T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ TEKNİKLERİ İLE SİBER SALDIRI TESPİTİ

Avni Avnullah Kaşıkçı

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Mart-2026
KONYA

TEZ KABUL VE ONAYI

Avni Avnullah Kaşıkçı tarafından hazırlanan "YAPAY ZEKÂ TEKNİKLERİ İLE SİBER SALDIRI TESPİTİ" adlı tez çalışması 06/03/2026 tarihinde aşağıdaki jüri tarafından oy birliği / oy çokluğu ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Doç. Dr. Murat KÖKLÜ (Selçuk Üniversitesi)	……..
Danışman Doç. Dr. Emine BAŞ (Konya Teknik Üniversitesi)	……..
Üye Doç. Dr. Özgür ÖKSÜZ (Konya Teknik Üniversitesi)	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

Avni Avnullah Kaşıkçı

Tarih:

I. ÖZET

YÜKSEK LİSANS TEZİ

YAPAY ZEKÂ TEKNİKLERİ İLE SİBER SALDIRI TESPİTİ

Avni Avnullah Kaşıkçı

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı
Danışman: Doç. Dr. Emine BAŞ

2026, 118 Sayfa

Jüri
Doç. Dr. Emine BAŞ
Doç. Dr. Özgür Öksüz
Doç. Dr. Murat Köklü

Bilgi teknolojilerinin hızlı gelişimiyle birlikte siber saldırılar, hem bireysel kullanıcıları hem de kurumsal altyapıları tehdit eden temel güvenlik sorunlarından biri hâline gelmiştir. Günümüzde saldırı türlerinin yalnızca sayıca artmakla kalmayıp aynı zamanda daha karmaşık ve dinamik bir yapıya bürünmesi, imza tabanlı ve geleneksel güvenlik yaklaşımlarının güncel tehditleri tespit etmede yetersiz kalmasına yol açmaktadır. Bu durum, makine öğrenimi ve derin öğrenme tabanlı saldırı tespit sistemlerini ağ güvenliği alanında vazgeçilmez hâle getirmiştir. Ancak bu tür modellerin yüksek boyutlu ve karmaşık ağ trafiği verileri üzerinde çalışması, gereksiz veya birbiriyle yüksek korelasyona sahip özelliklerin varlığı nedeniyle artan hesaplama maliyeti, uzun eğitim süreleri ve ölçeklenebilirlik sorunlarını da beraberinde getirmektedir. Bu çalışmada, söz konusu problemlere çözüm üretmek amacıyla CSE-CIC-IDS2018 veri seti üzerinde optimizasyon tabanlı bir özellik seçimi yaklaşımı ele alınmıştır. Özellik seçimi sürecinde, literatürde görece yeni ve siber saldırı tespit alanında sınırlı sayıda çalışmada ele alınmış olan ZOA başta olmak üzere farklı meta-sezgisel yöntemler kullanılmıştır. Yapılan kapsamlı literatür incelemesi, ZOA'nın ağ saldırı tespit sistemlerinde, özellikle derin öğrenme modelleri ile bütünleşik biçimde kullanıldığı çalışmaların oldukça sınırlı olduğunu; gerçek zamanlı uygulanabilirlik ve performans değerlendirmesi açısından ise önemli bir araştırma boşluğu bulunduğunu ortaya koymuştur. Bu yönüyle çalışma, ZOA'nın siber güvenlik bağlamındaki potansiyelini sistematik biçimde inceleyen özgün bir yaklaşım sunmaktadır. Elde edilen özellik alt kümeleri; KNN, RF, CNN, DT, RNN, LSTM ve CNN modelleri kullanılarak değerlendirilmiştir. Deneysel sonuçlar, optimizasyon tabanlı özellik seçiminin yalnızca sınıflandırma doğruluğunu artırmakla kalmadığını, aynı zamanda modellerin öğrenme süreçlerini daha kararlı ve verimli hâle getirdiğini göstermektedir. Özellikle ZOA ile seçilen özellikler kullanıldığında, derin öğrenme modellerinde anlamlı performans iyileşmeleri gözlemlenmiştir. LSTM modelinde doğruluk oranının %99,85'ten %99,94'e, CNN modelinde ise %99,82'den %99,95'e yükselmesi, seçilen özelliklerin temsil gücünü açık biçimde ortaya koymaktadır. RNN modeli ise özellik seçimi sonrasında doğruluk, kesinlik, duyarlılık ve F1 skoru ölçütlerinin tamamında %100 başarıya ulaşmıştır. Makine öğrenimi modelleri açısından değerlendirildiğinde, RF algoritmasının yalnızca 10–12 özellik kullanarak %99,99 doğruluk seviyesine erişmesi, gereksiz özelliklerin elenmesinin sınıflandırma performansı üzerindeki belirleyici etkisini açıkça göstermektedir. Benzer şekilde, MLP modelinin düşük iterasyon sayılarıyla yüksek doğruluk değerlerine ulaşması, ZOA tabanlı özellik seçiminin model öğrenme sürecini kolaylaştırdığını ve aşırı öğrenme riskini azalttığını doğrulamaktadır. Zaman karmaşıklığı açısından yapılan analizler, optimizasyon sonrası modellerde eğitim ve test sürelerinde ortalama %20 civarında bir iyileşme sağlandığını göstermektedir. Ayrıca çalışma kapsamında geliştirilen web tabanlı saldırı tespit uygulaması aracılığıyla eğitilen modeller gerçek zamanlı ağ trafiği üzerinde test edilmiş ve sistemin canlı ortamda potansiyel saldırıları başarılı bir şekilde tespit edebildiği gösterilmiştir. Bu durum, önerilen yaklaşımın yalnızca teorik değil, aynı zamanda pratik ve gerçek dünya uygulamalarına uygun olduğunu ortaya koymaktadır. ZOA tabanlı özellik seçimi yaklaşımının, siber saldırı tespit sistemlerinde hem yüksek doğruluk hem de gerçek zamanlı uygulanabilirlik açısından güçlü ve ölçeklenebilir bir çözüm sunduğu görülmektedir. Bu tez çalışması, literatürde sınırlı biçimde ele alınmış olan ZOA'nın ağ güvenliği alanındaki etkinliğini kapsamlı deneysel analizlerle ortaya

koyarak, yeni nesil optimizasyon algoritmalarının siber güvenlik uygulamalarında kullanılmasına yönelik önemli ve özgün bir katkı sağlamaktadır.

Anahtar kelimeler: Siber güvenlik, Saldırı tespiti, Özellik seçimi, Zebra Optimizasyon Algoritması, Makine öğrenimi, Derin Öğrenme, CSE-CIC-IDS2018, Derin öğrenme, Web tabanlı izleme, Anomali tespiti, Optimizasyon yöntemleri.

II. ABSTRACT

MS THESIS

CYBER ATTACK DETECTION WITH ARTIFICIAL ITELLIGENCE TECHNIQUES

Avni Avnullah Kaşıkçı

Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering

Advisor: Assoc. Prof. Dr. Emine BAŞ

2026, 118 Pages

Jury
Assoc. Prof. Dr. Emine BAŞ
Assoc. Prof. Dr. Özgür Öksüz
Assoc. Prof. Dr. Murat Köklü

