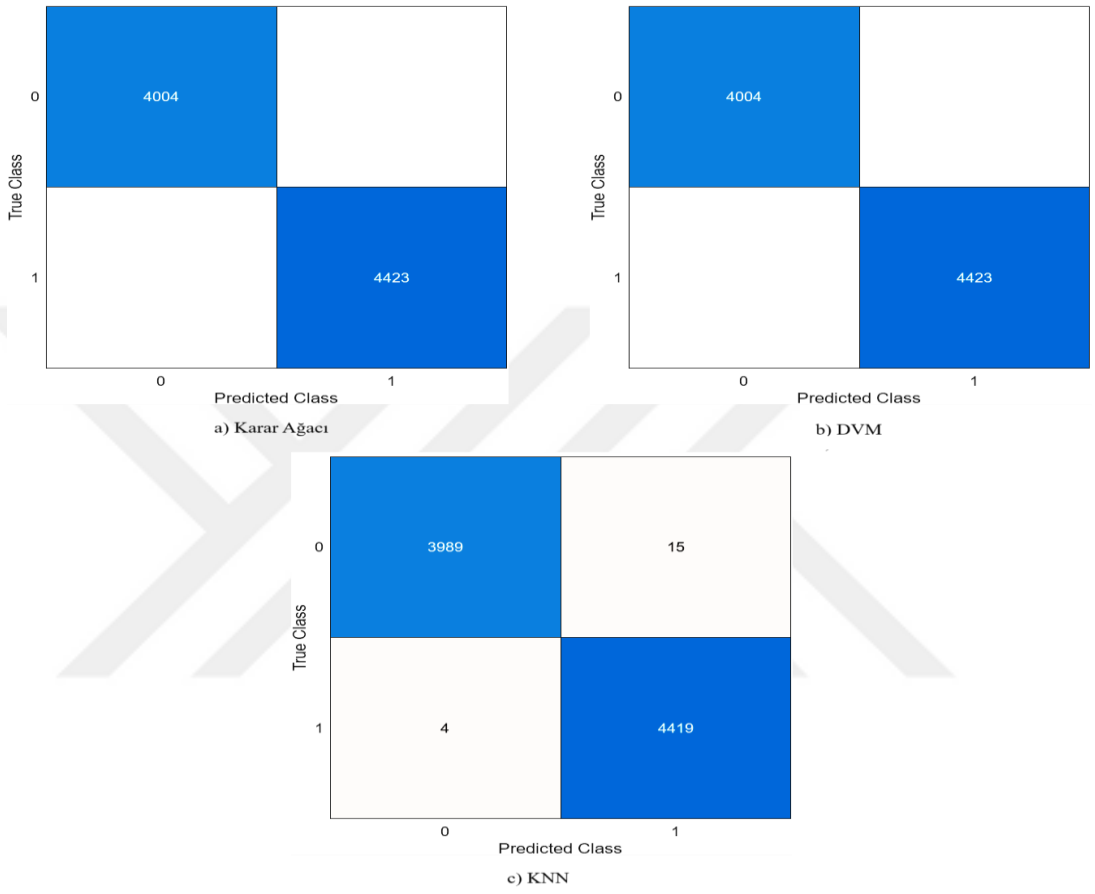
Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~64000	18.97
DVM	100	0	~86000	22.91
KNN	99.7	104	~1000	154.12

Tablo 4.5'e göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler Karar Ağacı ve DVM'dir. Toplam maliyet bakımından en başarılı yöntemler Karar Ağacı ve DVM'dir. Tahminleme hızı bakımından en başarılı yöntem, DVM yöntemi olmuştur. Tahminleme zamanı bakımından en başarılı yöntem Karar Ağacı olmuştur. mRmR veri Setine ait test başarım sonuçları Tablo 4.6'daki gibidir.

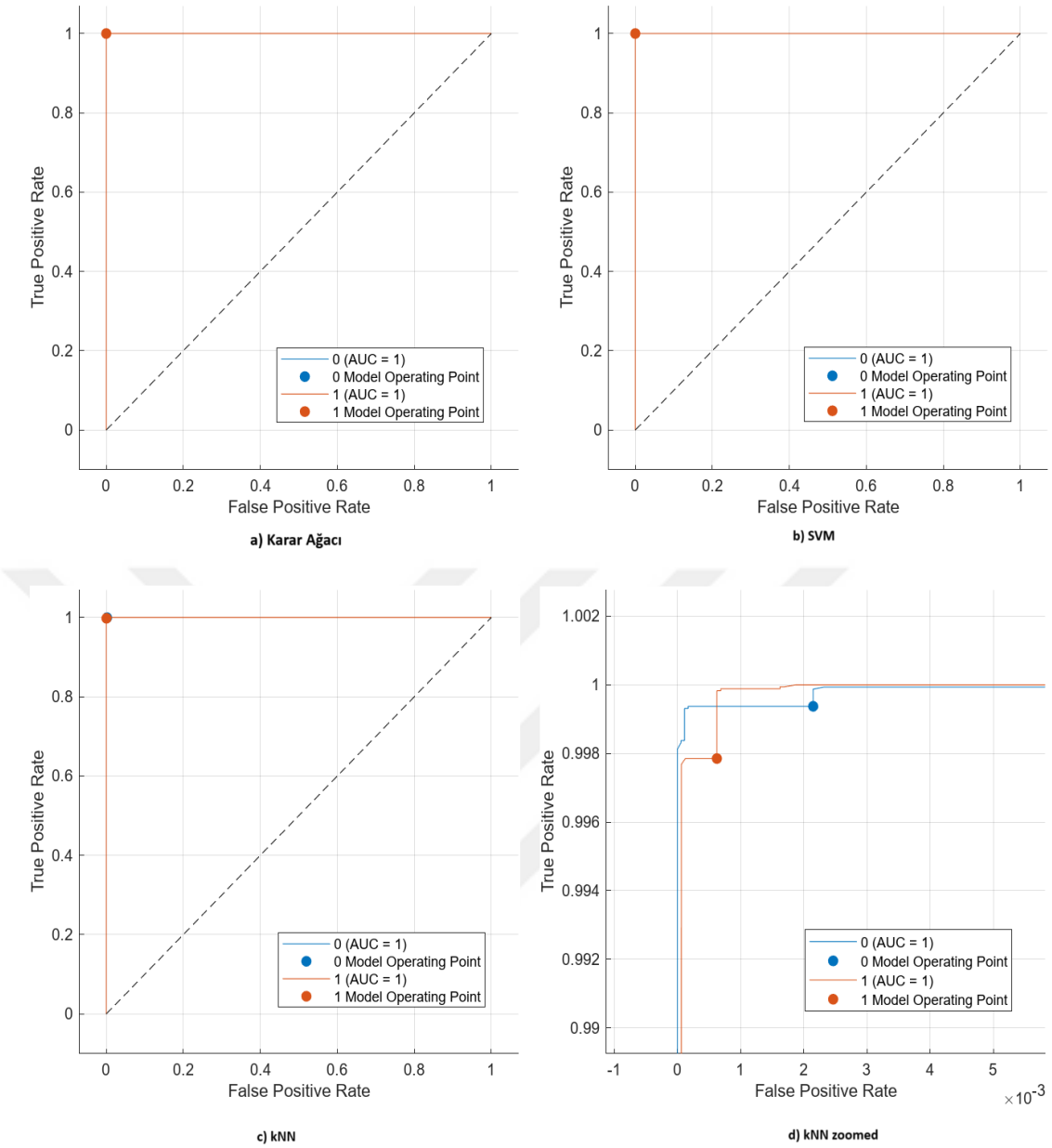
Tablo 4.6: mRmR yöntemi ile elde edilen veri setine ait test veri seti başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	100	100	100
KNN	99.66	99.50	99.79

Test başarımlarına göre karışıklık matrisi Şekil 4.1'de verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı = DVM > KNN olmuştur. Özgüllük başarımı Karar Ağacı = DVM > KNN olmuştur. Duyarlılık bakımında ise model başarımları Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.1: mRmR yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.2: mRmR yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.2'ye göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.1.2. RelieF Yöntemi ile Elde Edilen Deneysel Sonuçlar

Çift yönlü WiFi verisine RelieF özellik seçimi yöntemi uygulanarak anlamlı 10 özellik hesaplanmıştır. Seçilen özellikler Tablo 4.7'de verilmiştir.

Tablo 4.7: Çift Yönlü Trafik akış veri setinden ReflieF ile seçilen özellikler

Özellik Adı	**ReliefF skoru**
TF_PS_MIN	0.48
DF_PS_MIN	0.089
TF_PT_MIN	0.08
UF_PS_Skewness	0.07
UF_PS_MIN	0.07
DF_PS_median	0.06
DF_PT_Kurtois	0.05
UF_PS_mean	0.05
DF_PS_Kurtois	0.05
UF_PS_MeanSquare	0.05

Tablo 4.7'de UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 42136×10 olarak Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.8'de verilmiştir.