With the rapid advancement of information technologies, cyberattacks have become one of the fundamental security challenges threatening both individual users and corporate infrastructures. Today, attack types are not only increasing in number but also evolving into more complex and dynamic structures, rendering signature-based and traditional security approaches insufficient for detecting emerging threats. Consequently, machine learning and deep learning–based intrusion detection systems have become indispensable in the field of network security. However, the application of such models to high-dimensional and complex network traffic data introduces additional challenges, including increased computational cost, prolonged training times, and scalability issues due to redundant or highly correlated features.In this study, an optimization-based feature selection approach is proposed to address these challenges using the CSE-CIC-IDS2018 dataset. During the feature selection process, several meta-heuristic methods were employed, with particular emphasis on the Zebra Optimization Algorithm (ZOA), which is relatively recent in the literature and has been explored in only a limited number of studies within the domain of cyberattack detection. A comprehensive literature review reveals that studies integrating ZOA with deep learning models for network intrusion detection are scarce, and significant research gaps remain in terms of real-time applicability and systematic performance evaluation. In this respect, the present study offers an original and systematic investigation of the potential of ZOA in the context of cybersecurity.The selected feature subsets were evaluated using KNN, RF, CNN, DT, RNN, and LSTM models. Experimental results demonstrate that optimization-based feature selection not only improves classification accuracy but also enhances the stability and efficiency of the learning process. Notably, when features selected by ZOA were utilized, significant performance improvements were observed in deep learning models. The accuracy of the LSTM model increased from 99.85% to 99.94%, while the CNN model's accuracy improved from 99.82% to 99.95%, clearly highlighting the representational strength of the selected features. Furthermore, the RNN model achieved 100% performance across all evaluation metrics, including accuracy, precision, recall, and F1-score, following feature selection.From the perspective of machine learning algorithms, the Random Forest (RF) model achieved an accuracy of 99.99% using only 10–12 features, underscoring the decisive impact of eliminating redundant features on classification performance. Similarly, the MLP model reached high accuracy values with a relatively low number of training iterations, confirming that ZOA-based feature selection facilitates the learning process and reduces the risk of overfitting. Time complexity analysis indicates that the optimized models achieved an average improvement of approximately 20% in training and testing durations.Moreover, a web-based intrusion detection application developed within the scope of this study was used to evaluate the trained models on real-time network traffic. The system successfully detected potential attacks in a live environment, demonstrating that the proposed approach is not only theoretically robust but also practically applicable in real-world scenarios. Overall, the ZOA-based feature selection approach provides a powerful and scalable solution for cyber intrusion detection systems in terms of both high accuracy and real-time applicability. By comprehensively analyzing the effectiveness of ZOA in network security through extensive experimental validation, this thesis makes a significant and original contribution to the application of next-generation optimization algorithms in cybersecurity.

Keywords: Cybersecurity, Intrusion detection, Feature selection, Zebra Optimization Algorithm, Machine learning Deep learning, CSE-CIC-IDS2018, Deep learning, Web-based monitoring, Anomaly detection, Optimization methods.

III. ÖNSÖZ

Türk Dil Kurumu'na göre "optimizasyon" kelimesi, "en uygun duruma getirme" anlamına gelmektedir. Günlük hayatın olağan akışı içerisinde aslında sürekli bir optimizasyon süreci yaşamaktayız. Örneğin, seyahatlerimizi planlarken en uygun rotayı oluşturmaya çalışır, kıyafet seçimimizi hava durumu ve sıcaklığa göre yapar ya da ödev ve sorumluluklarımız için zamanı daha verimli kullanabilmek adına zaman planlaması gerçekleştiririz. Bu tür durumlarda, mevcut koşullar çerçevesinde en uygun seçeneği tercih etmeye çalışırız.

Geleneksel algoritmalar, insan zekâsı ile hesaplanabilen veya belirli kurallara ve bilgilere dayalı girişlerden oluşan küçük boyutlu problemlerin çözümünde etkili olabilmektedir. Ancak, büyük ölçekli veya karmaşık problemlerin çözümü için bilgisayarların gücünden yararlanılmaktadır. Bilim ve teknolojideki gelişmeler sayesinde, geleneksel algoritmalardan daha hızlı olan Tepe Tırmanma Algoritması, Açgözlü Yaklaşım Algoritması, Benzetimli Tavlama Algoritması gibi sezgisel algoritmalar geliştirilmiştir. Bununla birlikte, sezgisel algoritmalar her ne kadar daha hızlı olsalar da lokal optimuma takılma veya problem karşısında yeterli esneklik gösterememe gibi bazı sınırlılıklara sahiptir (Demirci, 2023).

Bu ihtiyaçlardan hareketle geliştirilen meta-sezgisel algoritmalar, arama uzayını daha etkin ve verimli bir şekilde incelemeyi amaçlamaktadır. Meta-sezgisel algoritmalar, sezgisel yöntemlere kıyasla daha hızlı çalışmakta ve belirli bir problem türüne bağımlı olmadan geniş bir yelpazede uygulanabilmektedir. Her ne kadar her zaman küresel en iyi çözümü garanti edemeseler de büyük ölçekli ve karmaşık problemlerde oldukça etkili sonuçlar üretebilmektedirler(Talbi, 2009).

Siber güvenlik, dijitalleşmenin hız kazanmasıyla birlikte yalnızca bilişim sistemlerini değil, aynı zamanda kritik altyapıları, kurumların iş sürekliliğini ve bireysel veri mahremiyetini doğrudan etkileyen stratejik bir alan hâline gelmiştir (Yel ve Atasoy, 2021). Ağ tabanlı saldırıların çeşitlenmesi, saldırı tekniklerinin sürekli evrilmesi ve günlük üretilen veri miktarının hızla artması, saldırı tespit sistemlerinin (Intrusion Detection Systems – IDS) daha güçlü ve esnek yöntemlere ihtiyaç duymasına yol açmaktadır (Taş, 2022). Özellikle gerçek ağ trafiği verilerinin yüksek boyutlu yapıda olması, gürültü ve tekrar içeren özellikler barındırması, sınıfların dengesiz dağılımı ve gerçek zamanlı işlem gereksinimi; saldırı tespitinde hem hesaplama maliyetini artırmakta hem de öğrenme modellerinin genelleme performansını olumsuz etkileyebilmektedir (Topbaş, 2024). Bu nedenle, saldırı tespiti probleminin yalnızca bir sınıflandırma problemi olarak değil, aynı zamanda yüksek boyutlu

veride en ayırt edici özelliklerin belirlenmesini hedefleyen bir optimizasyon problemi olarak ele alınması önem kazanmaktadır.

Bu tez çalışmasında, söz konusu zorlukların giderilmesine katkı sağlamak amacıyla yeni meta-sezgisel yaklaşımlardan biri olan ZOA temel alınmıştır. ZOA, zebraların sürü hâlindeki hareket dinamiklerinden, lider–takipçi etkileşimlerinden ve avcılardan kaçınma davranışlarından esinlenerek geliştirilmiş doğa temelli bir optimizasyon algoritmasıdır (Trojovska vd., 2022). Algoritmanın temelinde, arama uzayında keşif (exploration) ve sömürü (exploitation) süreçlerini dengeleyerek çözüm kalitesini artırma yaklaşımı bulunmaktadır. ZOA, popülasyon tabanlı bir yapı üzerinden aday çözümler üretmekte; sürü davranışı analojisi ile daha iyi çözümlere yönelme ve arama uzayında çeşitliliği koruma mekanizmalarını birlikte işletmektedir.

Bu kapsamda tezde, ZOA'nın siber saldırı tespiti probleminde özellikle özellik seçimi ve buna bağlı olarak sınıflandırma başarımını artırma üzerindeki etkisi incelenmektedir. ZOA ile seçilen özellik alt kümeleri sayesinde veri boyutunun azaltılması, gereksiz/tekrarlı özelliklerin elenmesi ve böylece hem eğitim–test süreçlerinin hızlandırılması hem de modellerin aşırı uyum (overfitting) eğiliminin azaltılması hedeflenmektedir. Elde edilen bulgular, ZOA'nın siber güvenlik alanında saldırı tespit sistemleri için güçlü bir optimizasyon alternatifi sunduğunu; ayrıca problemden bağımsız yapısı nedeniyle farklı optimizasyon gerektiren uygulama alanlarında da değerlendirilebileceğini göstermektedir.

Avni Avnullah Kaşıkçı
KONYA-2026

İçindekiler

IV. SİMGELER VE KISALTMALAR

Kısaltmalar

KNN: k-En Yakın Komşu

RF: Rastgele Orman (Random Forest)

MLP: Çok katmanlı algılayıcı (Multi-Layer Perceptron)

DT: Karar ağacı (Desicion Tree)

CNN: Evrişimsel Sinir Ağı (Convolutional neural network)

RNN: Tekrarlayan Sinir Ağı (Tekrarlayan Sinir Ağı)

LSTM: Uzun Kısa Süreli Bellek (Long short-term memory)

TP: Doğru Pozitif (True Positive)

TN: Doğru Negatif (True Negative)

FP: Yanlış Pozitif (False Positive)

FN: Yanlış Negatif (False Negative)

ZOA: Zebra Optimizasyon Algoritması (Zebra Optimization Algorihm)

AUC: Eğri Altındaki Alan (Area Under the Curve)

ROC: Alıcı İşletim Karekteristiği (Receiver Operating Charecteristic)

SVM: Destek Vektör Makinesi(Support Vector Machine)

1. GİRİŞ

Kelime olarak “en iyi duruma getirme” anlamına sahip olan optimizasyon, temelde sayısal problemleri çözmek için kullanılan matematiksel ilke ve yöntemlerin toplamı olarak ifade edilebilir. Basitçe ifade etmek gerekirse sayısal bir problemin minimum veya maksimum uygunluk değeri, bu değere karşılık gelen değişken veya değişkenlerin bulunması işlemine optimizasyon denilir (Okudan & Şen, 2023).

Sayısal optimizasyon problemlerine örnek olarak, bir sistemin en yüksek verimle çalışabilmesi için bakım ve işletme zamanlarının planlanması, iki nokta arasında en kısa mesafeyi ya da en düşük maliyeti sağlayan güzergâhın belirlenmesi ve haberleşme altyapılarında kapsama alanı maksimum, kurulum maliyeti minimum olacak şekilde verici yerleşimlerinin tasarlanması gösterilebilir. Bu tür problemlerin çözümünde, öncelikle ulaşılmak istenen hedef veya hedeflerin açık biçimde tanımlanması ve bu hedeflerin matematiksel ifadelerle modellenmesi gerekmektedir. Problemin olası tüm çözümlerini kapsayan alan çözüm uzayı olarak adlandırılmaktadır. Teorik olarak sınırsız büyüklükteki bir çözüm uzayında, doğrudan en uygun çözümün elde edilmesi mümkün değildir. Bu nedenle, optimizasyon problemleri belirli kısıtlar altında ele alınmakta; söz konusu kısıtlar genellikle eşitlik ve eşitsizlikler şeklinde ifade edilmektedir. Tanımlanan kısıtlar altında, amaç fonksiyonunun maksimum veya minimum değerini sağlayan çözümler aranmaktadır. Amaç fonksiyonunun tüm çözüm uzayı içerisindeki en yüksek veya en düşük değerini veren çözüm küresel optimum olarak tanımlanırken, yalnızca belirli bir bölge içerisinde optimum görünen ancak genel çözüm uzayı açısından en iyi olmayan çözümler yerel optimum olarak ifade edilmektedir.

Optimizasyon problemleri tanımlarına göre sınıflandırılırlar. Bir problemde eğer amacımız en yüksek değeri bulmak ise o problem bir en üst düzeye çıkarma (maksimizasyon) problemidir. Tersi şekilde en düşük değeri bulmamız isteniyorsa o problem minimizasyon problemi olur. Eğer problemde sadece değişken sınırları varsa bu tür problemler sınırlamasız problemler olarak adlandırılırken amaç fonksiyonu sınırlayan başka fonksiyonlar varsa sınırlamalı fonksiyon adını alır. Problemin tasarımında kullanılan değişkenlerin türü problemin sınıflandırılmasında rol oynar. Problemin yapısında yer alan değişkenlerin ayrık nitelikte olması; nesnelerin sınıflandırılması, sıralanması, seçilmesi ya da gruplanması gibi işlemleri içermesi durumunda, söz konusu problem ayrık (kesikli) optimizasyon problemi olarak tanımlanmaktadır. Buna karşılık, değişkenlerin belirli bir aralık içerisinde sürekli değerler alabilmesi halinde, bu tür problemler sürekli optimizasyon problemleri olarak sınıflandırılmaktadır. Bir problemde tek bir küresel çözüm var ancak hiç yerel çözüm yoksa bu problem tek modludur. Eğer problem bünyesinde tek bir küresel çözüm ve bir veya birden fazla

yerel çözüm barındırıyorsa çok modludur. Aynı şekilde problem bünyesindeki yerel çözüm miktarına bakılmaksızın birden fazla küresel çözüm bulunduruyorsa yine çok modludur denilir (Demirci, 2023).

Siber güvenlik, dijitalleşmenin hızla artmasıyla birlikte günümüzün en kritik araştırma ve uygulama alanlarından biri hâline gelmiştir. Kurumların, devletlerin ve bireylerin bilgi sistemlerine olan bağımlılığının artması, ağ tabanlı saldırıların hem sayısal hem de yapısal olarak daha karmaşık bir hâl almasına neden olmuştur. Özellikle hizmet engelleme saldırıları, yetkisiz erişim girişimleri, kötü amaçlı yazılımlar ve gelişmiş sürekli tehditler (APT) gibi saldırı türleri, geleneksel güvenlik mekanizmalarının yetersiz kalmasına yol açmaktadır. Bu durum, siber saldırıların daha erken, daha doğru ve daha düşük hesaplama maliyetiyle tespit edilebileceği akıllı ve uyarlanabilir saldırı tespit sistemlerine olan ihtiyacı artırmıştır. Bu tez çalışmasında, söz konusu ihtiyaca yanıt vermek amacıyla ZOA gibi meta-sezgisel algoritmalar, siber güvenlik alanındaki optimizasyon problemlerine çözüm sunmak üzere derinlemesine incelenmiştir.

Çalışma kapsamında ZOA, özellikle özellik seçimi problemi üzerine odaklanılarak ele alınmış; yüksek boyutlu ağ trafiği verileri içerisinden en ayırt edici özelliklerin belirlenmesi hedeflenmiştir. Bu yaklaşım sayesinde, gereksiz ve tekrarlı özellikler elenmiş, veri boyutu önemli ölçüde azaltılmış ve böylece sınıflandırma sürecinde ortaya çıkan yüksek hesaplama maliyetlerinin düşürülmesi sağlanmıştır. Tez kapsamında, ZOA'nın ikili optimizasyon problemlerine uyarlanabilmesi için transfer fonksiyon tabanlı bir yaklaşım benimsenmiş ve algoritmanın ikili arama uzayındaki başarımı ayrıntılı olarak değerlendirilmiştir. ZOA ile seçilen özellik alt kümeleri, KNN, RF, DT ve MLP gibi klasik makine öğrenmesi algoritmalarının yanı sıra, LSTM, RNN ve CNN gibi derin öğrenme modelleri ile test edilmiştir. Elde edilen sonuçlar, ZOA tabanlı özellik seçiminin hem klasik hem de derin öğrenme tabanlı modellerin performansını anlamlı ölçüde artırdığını ortaya koymuştur. Bununla birlikte, çalışma yalnızca deneysel analizlerle sınırlı kalmamış; tez kapsamında geliştirilen web tabanlı gerçek zamanlı saldırı tespit sistemi ile önerilen yaklaşımın pratikte uygulanabilirliği de gösterilmiştir. Geliştirilen sistem, ağ trafiğini belirli zaman aralıklarında otomatik olarak yakalayarak CSE-CIC-IDS2018 veri seti formatına dönüştürmekte ve eğitilmiş yapay zekâ modelleri aracılığıyla anlık saldırı tespiti gerçekleştirmektedir. Bu yönüyle çalışma, ZOA tabanlı optimizasyon yaklaşımının yalnızca akademik düzeyde değil, gerçek dünya siber güvenlik senaryolarında da etkin bir şekilde kullanılabileceğini ortaya koymaktadır. Sonuç olarak, bu tez çalışması; ZOA'nın siber güvenlik alanındaki özellik seçimi ve saldırı tespit problemlerinde güçlü, esnek ve uygulanabilir bir çözüm sunduğunu göstermektedir. Elde edilen

bulgular, ZOA'nın yalnızca siber güvenlik alanında değil, yüksek boyutlu ve karmaşık optimizasyon problemlerinin bulunduğu farklı disiplinlerde de önemli katkılar sağlayabilecek bir meta-sezgisel algoritma olduğunu ortaya koymaktadır (Demirci, 2023).

1.1. Meta-sezgisel Algoritmalar

Gerçek dünya optimizasyon problemleri karmaşık ve çözümü zor olarak nitelendirilen bir problemlerdir. Bu tür problemleri geleneksel yöntemlerle çözmeye çalışmak zaman alıcıdır. Ayrıca bir tür problemleri çözmek için tasarlanmış bir geleneksel yöntem benzer başka bir problemde çalışmayabilir. Bu sorunu aşmak için yürütme zamanı açısından daha hızlı ancak tam sonucu garanti etmeyen algoritmalar geliştirilmiştir. Bu tür algoritmalara meta sezgisel algoritmalar denilir (Reddy ve diğ., 2023).