Tablo 4.8: RelieF yöntemi ile elde edilen veri setine ait eğitim sonuçları

Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~140000	1321.6
DVM	100	0	~150000	1263.3
KNN	99.72	96	~11000	1261.3

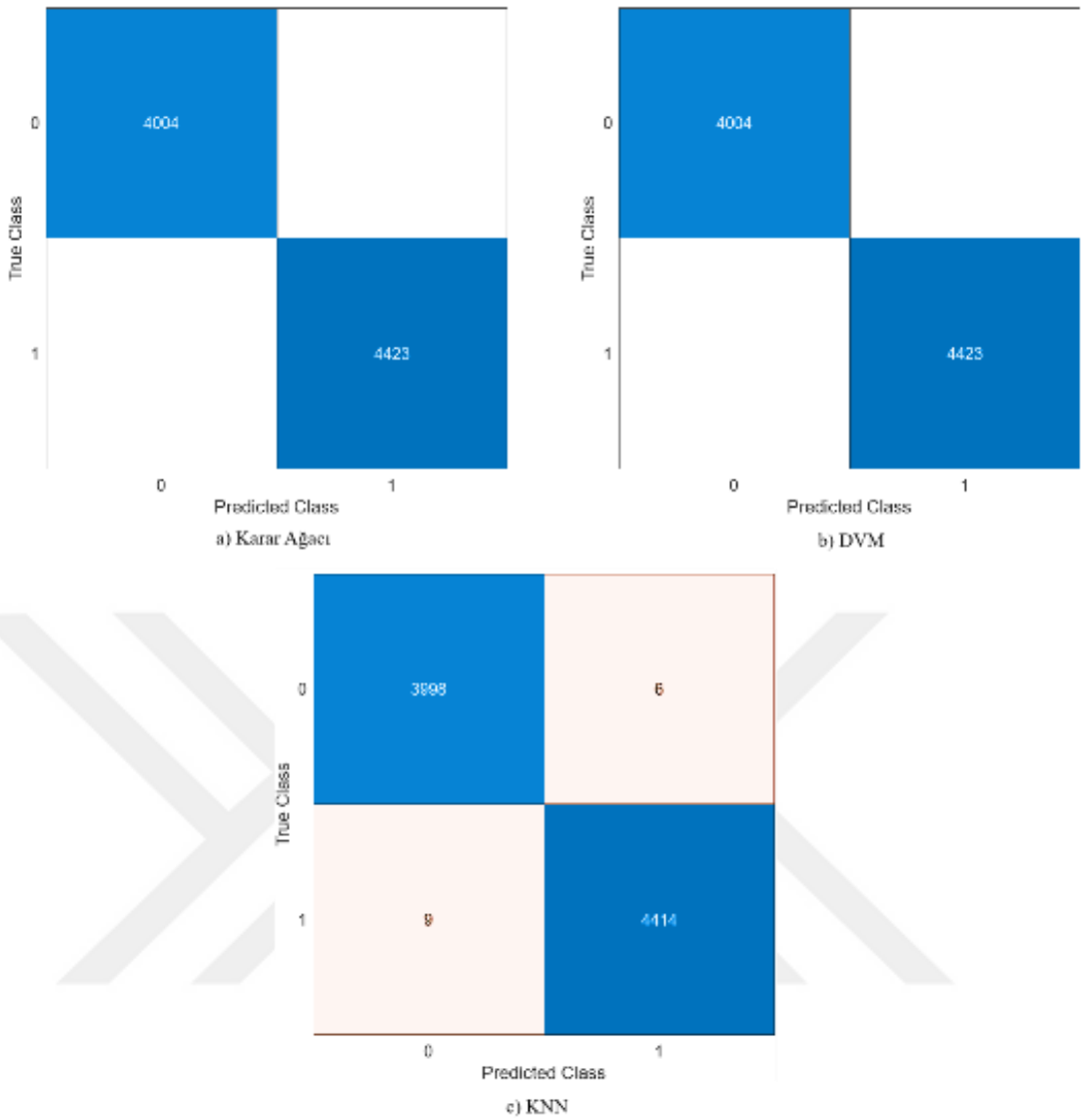
Tablo 4.8'e göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler Karar Ağacı ve DVM'dir. Toplam maliyet bakımından en başarılı yöntemler Karar Ağacı ve

DVM'dir. Tahminleme hızı ve tahminleme zamanı bakımından en başarılı yöntem DVM yöntemi olmuştur. RelieF yöntemi ile elde edilen veri setine ait test başarım sonuçları Tablo 4.9'deki gibidir.

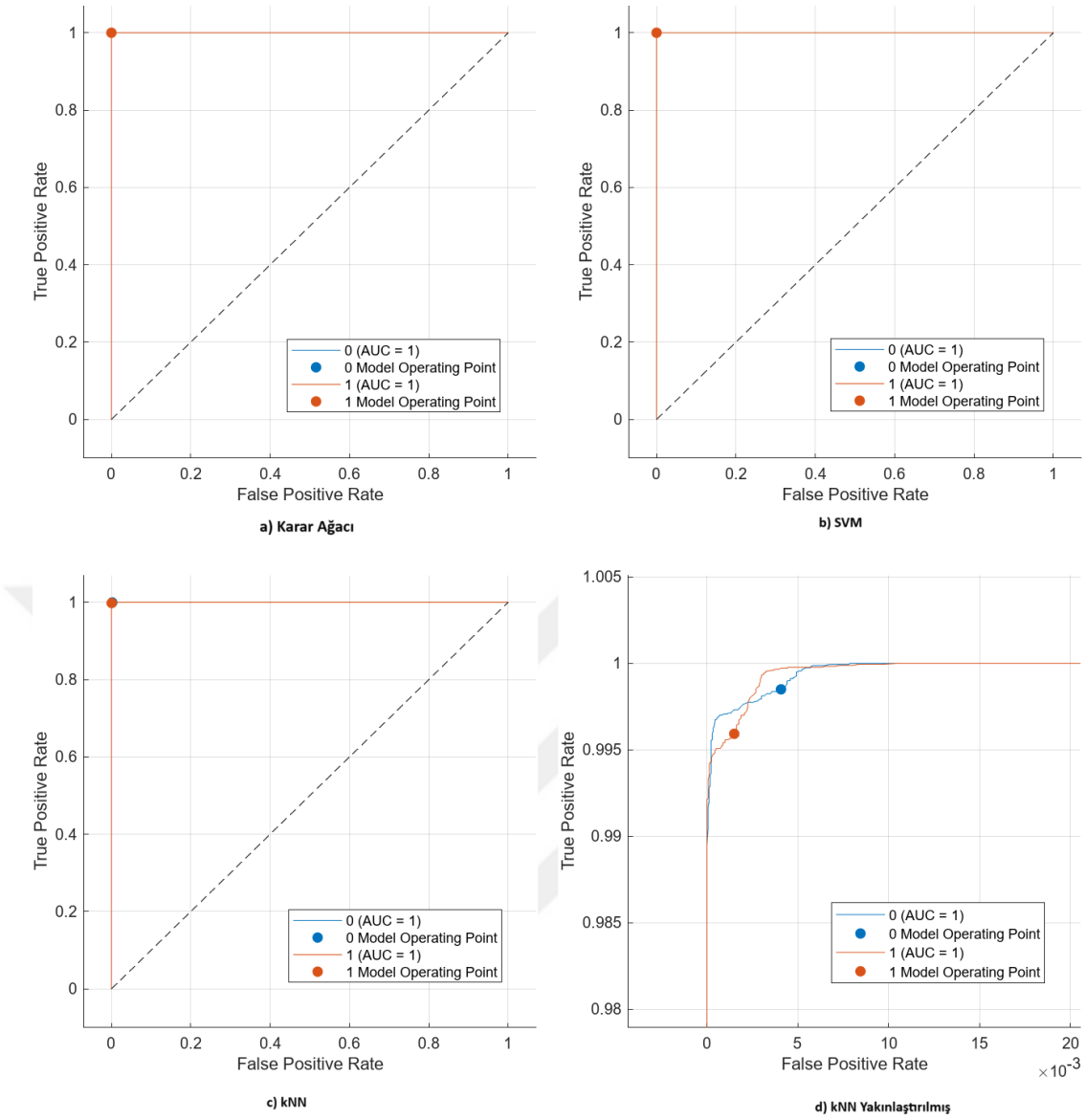
Tablo 4.9: ReliefF yöntemi ile elde edilen veri setine ait test başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	100	100	100
KNN	99.82	99.86	99.7

Test başarımlarına göre karışıklık matrisi Şekil 4.3'de verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı = DVM > KNN olmuştur. Özgüllük ve duyarlılık başarımları Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.3: RelieF yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.4: RelieF yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.4'e göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.1.3. ANOVA ile Elde Edilen Deneysel Sonuçlar

Çift yönlü WiFi verisine ANOVA özellik seçimi yöntemi uygulanarak anlamlı 21 özellik hesaplanmıştır. Seçilen özellikler

Tablo 4.10'da verilmiştir.

Tablo 4.10: Çift Yönlü Trafik akış veri setinden ANOVA ile seçilen özellikler

Özellik adı	ANOVA skoru
TF_PS_MIN	64745
DF_PT_Kurtois	2209
DF_PT_Skewness	2153
UF_PS_MAX	2034
TF_PS_MAX	1365
TF_PS_Skewness	957
TF_PS_MeanSquare	366
TF_PS_Kurtois	355
UF_PS_Skewness	326
DF_PS_Skewness	306
TF_PS_median	279
DF_PS_MAD	249
UF_PS_Kurtois	235
UF_PT_Kurtois	221
DF_PS_MIN	217
TF_PS_mean	190
TF_PT_Kurtois	186
DF_PS_STD	170
TF_PT_Skewness	161
UF_PT_Skewness	146
DF_PS_Kurtois	132

Tablo 4.10'da**,** UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 42136×21 olarak Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.11'de verilmiştir.

Tablo 4.11: ANOVA yöntemi ile elde edilen veri setine ait eğitim sonuçları

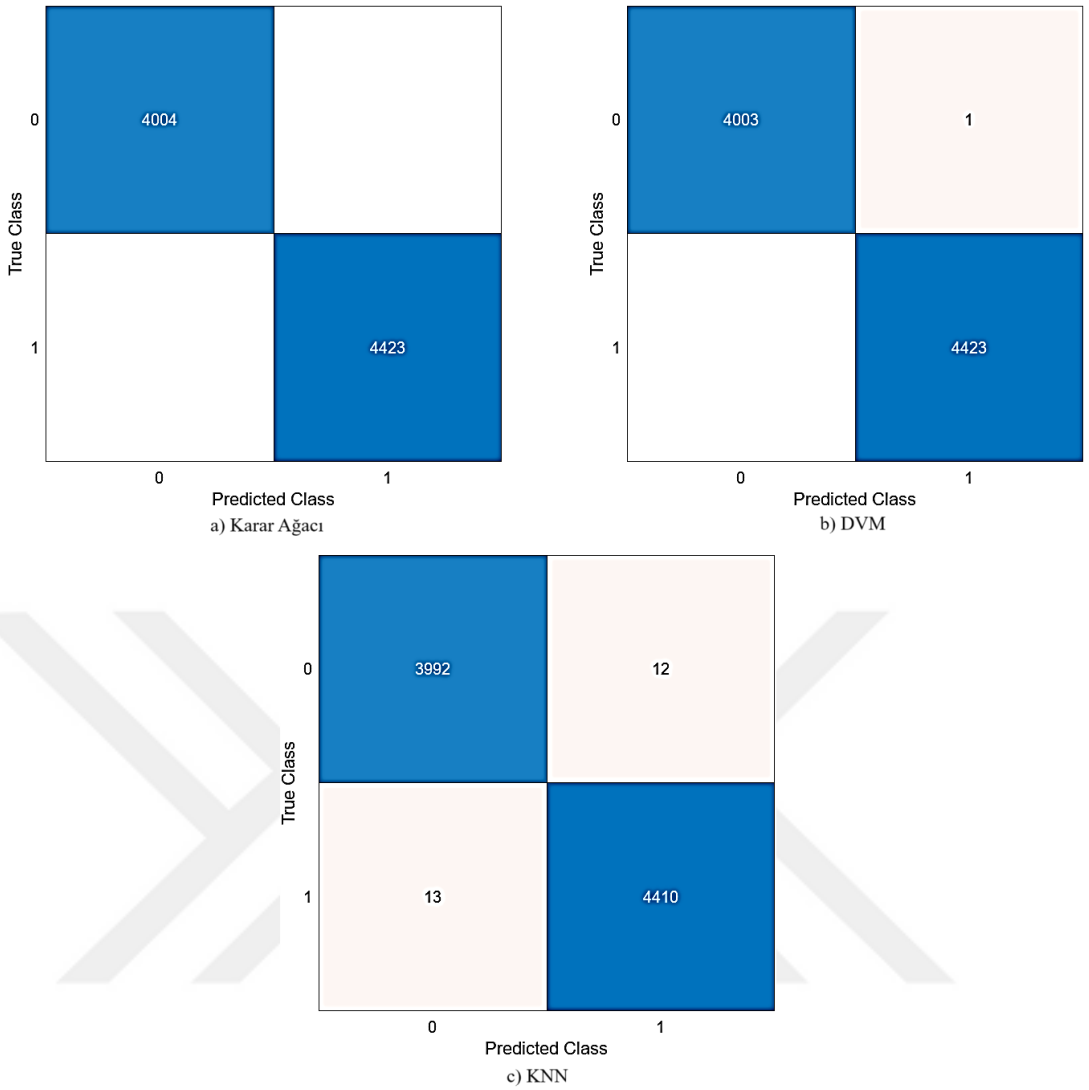
Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~110000	11.56
DVM	100	0	~86000	18.10
KNN	99.80	75	~2100	74.04

Tablo 4.11'e göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler Karar Ağacı ve DVM'dir. Toplam maliyet bakımından en başarılı yöntemler Karar Ağacı ve DVM'dir. Tahminleme hızı en başarılı yöntem Karar Ağacı yöntemi olmuştur. Tahminleme zamanı bakımından en başarılı yöntem yine Karar Ağacı yöntemi olmuştur. ANOVA yöntemi ile elde edilen veri setine ait test başarım sonuçları Tablo 4.12'deki gibidir.

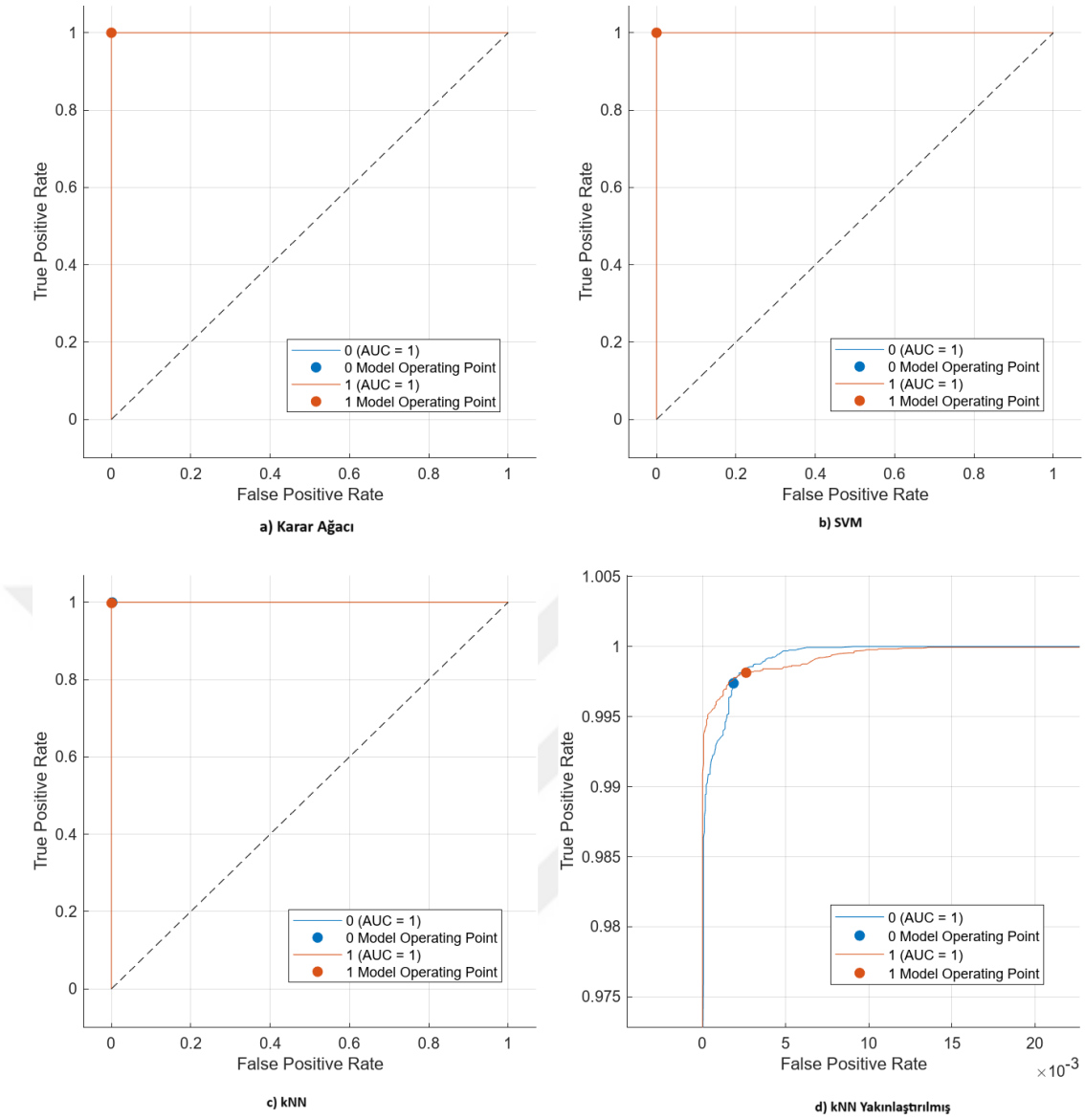
Tablo 4.12: ANOVA yöntemi ile elde edilen veri setine ait test başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	99.99	99.98	100
KNN	99.70	99.70	99.68