Meta-sezgisel algoritmalar genellikle doğadaki olgulardan esinlenerek tasarlanmıştır ve meta-sezgisel algoritmaları esinlendikleri olgulara göre kategorize edecek olursak; Sürü Zekâsı Algoritmaları, Fizik Temelli Algoritmalar, Evrim (Biyoloji) Temelli Algoritmalar ve Sosyal Olaylardan İlham Alan Algoritmalar olmak üzere dört gruba ayırabilmekteyiz (Jia vd., 2021). Sürü zekâsı algoritmalarına örnek verecek olursak Parçacık Sürü Optimizasyonu (PSO) kuşların kolektif yiyecek arama davranışından esinlenmiştir. Fizik temelli Algoritmalara örnek verecek olursak Yerçekimi arama algoritması yerçekimi kanunu temel alan bir optimizasyon algoritmasıdır (Rashedi vd., 2009). Her bir ajan popülasyonu oluşturan bir birey olarak düşünülmüş ve ajanların ağırlıkları sonuçlarla ilişkilendirilmiştir. Sonuç elde etme yöntemi ise yerçekimi kuvvetinin ajanların ağırlıklarına göre farklı düzeyde etki etmesi ve daha ağır kütleli nesnelere doğru hareket etmesiyle oluşmuştur. Evrim temelli Algoritmalara örnek verecek olursak Genetik Algoritması (GA) doğadaki en güçlü olanın hayatta kalması ilkesinden yola çıkarak üremedeki gen birleşimlerinden esinlenmiştir (Holland, 1992). Sosyal olaylardan ilham alan algoritmalar, insanın davranış alışkanlıklarından ilham alan meta-sezgisel algoritmalardır ve örnek verecek olursak mükemmel bir uyum sürecini bulmak için müzik performansını simüle eden Harmoni Arama (HS) ile ünlüdür (Geem vd., 2001).

Genel olarak meta sezgisel algoritmaların iki temel arama yapısı bulunmaktadır. Bunlar küresel arama ve yerel aramadır. Yerel arama, çözüm uzayının belirli bir alanındaki en iyi çözümleri ararken; küresel arama, genel en iyi çözümü bulmayı amaçlamaktadır (Hart, 1994). Bütün algoritmalar, zor optimizasyon problemlerinde yerel uç noktalara (ekstremumlara) takılabilir. Önemli olan, algoritmanın içinde bulunduğu çözümün bir yerel uç nokta olduğunun farkına varabilmesi ve bu noktadan uzaklaşarak arama uzayında küresel uç noktaya ya da farklı umut vaat eden bölgelere yönelebilmesidir (Demirci, 2023).

Algoritmaların yerel uç noktalardan kaçınmak için kullandıkları mekanizmalar algoritmaya özgüdür ve bu mekanizmalar, algoritmanın performansını belirleyen en önemli faktörlerden biridir. Genel olarak meta sezgisel algoritmaların yapıları beş aşamadan oluşur. Bunlar sırasıyla başlangıç çözümü oluşturma, küresel arama, yerel arama, yerel uç noktalardan kurtulma ve aramayı durdurma mekanizmasıdır. Algoritma çalışırken bu sırayı takip eder, eğer ki aramayı durdurma şartını sağlanmamışsa algoritma küresel arama aşamasına dönerek bu döngüyü tekrar başlatır (Abdel-Basset ve diğ., 2018).

1.2. Siber Güvenlik

Siber güvenlik, bilgi sistemlerinin gizlilik, bütünlük ve erişilebilirlik ilkelerini korumayı amaçlayan disiplinler arası bir araştırma ve uygulama alanıdır. Günümüzde dijitalleşmenin hızla artmasıyla birlikte kamu kurumları, özel sektör kuruluşları ve bireyler bilgi teknolojilerine giderek daha fazla bağımlı hâle gelmiştir. Bu durum, bilgisayar ağları ve ağ tabanlı uygulamalar üzerinden gerçekleştirilen siber saldırıların hem sayısal hem de yapısal olarak artmasına neden olmuş; siber güvenliği yalnızca teknik bir konu olmaktan çıkararak stratejik ve kritik bir gereklilik hâline getirmiştir. Nitekim 2019 yılında CNBC tarafından yayımlanan bir raporda, bir siber saldırının işletmelere ortalama maliyetinin yaklaşık 200.000 ABD doları olduğu belirtilmiştir (CNBC, 2019). Siber saldırıların artışıyla birlikte, bu saldırıların doğru ve zamanında tespit edilmesi büyük önem kazanmıştır. Saldırı tespiti (intrusion detection), bir bilgi sistemine zarar verebilecek ya da sistemi tehlikeye atabilecek saldırıların belirlenmesi süreci olarak tanımlanmaktadır. Bu amaçla geliştirilen Saldırı Tespit Sistemleri (Intrusion Detection Systems – IDS), ana makine tabanlı (host-based), ağ tabanlı (network-based) ya da her iki yaklaşımın birleşimi şeklinde sınıflandırılmaktadır (Singh vd., 2014). Ana makine tabanlı IDS'ler, bir sistemin içsel durumunu izleyerek günlük (log) analizi, dosya bütünlüğü denetimi ve kayıt defteri takibi gibi işlemleri gerçekleştirirken; ağ tabanlı IDS'ler, ağ trafiğini analiz ederek Hizmet Engelleme (DoS), Dağıtık Hizmet Engelleme (DDoS), SQL enjeksiyonu ve parola saldırıları gibi tehditleri tespit etmeyi amaçlamaktadır (Leevy vd., 2022).

IDS sistemleri ayrıca imza tabanlı ve anomali tabanlı olmak üzere iki temel yaklaşım altında ele alınmaktadır. İmza tabanlı IDS'ler, bilinen saldırılara ait örüntülere dayalı olarak çalışmakta ve daha önce karşılaşılmamış saldırı türlerini tespit edememektedir. Bu nedenle, imza tabanlı sistemlerin veritabanlarının sürekli olarak güncellenmesi gerekmektedir. Buna karşılık, anomali tabanlı IDS'ler, normal ağ trafiği davranışından sapmaları tespit etmeyi hedeflemektedir. Makine öğrenmesi ve derin öğrenme yöntemlerinin, özellikle anomali

tespitinde başarılı sonuçlar sunabilmesi nedeniyle, anomali tabanlı saldırı tespiti siber güvenlik alanında önemli bir araştırma konusu hâline gelmiştir (Patil vd., 2017).

Anomali tabanlı saldırı tespiti çalışmalarında kullanılan veri setleri, modellerin başarımını doğrudan etkileyen temel unsurlardan biridir. CSE-CIC-IDS2018 gibi veri setleri, ağ trafiği üzerinde saldırı tespitine yönelik kestirimci modellerin eğitilmesi amacıyla oluşturulmuştur. Bu veri seti, modern ve gerçekçi ağ trafiği üretimini hedefleyen bir projenin parçası olup, önceki çalışmalarda kullanılan ISCXIDS2012 ve CICIDS2017 veri setlerinin geliştirilmiş bir sürümüdür (Saxena vd., 2017). ISCXIDS2012 veri seti, yedi günlük ağ trafiği üzerinden oluşturulmuş; DoS, DDoS ve kaba kuvvet saldırıları gibi tehditleri içermesine rağmen tüm saldırıları tek bir "saldırı" etiketi altında toplamıştır (Sharafaldin vd., 2018). Ayrıca, saldırı örneklerinin toplam veri içindeki oranının düşük olması ve sınıf dengesizliği içermesi, bu veri seti üzerinde yapılan çalışmaların genellenebilirliğini sınırlandırmıştır. Bu eksiklikleri gidermek amacıyla geliştirilen CICIDS2017 veri seti, daha fazla saldırı türü ve daha yüksek özellik sayısı içermesine rağmen, hâlen sınıf dengesizliği ve gerçek dünya trafiğini yeterince temsil edememe gibi sınırlılıklara sahiptir (Shiravi vd., 2012). 2018 yılında Kanada İletişim Güvenliği Kuruluşu'nun (Communications Security Establishment – CSE) projeye dahil olmasıyla yayımlanan CSE-CIC-IDS2018 veri seti ise, çok daha büyük bir istemci–hedef ve saldırı makinesi altyapısı kullanılarak oluşturulmuş; 10 günlük ağ trafiğinden elde edilen yaklaşık 16 milyon örnek içermektedir (D'hooge vd., 2019). Veri seti, çok sayıda saldırı türü, yüksek boyutlu özellik uzayı ve sınıf dengesizliği barındırması nedeniyle, gerçekçi ve zorlu bir saldırı tespit ortamı sunmaktadır (He & Garcia, 2009) ancak CSE-CIC-IDS2018 gibi büyük ve karmaşık veri setleri üzerinde gerçekleştirilen çalışmalar incelendiğinde, literatürde bazı önemli eksikliklerin bulunduğu görülmektedir. Mevcut çalışmaların birçoğunda rapor edilen performans değerlerinin olağan dışı derecede yüksek olması, aşırı uyum (overfitting) riskine işaret etmektedir. Ayrıca, sınıf dengesizliği probleminin ve veri temizleme sürecinin çoğu çalışmada yeterince ele alınmadığı görülmektedir. Bu durum, deneylerin yeniden üretilebilirliğini ve gerçek dünya senaryolarına uygulanabilirliğini olumsuz yönde etkilemektedir (Ferrag vd., 2020).

Bu tez çalışması, söz konusu eksiklikleri göz önünde bulundurarak siber güvenlik alanındaki saldırı tespit problemini bir optimizasyon problemi olarak ele almakta ve ZOA gibi meta-sezgisel yaklaşımlarla çözmeyi hedeflemektedir. Özellikle yüksek boyutlu ağ trafiği verilerinde özellik seçimi probleminin ele alınması hem hesaplama maliyetlerinin azaltılması hem de sınıflandırma performansının artırılması açısından kritik bir ön işleme adımı olarak

değerlendirilmektedir (Abdelmalek vd., 2024a). Bu bağlamda, ZOA tabanlı yaklaşımın siber saldırı tespit sistemlerine katkısı, tez kapsamında ayrıntılı olarak incelenmiştir.

1.3.Tezde Kullanılan CSE-CIC-IDS2018 Veri Seti Detayları

CSE-CIC-IDS2018 veri seti, Kanada Siber Güvenlik Enstitüsü (CIC) ve İletişim Güvenliği Kurumu (CSE) tarafından, ağ tabanlı anomali tespiti ve siber tehdit analizi amacıyla oluşturulmuş, siber güvenlik alanında yaygın olarak kullanılan bir veri kümesidir. Gerçek dünya ağ trafiğini modelleyen bu kapsamlı veri seti, zararsız trafik ile birlikte Brute-force, DoS, DDoS, Botnet gibi 15 farklı saldırı türünü içermektedir. 150 GB boyutundaki orijinal veri seti, işlenmiş haliyle 6.9 GB büyüklüğünde olup, AWS Open Data üzerinden erişilebilir durumdadır. Veri seti, 80 farklı öznitelik barındırmakta olup, "Dst Port", "Protocol", "Flow Duration" gibi kritik değişkenleri içermektedir. Etiketli ağ verisi sağlaması nedeniyle, makine öğrenimi tabanlı anomali tespit sistemlerinin geliştirilmesi ve saldırı analizi çalışmalarında önemli bir referans noktasıdır. Teknik altyapı açısından, Python, PyTorch ve TensorFlow gibi kütüphanelerle analiz edilen veri seti, DDoS attack-HOIC (686,012 kayıt), DoS attacks-Hulk (461,912 kayıt) ve DDoS attacks-LOIC-HTTP (576,191 kayıt) gibi çeşitli saldırı türlerini içeren etiketli kayıtlarla zenginleştirilmiştir. Ayrıca, veri ön işleme aşamalarından geçirilerek makine öğrenimi algoritmalarının kolay uygulanabilmesi için optimize edilmiştir. Özellikle Flow Duration (Akış Süresi) ve Tot Fwd Pkts (Toplam İleri Paket Sayısı) gibi değişkenler, ağ trafiğinin karakterizasyonunda kritik bir rol oynarken, saldırı tespiti süreçlerinde temel öznitelikler olarak değerlendirilmektedir. CSE-CIC-IDS2018, geniş kapsamı ve yüksek kaliteli verileriyle, siber güvenlik sistemlerinin geliştirilmesi, değerlendirilmesi ve saldırı türlerinin modellenmesi için güçlü bir test ortamı sunmaktadır. Aşağıda Çizelge 1.1. CSE-CIC-IDS2018 veri setinin saldırırı türlerine göre dağılımı verilmiştir. Bu dağılımlar incelendiğinde veri setindeki dengesizlikler tespit edilmiştir.

Çizelge 1.1 CSE-CIC-IDS2018 veri setinin saldırı türlerine göre dağılımı (Leevy vd., 2022)

Saldırı Türü	**Toplam Kayıt**
Benign	13484708
FTP-BruteForce	193360
SSH-Bruteforce	187589
DoS attacks-GoldenEye	41508
DoS attacks-Slowloris	10990
DoS attacks-Hulk	461912
DoS attacks-SlowHTTPTest	139890
DDoS attacks-LOIC-HTTP	576191
DDOS attack-HOIC	686012
DDOS attack-LOIC-UDP	1730
Brute Force -Web	611

Brute Force -XSS	230
SQL Injection	87
Infilteration	161934
Bot	286191
Benign	13484708
FTP-BruteForce	193360
Toplam	16232943

Bu çalışmada, CSE-CIC-IDS2018 veri setinin işlenmesi ve optimize edilmesi sürecinde uygulanan kapsamlı veri ön işleme adımlarıyla, literatürdeki benzer çalışmaların ötesine geçerek siber saldırı tespitinde daha dengeli ve verimli bir veri kümesi sunmaktadır. Öncelikle, veri setinin büyük boyutlu ve dengesiz yapısı dikkate alınarak detaylı bir analiz gerçekleştirilmiş, saldırı ve normal trafik arasındaki dağılım incelenmiştir. Daha önce yapılan çalışmalarda (García vd., 2015; Roy vd., 2018), sınıf dengesizliğinin makine öğrenimi modellerinin performansı üzerinde olumsuz etkiler yarattığı gösterilmiş olup, bu çalışmada her sınıftan maksimum 10.500 örnek seçilerek veri setinin dengeli hale getirilmesi sağlanmıştır. Bu sayede, azınlık sınıflarının model tarafından göz ardı edilmesinin önüne geçilmiş ve tüm saldırı türlerinin eşit temsil edildiği bir yapı oluşturulmuştur.

Bunun yanı sıra, veri temizleme aşamasında gereksiz veya öğrenme sürecine katkı sağlamayan değişkenlerin (özelliklerin) belirlenerek çıkarılması, veri setinin gereksiz yüklerden arındırılmasına katkı sağlamıştır (Chu vd., 2016). Özellikle, Flow ID, Src IP ve Dst IP gibi değişkenlerin kaldırılması, ağ izleme sürecinde anlamlı olmasına rağmen makine öğrenimi modellerinin karar verme mekanizmasına doğrudan katkı sağlamadığı için veri setinin daha verimli kullanılmasına olanak tanımıştır. Aynı zamanda, eksik ve hatalı verilerin temizlenmesi, veri setinin doğruluğunu artırarak modelin güvenilirliğini iyileştirmiştir (Rossmann & Van Beek, 1999).

Özellik seçimi ve veri optimizasyonu kapsamında uygulanan Pearson korelasyon analizi, yüksek derecede ilişkili olan değişkenlerin belirlenerek yalnızca en anlamlı değişkenlerin modele dahil edilmesini sağlamıştır. Bu işlem, modelin öğrenme sürecini hızlandırmakla kalmamış, aynı zamanda aşırı öğrenmenin (overfitting) önüne geçerek genelleme performansını artırmıştır (El-Hasnony vd., 2020). Daha önceki çalışmalarda yüksek korelasyonlu değişkenlerin model doğruluğunu olumsuz etkilediği belirtilmiş olup (Piao vd., 2015), bu çalışmada belirlenen 80 özelliğin 38'e indirilmesiyle modelin daha dengeli ve verimli çalışması sağlanmıştır.

Ayrıca, saldırı türlerinin belirli bir öncelik sırasına göre ele alınarak veri setinden bazı saldırı türlerinin çıkarılması, literatürde sıkça karşılaşılan veri fazlalığı ve düzensizliğin önüne geçmiştir. Özellikle, az sayıda örneğe sahip olan saldırı türlerinin çıkarılması, modelin yanlış öğrenme yapısını engelleyerek genelleme kapasitesini artırmıştır (Roy vd., 2018). Bunun yanı sıra, saldırı türlerinin LabelEncoder kullanılarak sayısal değerlere dönüştürülmesi, makine öğrenimi algoritmalarının veri setiyle daha etkin bir şekilde çalışmasını sağlamıştır.