Test başarımlarına göre karışıklık matrisi Şekil 4.5'te verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı > DVM > KNN olmuştur. Özgüllük başarımları Karar Ağacı > DVM > KNN olmuştur. Duyarlılık başarımları Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.5: ANOVA yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.6: ANOVA yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.6'ya göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.2. Tek Yönlü Şifrelenmiş WiFi Verisinden Elde Edilen Sonuçlar

Tek yönlü WiFi verisine mRmR, RelieF, ANOVA özellik seçimi yöntemleri uygulanarak anlamlı özellikler her bir yöntem için hesaplanmıştır. Özellik seçim yöntemleri ile elde edilen veri setlerine Karar Ağacı, DVM, KNN yöntemleri

uygulanarak sınıflandırılmıştır. 5–kat çaprazlama yöntemi yanında, veri seti eğitim ve test olarak ayrılmıştır. Oluşturulan eğitim ve test veri seti Tablo 4.13'deki gibidir.

Tablo 4.13: Tek yönlü şifrelenmiş WiFi eğitim ve test verisi

Veri seti	Gözlemler
Eğitim	21280
Test	5320

4.2.1. mRmR Yöntem ile Elde Edilen Deneysel Sonuçlar

Tek yönlü WiFi verisine mRmR özellik seçimi yöntemi uygulanarak anlamlı 6 özellik hesaplanmıştır. Seçilen özellikler Tablo 4.14'de verilmiştir.

Tablo 4.14: Tek Yönlü Trafik akış veri setinden mRmR ile seçilen özellikler

Özellik adı	**mRMR skoru**
TF_PS_MIN	0.69
TF_PT_MIN	0.61
TF_PS_mean	0.54
TF_PS_MAX	0.53
TF_PT_MAD	0.53
TF_PT_MAX	0.50

Tablo 4.14'de UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 21280×6 olarak Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.15'de verilmiştir.

Tablo 4.15: mRmR yöntemi ile elde edilen veri setine ait eğitim sonuçları

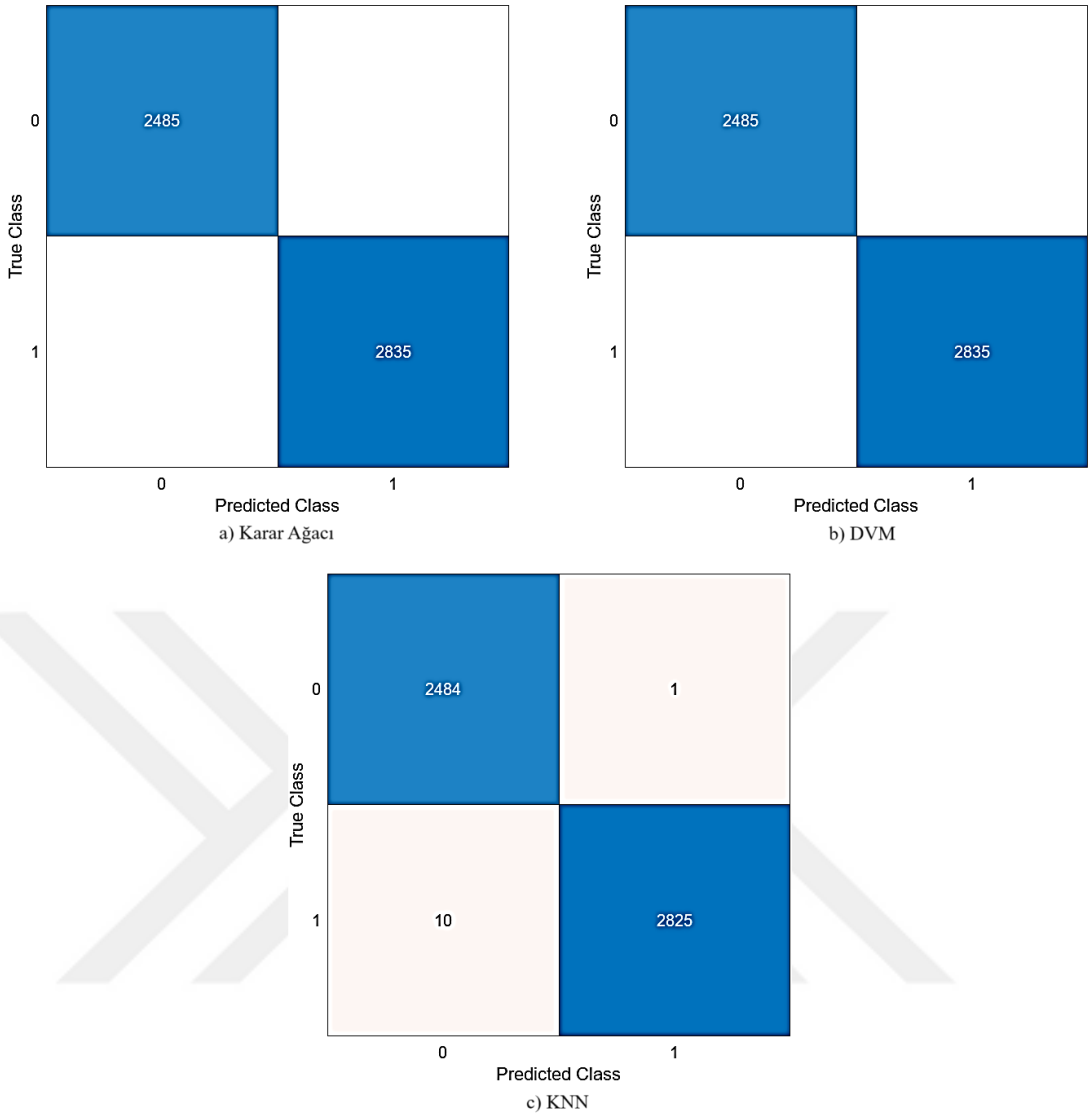
Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~420000	3.47
DVM	100	0	~280000	4.96
KNN	99.82	39	~20000	7.46

Tablo 4.15'e göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler büyükten küçüğe sırasıyla Karar Ağacı = DVM > KNN'dir. Toplam maliyet bakımından en başarılı yöntemler Karar Ağacı ve DVM'dir. Tahminleme hızı bakımından en başarılı model, Karar Ağacı yöntemi olmuştur. Tahminleme zamanı bakımından en başarılı yöntem yine Karar Ağacı olmuştur. mRmR veri setine ait test başarım sonuçları Tablo 4.16'deki gibidir.