ZOA kullanılarak özellik seçimi gerçekleştirilmiş ve belirlenen en iyi özellikler dört farklı sınıflandırıcı ile test edilmiştir. Bu kapsamda, KNN, MLP, RF, DT, LSTM, RNN ve CNN algoritmaları kullanılarak sınıflandırma deneyleri gerçekleştirilmiş ve ZOA ile optimize edilen özelliklerin model performansına etkisi ayrıntılı bir şekilde analiz edilmiştir. ZOA algoritması, transfer fonksiyonu kullanılarak optimize edilmiş ve gereksiz özellikler elenerek en anlamlı özelliklerin seçilmesi sağlanmıştır. Özellik seçimi sonrasında, sınıflandırma performansı
En iyi uygunluk değeri (Best Fitness), en kötü uygunluk değeri (Worst Fitness), medyan uygunluk değeri (Median Fitness), ortalama uygunluk değeri (Mean Fitness), standart sapma uygunluk değeri (Standard Deviation Fitness), ortalama doğruluk (Mean Accuracy), ortalama seçilen özellik sayısı (Mean Feature Selection) ve ortalama işlem süresi (Mean Time) gibi metrikler ile değerlendirilmiştir.Her bir sınıflandırıcı, sonuçların güvenilirliğini test etmek amacıyla 20 farklı tekrar ile çalıştırılmış ve algoritmaların istikrarı analiz edilmiştir. Elde edilen bulgular, ZOA ile optimize edilen özellik seçiminin sınıflandırma performansını artırırken aynı zamanda işlem süresini önemli ölçüde azalttığını göstermektedir. Çalışmanın devamında KNN, RF, MLP, DT algoritmaların sonucunda oluşan en iyi özellikler çıkartılarak derin öğrenme modellerine verilmiştir. Bu doğrultuda, CNN, LSTM, RNN derin öğrenme algoritmaları ile sınıflandırma performansı ileri düzeye çıkartılmıştır.

Bu çalışma, literatürde yaygın olarak kullanılan CSE-CIC-IDS2018 veri setine yönelik derinlemesine bir ön işleme süreci sunarak, siber güvenlik alanında daha güvenilir ve dengeli makine öğrenimi ve derin öğrenme modellerinin geliştirilmesine katkıda bulunmaktadır. Önerilen veri işleme teknikleri, modelin eğitim sürecini iyileştirerek saldırı tespit performansını artırmakta ve siber tehditlerin daha etkili bir şekilde belirlenmesine olanak tanımaktadır.

1.4. Problem Tanımı

Siber güvenlik, her geçen gün daha büyük tehditlerle karşı karşıya kalmaktadır. Gelişen teknolojiyle birlikte, siber saldırılar daha sofistike ve karmaşık hale gelirken, bu saldırıları tespit etmek için geliştirilen yöntemler de önem kazanmaktadır. Özellikle, ağ tabanlı saldırıların tespiti, siber güvenlik alanında önemli bir araştırma konusu haline gelmiştir. Ancak, siber saldırı tespiti yapmak için kullanılan veri setleri genellikle birçok zorlukla karşı karşıyadır. Bunlar arasında en belirgin olanlar, veri setlerinin dengesiz sınıf dağılımına sahip olması, yani normal trafik (Benign) sınıfının büyük bir çoğunluğu oluşturması, saldırı sınıflarının ise az sayıda örnek içermesidir. Bu tür dengesizlikler, makine öğrenimi algoritmalarının doğru bir şekilde saldırıları tespit etmesini zorlaştırmaktadır (García vd., 2015). Örneğin, CSE-CIC-IDS2018 veri setinde Benign sınıfı, toplam verinin %80'inden fazlasını oluştururken, bazı saldırı türleri yalnızca birkaç bin örnekle sınırlıdır. Bu durum, dengesiz veri setleri üzerinde yapılan modellemelerde genellikle yanlış sınıflandırmalara neden olabilir (Roy vd., 2018).

Bir diğer önemli problem, yüksek boyutlu veri setlerinde yer alan gereksiz özellikler ve yüksek korelasyonlu sütunlardır. Özellikle, veri setindeki bazı özellikler birbirleriyle yüksek korelasyona sahip olabilir ve bu durum, modelin doğruluğunu olumsuz yönde etkileyebilir. Ayrıca, bazı sütunlar ağ izleme sırasında kimlik belirleme amacıyla kullanıldığından, makine öğrenimi için anlamlı bilgi taşımamaktadır. Bu tür gereksiz özellikler, modelin öğrenme sürecini karmaşıklaştırabilir ve işlem yükünü artırabilir (El-Hasnony vd., 2020). Bu tür zorlukları aşmak için, veri setlerinden gereksiz ve yüksek korelasyonlu sütunlar çıkarılmalıdır. Bu adım hem modelin eğitim süresini kısaltır hem de doğruluğunu artırır.

Bunun yanı sıra, siber güvenlik alanında saldırı türlerinin sınıflandırılması da büyük önem taşımaktadır. Bazı saldırı türleri, veri setlerinde çok az sayıda örneğe sahip olup, bu durum modelin doğru bir şekilde bu saldırıları tespit etmesini zorlaştırır. Bu tür saldırıların sınıf dengesizliği oluşturduğundan, sınıf dengesini sağlamak amacıyla örnekleme işlemleri uygulanması gereklidir. Ayrıca, bazı saldırı türleri analiz açısından öncelikli olmayabilir ve bu nedenle veri setinden çıkarılabilir. Bu işlemler, veri setinin daha anlamlı hale gelmesini sağlar ve analiz sürecini kolaylaştırır (Chu vd., 2016).

Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen ön işleme adımları, siber saldırıların tespitinde karşılaşılan bu zorlukların üstesinden gelmek için tasarlanmıştır. Veri setindeki dengesizlikler, gereksiz özellikler ve düşük örnek sayısına sahip saldırı türleri gibi problemler çözüme kavuşturulmuş, böylece daha dengeli ve verimli bir veri seti oluşturulmuştur. Kaynak araştırmaları, siber güvenlik ve makine öğrenimi alanlarındaki

literatürde bu tür optimizasyonların ne kadar önemli olduğunu göstermektedir. Verilen kaynaklarda, saldırı tespit sistemlerinin etkinliğini artırmak için kullanılan veri işleme yöntemlerinin literatürle uyumlu olarak uygulandığı ve bu sayede saldırı tespitinde daha yüksek başarı oranlarına ulaşılabileceği belirtilmiştir (Piao vd., 2015). Bu süreç, makine öğrenimi modellerinin eğitiminde kullanılan veri setlerinin kalitesinin arttırılması açısından önemli bir adım olarak değerlendirilmektedir.

1.5.Tezin Amacı ve Hedefleri

Bu çalışmanın temel amacı, meta sezgisel algoritmalar kullanarak siber güvenlik alanında kapsamlı bir analiz gerçekleştirmek, optimizasyon algoritmalarının saldırı tespit sistemlerindeki teorik ve uygulamalı rollerini incelemek ve literatürdeki boşluklara katkı sağlamaktır. Günümüzde siber saldırılar, bireylerden kurumlara kadar geniş bir yelpazede ciddi güvenlik tehditleri oluşturmakta ve bu tehditlerle etkin mücadele edebilmek için doğru özellik seçimi ile güçlü sınıflandırma yöntemlerinin bir arada kullanılması gerekmektedir. Bu bağlamda, bu çalışmada siber güvenlik araştırmalarında yaygın olarak kullanılan, geniş kapsamlı ve gerçeğe yakın bir veri kaynağı olan CSE-CIC-IDS2018 veri seti kullanılarak son yıllarda yeni önerilmiş bir meta sezgisel algoritma olan ZOA saldırı tespitindeki etkinliği değerlendirilmiştir. Özellikle veri boyutunu küçültme, sınıflandırma performansını artırma, hesaplama maliyetlerini düşürme ve doğruluk oranını yükseltme gibi kritik faktörler üzerine ayrıntılı analizler yapılmıştır.