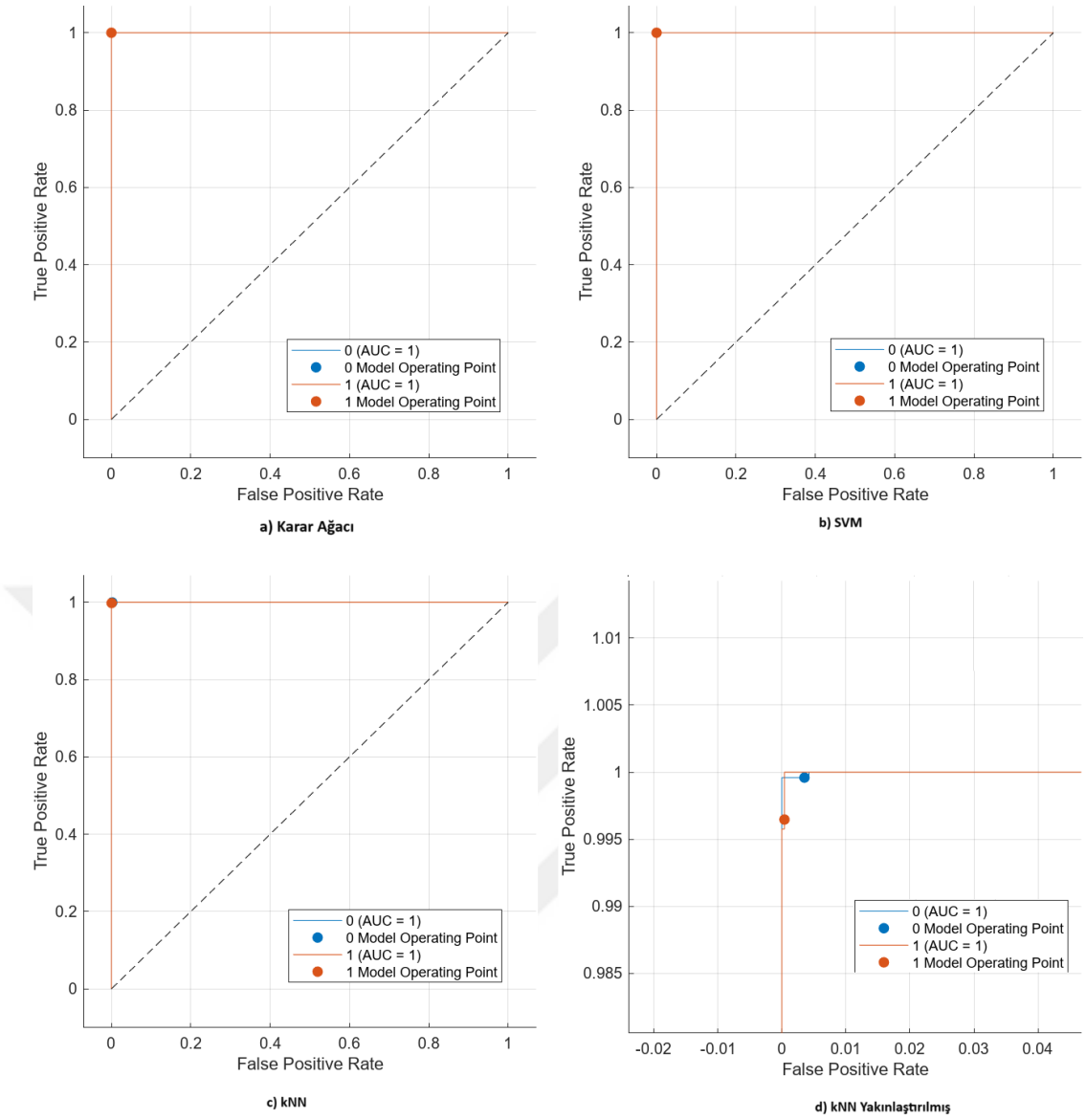
Tablo 4.16: mRmR yöntemi ile elde edilen veri setine ait test veri seti başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	100	100	100
KNN	99.79	99.96	99.60

Test başarımlarına göre karışıklık matrisi Şekil 4.7'de verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı = DVM > KNN olmuştur. Özgüllük başarımı Karar Ağacı = DVM > KNN olmuştur. Duyarlılık bakımında ise model başarımları Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.7: mRmR yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.8: mRmR yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.8'e göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.2.2. RelieF Yöntem ile Elde Edilen Deneysel Sonuçlar

Tek yönlü WiFi verisine RelieF özellik seçimi yöntemi uygulanarak anlamlı 5 özellik hesaplanmıştır. Seçilen özellikler

Tablo 4.17'de verilmiştir.

Tablo 4.17: Tek Yönlü Trafik akış veri setinden RelieF ile seçilen özellikler

Özellik adı	**RelieF skoru**
TF_PS_MIN	0.08
TF_PS_Kurtois	0.04
TF_PS_MAD	0.03
TF_PS_median	0.02
TF_PS_mean	0.02

Tablo 4.17'ye göre, UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 42136×5 olarak Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.18'de verilmiştir.

Tablo 4.18: RelieF yöntemi ile elde edilen veri setine ait eğitim sonuçları

Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~330000	245.13
DVM	100	1	~290000	261.96
KNN	99.90	21	~28000	262.45

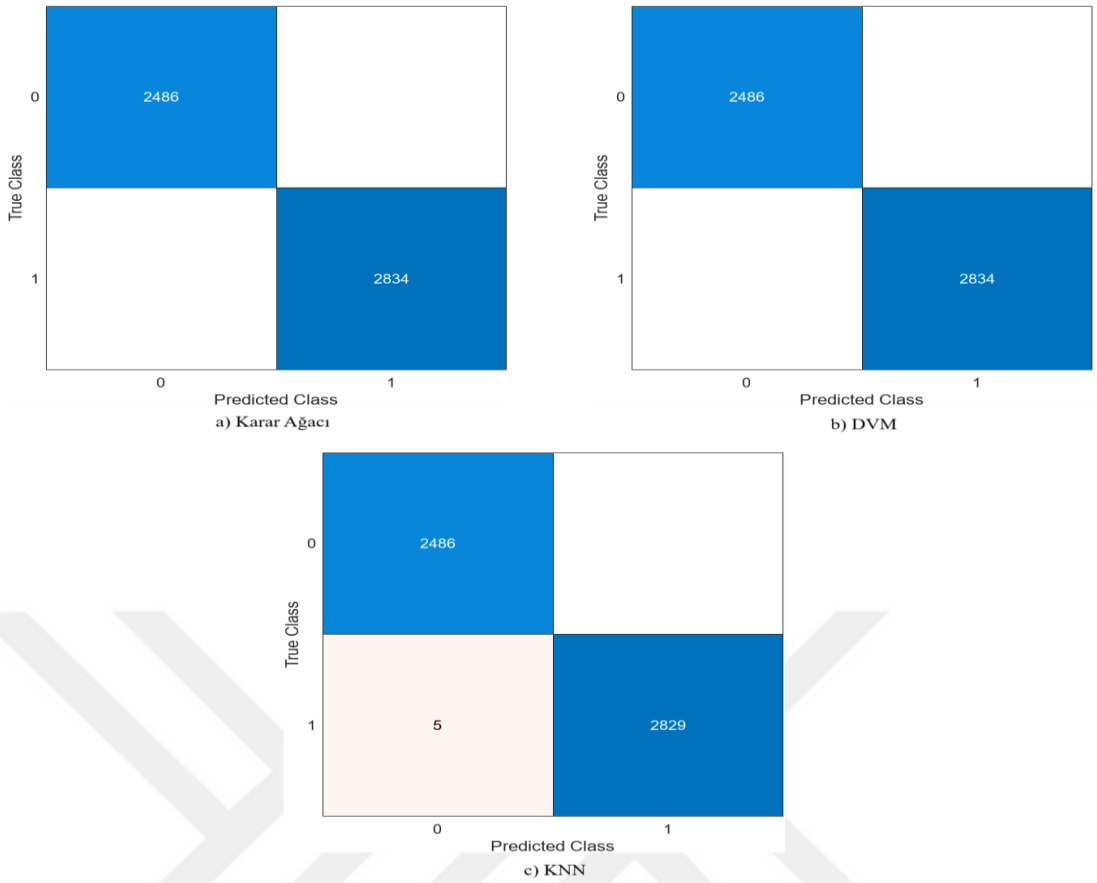
Tablo 4.18'e göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler Karar Ağacı ve DVM'dir. Toplam maliyet bakımından en başarılı yöntemler Karar Ağacı yöntemidir. Tahminleme hızı bakımından en başarılı yöntem Karar Ağacı yöntemi olmuştur. Tahminleme zamanı bakımından Karar Ağacı yöntemi en başarılıdır. RelieF yöntemi ile elde edilen veri setine ait test başarım sonuçları

Tablo **4.19**'daki gibidir.