Tez kapsamında, ZOA'nın siber güvenlik alanındaki rolü derinlemesine incelenmiş ve algoritmanın saldırı tespitindeki başarımı farklı meta-sezgisel yöntemlerle karşılaştırmalı olarak değerlendirilmiştir. Özellik seçimi sürecinden önce, veri setinin yüksek boyutluluğunun olası olumsuz etkilerini azaltmak amacıyla kapsamlı bir veri küçültme ve ön işleme süreci uygulanmıştır. Bu süreçte, eksik veya hatalı değer içeren kayıtlar temizlenmiş, saldırı tespitine katkısı sınırlı olan gereksiz özellikler veri setinden çıkarılmıştır. Ayrıca, özellikler arasındaki ilişkiler korelasyon analizleri ile incelenmiş; düşük korelasyona sahip veya yüksek oranda tekrarlayan bilgiyi temsil eden sütunlar elenerek veri setinin daha temsil edici ve sade bir yapıya kavuşturulması sağlanmıştır. Bu ön işlemler sayesinde hem optimizasyon algoritmalarının arama uzayı daraltılmış hem de sınıflandırma modellerinin daha kararlı ve verimli şekilde eğitilmesi mümkün hâle gelmiştir. Çalışmada, ZOA'nın veri boyutunu küçültme, hesaplama maliyetlerini azaltma ve sınıflandırma doğruluğunu artırma konusundaki etkileri ayrıntılı olarak

analiz edilmiştir. Ayrıca bu çalışma, yalnızca siber güvenlik alanına değil, optimizasyon gerektiren farklı problem alanlarına da katkı sağlamayı amaçlamakta ve ZOA'nın geniş bir uygulama yelpazesinde kullanım potansiyelini ortaya koymaktadır (Demirci, 2023).

Bu çalışma kapsamında bir web sitesi geliştirilmiş ve ağ üzerindeki veri trafiği analiz edilerek gerçek zamanlı saldırı durumları test edilmiştir. Geliştirilen web tabanlı yapı üzerinden elde edilen ağ trafiği verileri, CSE-CIC-IDS2018 veri seti ile uyumlu olacak şekilde işlenmiş ve analiz süreçlerinde kullanılmıştır. Ağ trafiği, Wireshark program aracı yardımıyla yakalanmış; elde edilen ham veriler gerekli ön işleme adımlarından geçirilerek saldırı tespitine uygun hâle getirilmiştir. Ayrıca, geliştirilen web tabanlı uygulama aracılığıyla gerçek zamanlı siber saldırıların tespit edilebildiği bütünleşik bir sistem ortaya konulmuştur.

2. LİTERATÜR TARAMASI

Bu bölümde, CSE-CIC-IDS2018 veri seti üzerine yapılmış olan araştırmalar ve kullanılan yaklaşımlar incelenmiştir. CSE-CIC-IDS2018, siber güvenlik tehditlerinin tespiti ve sınıflandırılmasında önemli bir veri seti olup, birçok araştırmacı bu veri setini kullanarak farklı makine öğrenimi ve derin öğrenme algoritmalarının performanslarını değerlendirmiştir. Yapılan bu çalışmalar, kullanılan modellerin doğruluk, hassasiyet, geri çağırma gibi metriklerle karşılaştırılmasını sağlamakta ve siber güvenlikte daha etkili çözümler geliştirmek için bir temel oluşturmaktadır.

Yazarlar, CSE-CIC-IDS2018 veri setini kullanarak çeşitli algoritmalar ve yöntemler üzerinde çalışmışlardır. Bu çalışmalarda genellikle derin öğrenme modelleri, ensemble yöntemleri ve klasik makine öğrenimi algoritmaları tercih edilmiştir. Örneğin, (Atefinia & Ahmadi, 2021) tarafından önerilen "Modular Deep Neural Network" modelinin başarılı olduğu görülürken (Basnet vd., 2019), "MLP" ve "Deep Autoencoder" gibi derin öğrenme yöntemlerini kullanarak yüksek doğruluk oranlarına ulaşmışlardır. Diğer taraftan Ferrag ve arkadaşlarının (Ferrag vd., 2020) çalışmasında RF ve "Logistic Regression + DT" kombinasyonunu kullanarak dikkat çekici sonuçlar elde etmiştir. Bu çeşitlilik, farklı algoritmaların ve parametrelerin siber güvenlik alanındaki başarısını test etmek ve karşılaştırmak açısından önemli fırsatlar sunmaktadır.

Çalışmalar, sadece kullanılan algoritmalarla sınırlı kalmayıp, aynı zamanda bu algoritmaların hangi ortamda çalıştığına dair de değerli bilgiler sunmaktadır. Python, Scikit-learn, TensorFlow, Keras gibi yaygın yazılım çerçeveleri ve araçlar, siber güvenlik araştırmalarında sıklıkla kullanılmaktadır. Ayrıca, bazı araştırmalar GPU kullanımına veya Google Colab gibi bulut tabanlı platformlara odaklanarak hesaplama gücünden yararlanmayı tercih etmiştir. Bu faktörler, modellerin eğitim süresini ve sonuçların doğruluğunu etkileyebileceği için araştırmaların çevresel faktörler açısından da dikkatle incelenmesi önemlidir.

CSE-CIC-IDS2018 veri seti üzerine yapılan araştırmalar, siber güvenlik saldırılarının tespitinde kullanılan farklı algoritmaların etkinliğini ölçme açısından önemli bilgiler sunmaktadır. Gelecekteki çalışmalar, bu araştırmaların bulgularını kullanarak daha verimli ve doğru sistemler geliştirmeyi hedefleyecektir. Bu bağlamda, kullanılan algoritmalar, metrikler

ve hesaplama ortamlarının derinlemesine karşılaştırılması, siber güvenlikte yeni nesil tehditlere karşı daha sağlam çözümler üretme yolunda önemli bir adım olacaktır.

Mevcut çalışmalardan farklı olarak, önerilen model ZOA kullanılarak yalnızca belirli saldırı türlerine odaklanmak yerine, veri seti içerisinde yer alan tüm saldırı türlerini kapsayacak biçimde tasarlanmıştır. ZOA, özellik seçimi aşamasında kullanılarak yüksek boyutlu ağ trafiği verilerinden en ayırt edici özelliklerin belirlenmesini sağlamış ve böylece sınıflandırma sürecinin etkinliği artırılmıştır. Seçilen özellik alt kümeleri; KNN, RF, DT, MLP gibi geleneksel makine öğrenmesi algoritmalarının yanı sıra LSTM, CNN ve RNN gibi derin öğrenme modelleri kullanılarak değerlendirilmiştir. Bununla birlikte, geliştirilen yaklaşım yalnızca statik veri kümeleri üzerinde test edilen bir sınıflandırma modeli olmanın ötesine geçerek, gerçek zamanlı saldırı tespit senaryolarına uygulanabilir bir yapı sunmaktadır. Bu kapsamda model, dinamik ağ trafiği verileri üzerinde çalışabilecek şekilde tasarlanmış ve değişen tehdit ortamına uyum sağlayabilecek esnek bir mimari ortaya konulmuştur. Elde edilen sonuçlar, önerilen yöntemin yalnızca yüksek sınıflandırma başarımı sunduğunu değil, aynı zamanda gerçek zamanlı olarak uygulanabilir ve sürekli güncellenebilir bir siber saldırı tespit çözümü olduğunu göstermektedir.

2.1 CSE-CIC-IDS2018 veri seti ile yapılmış çalışmalar

Bu bölümde, CSE-CIC-IDS2018 veri setini kullanan araştırma çalışmaları incelenmiştir. İlgili çalışmalar, yazarların soyadlarına göre alfabetik sırayla sunulmuştur. Doğruluk, geri çağırma ve benzeri performans metriklerine ait değerler, her bir çalışma kapsamında ikili sınıflandırma senaryosunda raporlanan en iyi sonuçları temsil etmektedir. Çizelge 2.1'de, bu bölümde ele alınan çalışmaların yazar soyadına göre alfabetik listesi ile birlikte, CSE-CIC-IDS2018 veri seti üzerinde elde edilen en yüksek performans değerleri, algoritmalar ve bilgisayar ortamları sunulmaktadır. Bununla birlikte, farklı araştırma çalışmaları arasında veya aynı çalışma kapsamında gerçekleştirilen farklı deneyler arasında doğrudan performans karşılaştırmalarının yapılması her zaman uygun olmayabilir. Bunun temel nedenleri arasında veri kümelerinin örnek sayısı, özellik boyutları ve kullanılan hesaplama çerçevelerindeki farklılıklar yer almaktadır. Ayrıca, aynı veri kümesi üzerinde gerçekleştirilen deneylerin farklı varyasyonları da bu sonuçlar üzerinde etkili olabilmektedir.