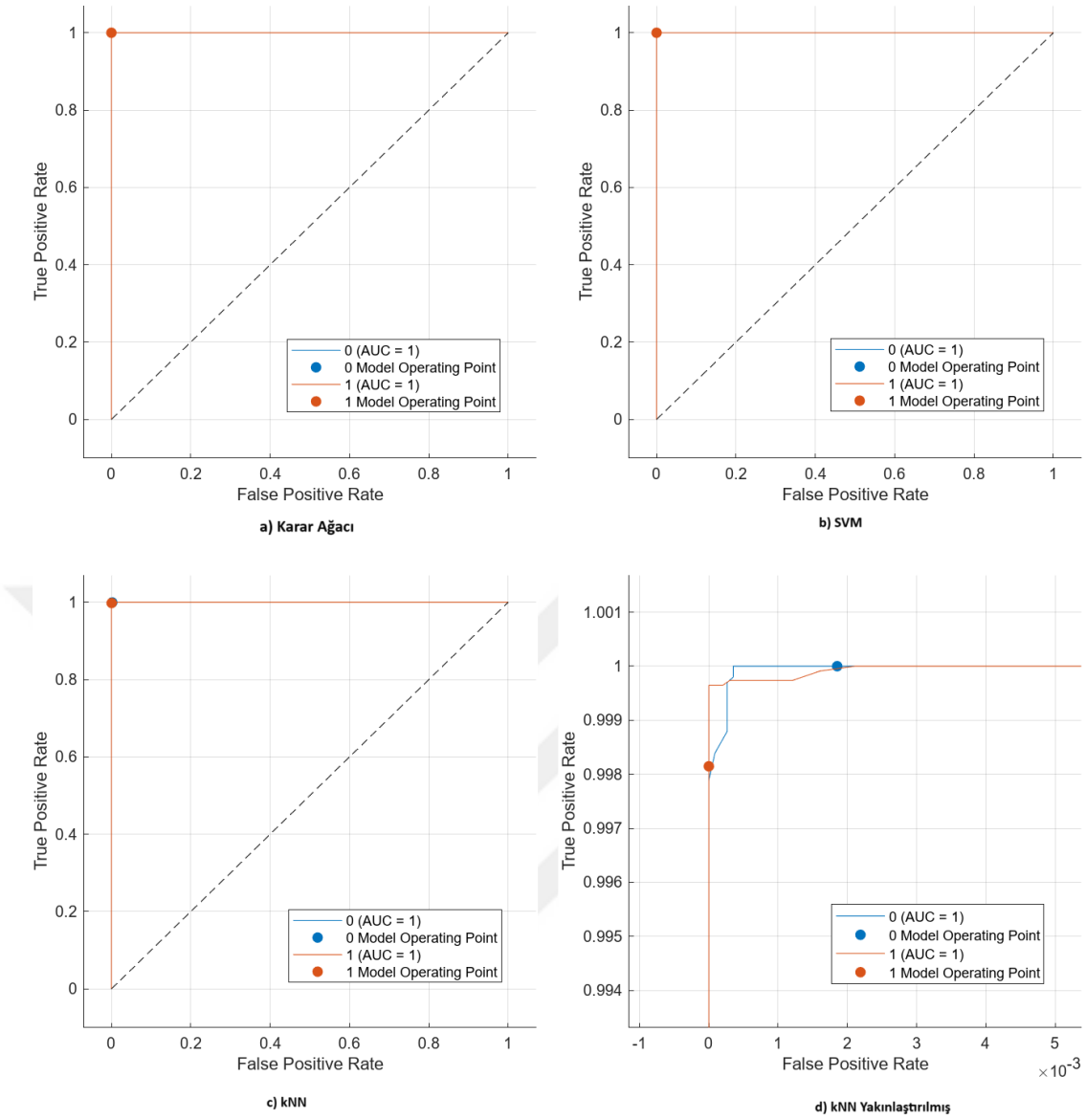
Tablo 4.19: RelieF yöntemi ile elde edilen veri setine ait test başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	100	100	100
KNN	99.91	100	99.80

Test başarımlarına göre karışıklık matrisi Şekil 4.9'te verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı = DVM > KNN olmuştur. Özgüllük başarımları 3 yöntem için de %100 başarılıdır. Yöntemlere göre test duyarlılığı Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.9: RelieF yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.10: RelieF yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.10'a göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.2.3. ANOVA ile Elde Edilen Deneysel Sonuçlar

Tek yönlü WiFi verisine ANOVA özellik seçimi yöntemi uygulanarak anlamlı 8 özellik hesaplanmıştır. Seçilen özellikler Tablo 4.20'de verilmiştir.

Tablo 4.20: Tek Yönlü Trafik akış veri setinden ANOVA ile seçilen özellikler

Özellik Adı	**ANOVA skoru**
TF_PS_MAD	10827
TF_PS_mean	6692
TF_PS_MeanSquare	5063
TF_PS_MIN	2956
TF_PS_median	2823
TF_PS_MAX	2557
TF_PT_MIN	1209
TF_PS_STD	1046

Tablo 4.20'de UF: Yukarı Akış, DF: Aşağı Akış, TF: Toplam Akış PS: Paket Boyutu, PT: Paket Varışlar Arası Süre özelliklerini temsil etmektedir.

Elde edilen veri seti 42136×8 olarak Karar Ağacı, DVM, KNN yöntemleri uygulanarak sınıflandırılmıştır. Eğitim veri setinden elde edilen validasyon sonuçları Tablo 4.21'de verilmiştir.

Tablo 4.21: ANOVA yöntemi ile elde edilen veri setine ait eğitim sonuçları

Sınıflandırma Yöntemi	Validasyon Doğruluğu (%)	Validasyon Toplam Maliyet	Tahminleme Hızı (obs/ sec)	Tahminleme Zamanı (sec)
Karar Ağacı	100	0	~420000	3.23
DVM	99.98	5	~350000	9.12
KNN	99.76	52	~24000	5.74

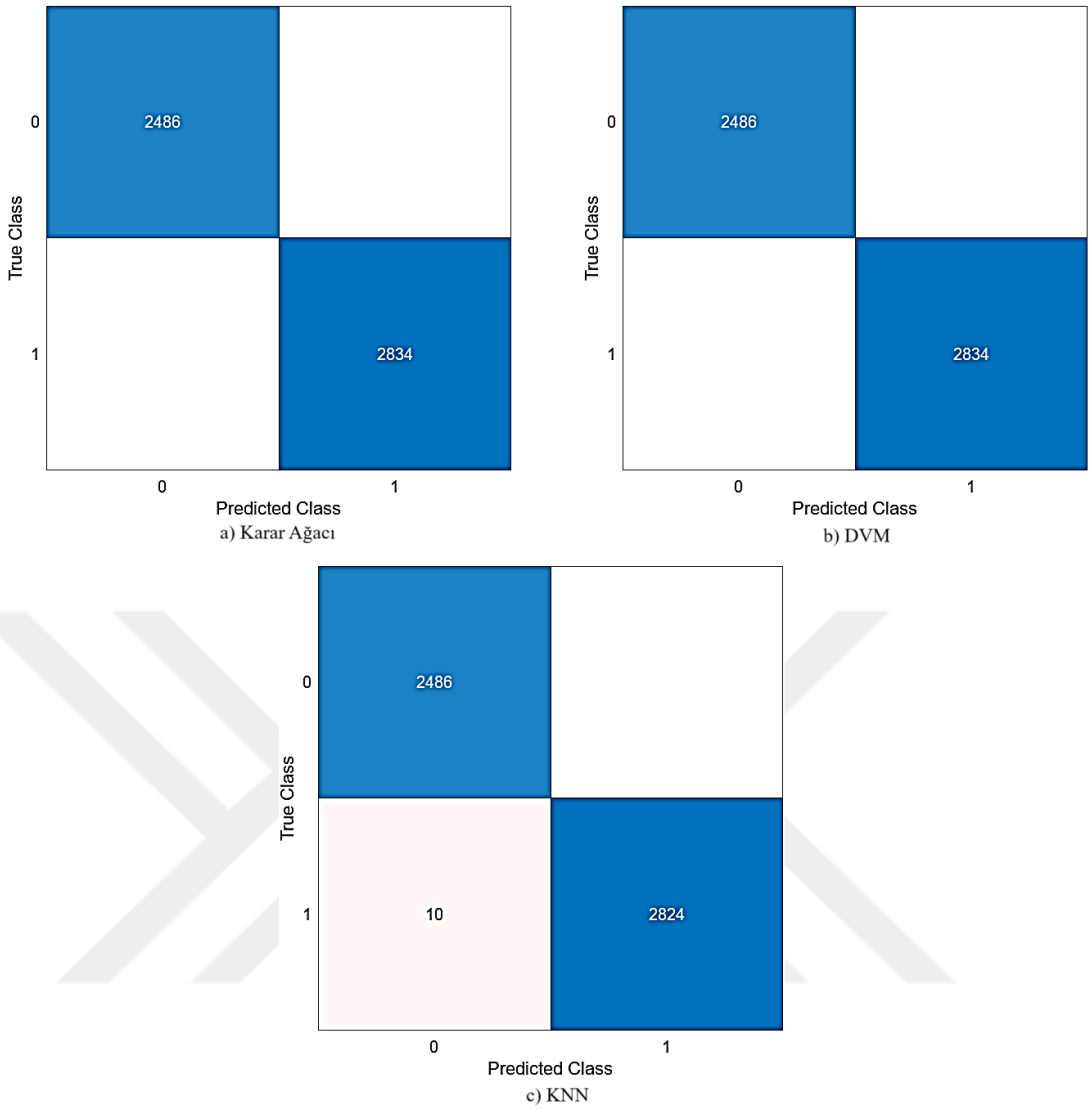
Tablo 4.21'ye göre, eğitim aşamasında validasyon doğruluğu en başarılı yöntemler Karar Ağacı > DVM > KNN yöntemleridir. Toplam maliyet bakımından en iyi sonuç Karar Ağacı yönteminden elde edilmiştir. Tahminleme hızı en başarılı yöntem Karar Ağacı yöntemi olmuştur. Tahminleme zamanı bakımından en başarılı yöntem yine

Karar Ağacı yöntemi olmuştur. ANOVA yöntemi ile elde edilen veri setine ait test başarım sonuçları Tablo 4.22'deki gibidir.

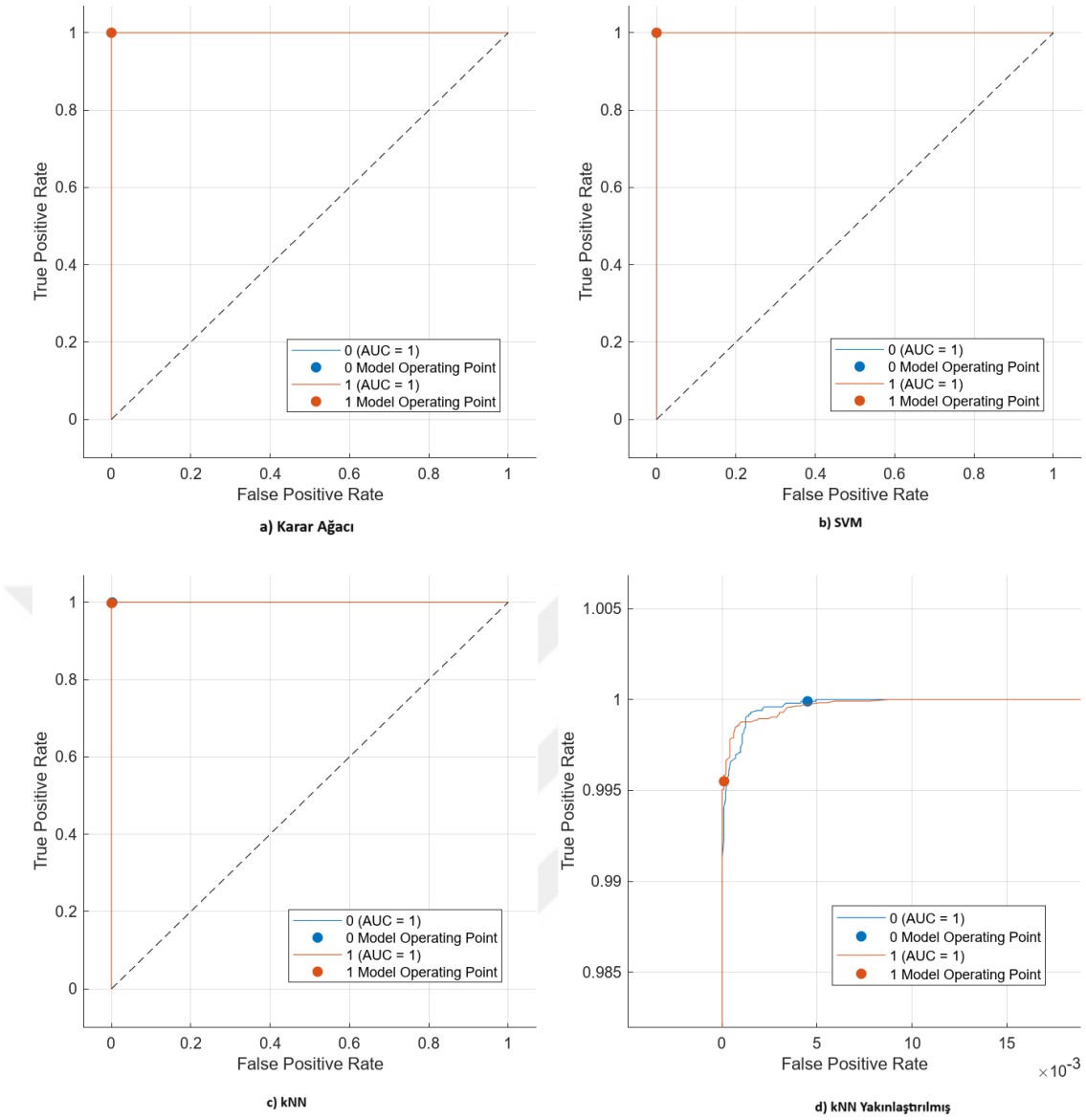
Tablo 4.22: ANOVA yöntemi ile elde edilen veri setine ait test başarımları

Sınıflandırma Yöntemi	Doğruluk (%)	Özgüllük(%)	Duyarlılık (%)
Karar Ağacı	100	100	100
DVM	100	100	100
KNN	99.81	100	99.60

Test başarımlarına göre karışıklık matrisi Şekil 4.11'de verilmiştir. Yöntemlere göre test doğruluğu Karar Ağacı = DVM > KNN olmuştur. Özgüllük başarımları her üç yöntemde de %100 başarılı olmuştur. Duyarlılık başarımları Karar Ağacı = DVM > KNN olmuştur.



Şekil 4.11: ANOVA yöntemi ile elde edilen veri setine ait karışıklık matrisi



Şekil 4.12: ANOVA yöntemi ile elde edilen veri setine ait ROC eğrileri

Şekil 4.12'ye göre Karar Ağacı, DVM ve KNN yöntemlerine ait ROC eğrilerinde AUC değeri 1'e çok yakındır ve modellerin test başarımları iyi bir performans göstermektedir.

4.3. Deneysel Sonuçların Yorumlanması

Bu çalışmada, Parrot Bebop I, DBPower UDI, DJI Spark İHA'larından elde edilen çift yönlü şifrelenmiş WiFi verileri ve tek yönlü şifrelenmiş WiFi verilerine mRMR, RelieF, ANOVA özellik seçim yöntemleri uygulanmıştır. İki farklı şifrelenmiş veri

için kamu güvenliği ve kişisel mahremiyetin sağlanmasında anlamlı özelliklerin seçilmesi hedeflenmiştir. Elde edilen özellik setleri Karar Ağacı, DVM ve KNN makine öğrenmesi yöntemlerine uygulanarak modellenmiştir. Model başarımları 5–kat çaprazlama yöntemi ile test edilmiştir. Modeller arası eğitim başarımı validasyon doğruluğu, validasyon toplam maliyet, tahminleme hızı ve tahminleme zamanı bakımından karşılaştırılmıştır. Modellerin test başarımları ise doğruluğu, özgüllük ve duyarlılık oranları bakımından karşılaştırılmıştır.

Çift yönlü şifrelenmiş WiFi veri seti için validayon doğruluğu en başarılı yöntem her üç özellik seçiminde de Karar Ağacı %100 başarılı olmuştur. Validasyon toplam maliyet bakımından her üç özellik seçimi yönteminde Karar Ağacı ve DVM yöntemleri en başarılı olmuştur. Tahminleme hızı bakımından RelieF yöntemi ile seçilen özellikler ile DVM sınıflandırma yöntemi ~150000 obs/sec ile en başarılı yöntem olmuştur. Tahminleme hızı bakımından ANOVA yöntemi ile elde edilen özellik setinde Karar Ağacı yöntemi 11,56 saniye ile başarılı olmuştur.

Çift yönlü WiFi verileri için mRmR, RelieF, ANOVA yöntemlerinden elde edilen özellik setlerinin test başarımlarında ortak en başarılı yöntem %100 doğrulukla Karar Ağacı olmuştur.

Tek yönlü şifrelenmiş WiFi veri seti için validayon doğruluğu en başarılı yöntem her üç özellik seçiminde de Karar Ağacı %100 başarılı olmuştur. Validasyon toplam maliyet bakımından her üç özellik seçimi yönteminde Karar Ağacı yöntemi en başarılı yöntem olmuştur. Tahminleme hızı bakımından hem mRmR ve hem ANOVA yöntemi ile elde edilen veri seti Karar Ağacı yönteminde ~420000 obs/sec ile en başarılı yöntem olmuştur. Tahminleme hızı bakımından ANOVA yöntem ile elde edilen özellik setinde Karar Ağacı yöntemi 3,23 saniye ile başarılı olmuştur.

Tek yönlü WiFi verileri için mRmR, RelieF, ANOVA yöntemlerinden elde edilen özellik setlerinin test başarımlarında ortak en başarılı yöntem %100 doğrulukla Karar Ağacı ve DVM olmuştur.

Amir Alipour-Fanid ve arkadaşlarının aynı veri seti ile yaptığı çalışmada DVM tabanlı bir sınıflandırma yöntemi ile en yüksek %95,2 doğruluk başarım elde edilmesine karşın (Amir Alipour-Fanid, Ning Wang, Liang Zhao, 2020) bu tezde yapılan çalışmada gerçekleştirilen özellik seçimi yöntemleri ile daha küçük boyutlu özellik uzayı ile Karar Ağacı yöntemi ile %100 doğruluk başarımı elde edilmiştir.

BEŞİNCİ BÖLÜM
SONUÇ

Bu tez, İHA'ların toplum güvenliğine olumlu bir katkı sağlamak amacıyla yeni bir İHA tespiti yaklaşımının geliştirilmesi açısından büyük bir öneme sahiptir. Aynı zamanda, özel hayat gizliliğini koruma amacı güdülen özgün bir yaklaşım sunarak bu alandaki araştırmalara yeni bir boyut kazandırmayı hedeflemektedir. İHA'ların giderek yaygınlaşması ve potansiyel güvenlik riskleri nedeniyle, bu tezin sonuçları toplumun güvenliğine ve özel hayat gizliliğine yönelik önemli katkılarda bulunması hedeflenmektedir. Bu doğrultuda Parrot Bebop I, DBPower UDI, DJI Spark İHA'larından elde edilen çift yönlü şifrelenmiş WiFi verileri ve tek yönlü şifrelenmiş WiFi verilerine mRMR, RelieF, ANOVA özellik seçim yöntemleri uygulanmıştır. Elde edilen özellik setleri Karar Ağacı, DVM ve KNN makine öğrenmesi yöntemlerine uygulanarak modellenmiştir. Model başarımları 5–kat çaprazlama yöntemi ile test edilmiştir. Modeller arası eğitim başarımı validasyon doğruluğu, validasyon toplam maliyet, tahminleme hızı ve tahminleme zamanı bakımından karşılaştırılmıştır. Modellerin test başarımları ise doğruluğu, özgüllük, duyarlılık oranları ve ROC analizi bakımından karşılaştırılmıştır. Çift yönlü şifrelenmiş WiFi verisi için mRMR, RelieF, ANOVA yöntemlerinden elde edilen test doğrulukları, özgüllük ve duyarlılık oranları bakımında Karar Ağacı yönteminin başarımı dikkat çekicidir, tek yönlü şifrelenmiş WiFi verisi için Karar Ağacı ve DVM başarımı dikkat çekici olmuştur. Aynı veri seti ile yapılan başka bir çalışmada DVM yöntemi %95,2 doğruluk başarım elde edilmesine karşın bu tezde yapılan çalışmada daha küçük boyutlu özellik uzayı ile Karar Ağacı yöntemi ile %100 doğruluk başarımı elde edilmiştir. Gelecekte farklı yöntemlerin uygulandığı, özgün ve daha büyük veri setleri ile çalışılması hedeflenmektedir.

KAYNAKÇA

AJPAS, A. (2016). A Feature Selection Based on One-Way-Anova for Microarray Data Classification. *AJPAS JOURNAL, 3*, 1-6.

Akyıldız, M. (2009, 04 12). *Tek Faktörlü Varyans Analizi (One-Way Anova) ve bir spss örneği*. (istatistik.gen.tr) 12 10, 2023 tarihinde https://www.istatistik.gen.tr/?p=29 adresinden alındı

Amir Alipour-Fanid, Ning Wang, Liang Zhao. (2020). Machine Learning-Based Delay-Aware UAV Detection and Operation Mode Identification Over Encrypted Wi-Fi Traffic. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, 15*, 2346 - 2360.

Bar-Yanai, R., Langberg, M., Peleg, D., & Roditty, L. (2010). Realtime Classification for Encrypted Traffic. *Experimental Algorithms. Springer, 6049*, 373-385.

Bayrak, S. (2021). Farklı veri türlerinin birleştirilmesi ile epilepsi hastalığının otomatik tespiti ve sınıflandırılması. *Doktora Tezi, İstanbul Üniversitesi – Cerrahpaşa*.

Bayrak, S., & Yucel, E. (2022). Methods for the Recognition of Multisource Data in Intelligent Medicine: A Review and Next-Generation Trends. *Next Generation Healthcare Informatics*, 1-25.

Bayrak, S., Yucel, E., & Takci, H. (2019). Classification of extracranial and intracranial EEG signals by using finite impulse response filter through ensemble learning. *27th Signal Processing and Communications Applications Conference (SIU)* (s. 1-4). IEEE.

Bayrak, S., Yucel, E., & Takci, H. (2022). Epilepsy Radiology Reports Classification Using Deep Learning Networks. *Computers, Materials and Continua: Tech Science Press, 70*(2), 3589-607.

Busset, J., Perrodin, F., Wellig, P., Ott, B., Heutschi, K., Rühl, T., & Nussbaumer, T. (2015). Detection and tracking of drones using advanced acoustic cameras. *Proc. SPIE, 9647*. doi:10.1117/12.2194309

Contributors, W. (2023, 12 11). *Relief (feature selection)*. (Wikipedia, The Free Encyclopedia.) 12 16, 2023 tarihinde

https://en.wikipedia.org/w/index.php?title=Relief_(feature_selection)&oldid=1189402629 adresinden alındı

F. Gökçe, G. Üçoluk, E. ¸ Sahin, and S. Kalkan. (2015). Vision-based detection and distance estimation of micro unmanned aerial vehicles. *Sensors, 15*(9), 23805–23846.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874.

Fiori, L. (2020, 5 22). *Distance metrics and K-nearest neighbor (KNN)*. Medium: https://medium.com/@luigi.fiori.lf0303/distance-metrics-and-k-nearest-neighbor-knn-1b840969c0f4 adresinden alındı

G. J. Mendis, T. Randeny, J. Wei, and A. Madanayake. (2016). Deep learning based Doppler radar for micro UAS detection and classification. *IEEE Military Communications Conference.* Baltimore, MD, USA.

Gajawada, S. K. (2019, October 19). *ANOVA for Feature Selection in Machine Learning*. (Towards Data Science) November 11, 2023 tarihinde https://towardsdatascience.com/anova-for-feature-selection-in-machine-learning-d9305e228476 adresinden alındı

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN Model-Based Approach in Classification. *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences* (s. 986-996). Catania, Sicily, Italy: Springer.

H. Peng, F. Long and C. Ding. (Aug. 2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(8), 1226-1238.

Jing, N., Yang, M., Cheng, S., Dong, Q., & Xiong, H. (Nov. 2011). An efficient svm-based method for multi-class network traffic classification. *30th IEEE International Performance Computing and Communications Conference.* Orlando, FL.

Küçüksille, E., & Ateş, N. (2016). Spam e-mail Filtering Using Support Vector Machine. *TBV-BBMD, 6*(1).

Liang Zhao, Amir Alipour-Fanid, Martin Slawski and Kai Zeng. (2018). Prediction-time Efficient Classification Using Feature Computational Dependencies. *Proceedings of the 24st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2018)* (s. 2787-2796). London, United Kingdom: ACM.

M. Messina and G. Pinelli. (2019). Classification of drones with a surveillance. *Proceedings of ICCV*. Seoul, Korea.

Marmaroli, P., Falourd, X., & Lissek, H. (Apr. 2012). A UAV motor denoising technique to improve localization of surrounding noisy aircrafts: Proof of concept for anti-collision systems. *Proc. Acoustic*.

McGregor, A., Hall, M., Lorier, P., & Brunskill, J. (2004). Flow Clustering Using Machine Learning Techniques. *Passive and Active Network Measurement Conference.*

P. Zhang, L. Yang, G. Chen, and G. Li. (2017). Classification of drones based on micro-doppler signatures with dual-band radar sensors. *Proceedings of 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL) : PIERS 2017*. Singapore.

R. Altawy and A. M. Youssef. (2016). Security, privacy, and safety aspects of civilian drones: A survey. *ACM Transactions on Cyber-Physical Systems, 1*, 1 - 25.

Sciancalepore, S., Ibrahim, O., Oligeri, G., & Di Pietro, R. (2019). Detecting drones status via encrypted traffic analysis. *Proc. ACM Workshop Wireless Secur. Mach. Learn. (WiseML).* New York, NY, USA.

T. Daniya, M. Geetha,& Dr, K. Suresh Kumar. (2020). Classification and regression trees with gini index. *Advances in Mathematics Scientific Journal, 9*, 1857 - 8438.

Townsend, J. T. (1971). Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics, 9*, 40-50.

Zhao, L. (2015, 09 05). *Unmanned Aerial Vehicle (UAV) Intrusion Detection Datasets*. (GEORGE MASON UNIVERSITY) 09 25, 2023 tarihinde https://mason.gmu.edu/~lzhao9/materials/data/UAV/ adresinden alındı

ÖZGEÇMİŞ

Kişisel Bilgiler

Ad Soyadı: Tansel ÖZTÜRK

Doğum Yeri / Tarihi:

Uyruk: Türkiye

Eğitim Bilgileri

Yüksek Lisans	İstanbul Sabahattin Zaim Üniversitesi, Bilgisayar Bilimleri Ve Mühendisliği Bilim Dalı, Tezli Yüksek Lisans Öğrencisi 2024, İstanbul
Lisans	İstanbul Teknik Üniversitesi, Metalurji Mühendisliği, 1998, İstanbul

Mesleki Deneyim

01.2022 -	HAVELSAN A.Ş. – Kurumsal Mimari Takım Lideri
12.2019 – 12.2021	ASFAT A.Ş. – Yazılım Sistemleri Proje Yöneticisi
06.2007 – 11.2019	İSBAK A.Ş. – Ar&Ge Yazılım Teknolojileri Şefi
10.2002 – 05.2007	MİKRO ELEKTRONİK Ltd. – Ar&Ge Mühendisi
05.2000 – 09.2002	SENTİM TEKNİK A.Ş. – Sistem Destek Mühendisi
12.1996 – 04.2000	DATATEKNİK A.Ş. – Sistem Destek Mühendisi

Yayın Bilgileri

Ozturk, T., Bayrak, S. (2023). Unmanned Aerial Vehicle Anomaly Detection using Machine Learning Methods, 2nd International Conference on Contemporary Academic Research, Abstract Book of ICCAR 2023, p: 120