Çizelge 2.1 CSE-CIC-IDS2018 veri setinin performans puanları

Yazarlar	Accuary	Precision	Recall	AUC	Algoritmalar	Bilgisayar Ortamı

(Amin vd., 2024)	99,80	99,85	99,82	0,99	HXG-BLSTM	Python
(Appiahene vd., 2026a)	96,94	n/a	n/a	n/a	GConvTrans	Google Colab
(Atefinia & Ahmadi, 2021)	100,00	100,00	100,00	n/a	Modular Deep Neural Network	Python, Scikit-learn
(Abu Bakar vd., 2025)	95,00	n/a	n/a	n/a	IDS-NGNN	Python
(Basnet vd., 2019)	99,00	n/a	n/a	n/a	MLP	fast.ai, GPU, Python
(Bacevicius vd., 2026a)	99,85	98,34	98,66	n/a	CoLIME	Python
(Chi, 2025)	98,20	98,50	97,90	n/a	ST-CCNet	NVDIA Tesla
(Catillo vd., 2020)	99,20	95,00	98,90	n/a	Deep Autoencoder	Keras, Python, Tensorflow
(Chadza vd., 2019)	97,00	n/a	n/a	n/a	Baum Welch, Viterbi	MATLAB 2019a, Snort IDS
(Gupta vd., 2019)	n/a	n/a	n/a	n/a	LSTM	n/a
(D'hooge vd., 2019)	96,00	99,00	79,00	n/a	XGBoost	Python, Scikit-learn, XGBoost
(Ferrag vd., 2020)	97,38	n/a	98,18	n/a	RNN, Deep Autoencoder	Google Colab, GPU, Python,Tensorflow
(Lima Filho vd., 2019)	n/a	100,00	100,00	n/a	Random Forest	Python, Scikit-learn
(Fitni & Ramli, 2020)	98,80	98,80	97,10	0,94	Logistic Regression + Decision Tree + Gradient Boosting	GPU (on some PCs)
(Gamage & Samarabandu, 2020)	98,40	97,79	98,27	n/a	Deep Feed-forward Neural Network	Python, Scikit-learn
(Göcs & Johanyák, 2023)	99,99	99,95	99,90	n/a	DT/RF/SVM	Orange Visual Programing,Python
(Habibipour vd., 2026a)	99,30	n/a	n/a	n/a	ATS-IDS	Python
(Hua, 2020)	98,37	98,14	98,37	n/a	LightGBM	Scikit-learn, Tensorflow
(Ramos vd., 2020)	99,99	100,00	99,99	n/a	Random Forest, Decision Tree	Python, Scikit-learn
(Jagathbala vd., 2025a)	97,80	98,10	97,50	n/a	ForenXplorer	Docker,Python
(Panse & Bandhu, 2025)	98,21	97,88	97,94	n/a	Random Forest	Python
(Zhang vd., 2024)	84,7	96,75	96,94	n/a	GConvTrans	Python
(Ponkumar vd., 2025)	94,62	n/a	n/a	n/a	Random Forest + PCA + SMOTE	Python,Sckit-learn
(Simhon vd., 2026)	94,35	82,49	77,98	n/a	D-MAGIC	Python
(Kanimozhi & Jacob, 2019)	100,00	100,00	100,00	1	MLP	Python, Scikit-learn
(Kanimozhi & Jacob, 2019)	99,97	99,96	100,00	1	MLP	Python, Scikit-learn
(Karatas vd., 2020)	99,69	99,70	99,69	n/a	Adaboost	GPU, Keras, Python, Scikit-learn, Tensorflow
(Kim vd., 2020)	99,99	81,75	82,25	n/a	CNN	Python, Tensorflow
(Swain vd., 2025)	97,50	n/a	n/a	n/a	Random Forest	Python,Scikit-learn
(Li vd., 2020)	n/a	n/a	100,00	1	Deep Autoencoder	Python
(Lin vd., 2019)	96,20	96,00	96,00	n/a	LSTM	Tensorflow
(F. Zhao vd., 2020)	97,90	98,00	98,00	n/a	Deep Autoencoder	Python, Tensorflow

2.1.1. Modüler Derin Sinir Ağları ile Ağ Saldırı Tespiti: Atefinia ve Ahmadi'inin Yöntemi

Bu çalışmada (Atefinia & Ahmadi, 2021) dört farklı ağ mimarisini birleştirmek için bir toplayıcı modül (Happel vd., 1994) kullanılmıştır. Yazarlar, makalelerinde açıklanan ilgili çalışmalarda elde edilenlerden daha yüksek bir doğruluk oranı elde etmeyi hedeflemiştir. Bu dört ağ bileşeni, bir sınırlı Boltzmann makinesi (Lu vd., 2017), bir derin ileri beslemeli sinir ağı (Varsamopoulos vd., 2017) ve RNN, özellikle bir LSTM (Madan vd., 2018) ve Gated Recurrent Unit (GRU) (Lee vd., 2019) içerir.

Modeller Python ve Scikit-learn kütüphanesi kullanılarak uygulanmıştır. Veri ön işleme sırasında kaynak ve hedef IP adresleri ile kaynak port numaraları kaldırılmıştır. String değerleri içeren etiketler one-hot encoding yöntemiyle kodlanmış ve özellik uzayındaki tüm öznitelikler, 0 ile 1 arasında normalize edilmiştir. Eksik değerlere sahip satırlar ve çok fazla eksik değeri olan sütunlar CSE-CIC-IDS2018 veri setinden çıkarılmıştır, ancak kaç satır ve sütunun kaldırıldığına dair bir bilgi verilmemiştir. Her bir modül için %80 eğitim - %20 test oranında tabakalı örnekleme yapılmıştır. Daha sonra bilgiler, ağırlıklı ortalama tekniğini kullanan toplayıcı modüle aktarılmış ve modüler ağın çıktısı üretilmiştir.

DoS, DDoS ve kaba kuvvet saldırı türleri için en yüksek doğruluk oranları (%100) elde edilmiştir. Bu doğruluk oranları, %100 kesinlik ve geri çağırma skorları ile ilişkilendirilmiştir. Ancak, çalışmanın bir dezavantajı, yazarların sonuçlarını, ilgili çalışmalardaki sonuçlarla karşılaştırmasıdır. Daha iyi bir yaklaşım, en az iki modeli, biri önerilen modüler ağ olmak üzere, ampirik olarak değerlendirmek olacaktır. Bir diğer eksiklik, toplu saldırı türlerini kapsayan performans skorlarının bulunmamasıdır. Başka bir deyişle, saldırıların kombinasyonu için kesinlik, geri çağırma vb. skorlar ek bilgi sağlayabilir. Ancak bu, her bir saldırı türü için kesinlik, geri çağırma vb. raporlanmasının önemini azaltmaz.

2.1.2 Ağ Saldırı Trafiğinin Derin Öğrenme Çerçeveleri ile Tespiti ve Sınıflandırması: Basnet ve Arkadaşlarının Yaklaşımı

Bu çalışmada (Basnet vd., 2019), ağ saldırı trafiğini tespit etmek ve saldırı türlerini sınıflandırmak amacıyla çeşitli derin öğrenme çerçeveleri (fast.ai, Keras, PyTorch, TensorFlow, Theano) üzerinde deneyler gerçekleştirmiştir. Veri ön işleme aşamasında, "Infinity", "NaN" veya eksik değerlere sahip örnekler temizlenmiş ve zaman damgaları Unix

epoch sayısal değerlerine (1 Ocak 1970'ten itibaren geçen saniye sayısı) dönüştürülmüştür. Veri temizleme süreci sonucunda yaklaşık 20.000 örnek veri setinden çıkarılmıştır. Hedef port ve protokol özellikleri kategorik veri olarak ele alınırken, geri kalan özellikler sayısal veri olarak işlenmiştir.

Model eğitimi ve testinde %80–20 ve %70–30 oranıyla on katlı çapraz doğrulama yöntemi kullanılmıştır. İkili sınıflandırma ve çoklu sınıf sınıflandırma (Chaudhary vd., 2016) olmak üzere her iki senaryo da değerlendirilmiştir. Sınıflandırıcı olarak yalnızca MLP (Rynkiewicz, 2019) kullanılmıştır. GPU hızlandırmasının yardımıyla, fast.ai çerçevesinin tüm deneylerde diğer çerçevelerden tutarlı bir şekilde daha iyi performans gösterdiği ve ikili sınıflandırmada %99 gibi optimum bir doğruluk oranına ulaştığı gözlemlenmiştir.

Bu çalışmanın en önemli sınırlılığı, yalnızca bir sınıflandırıcının kullanılmış olmasıdır. Çeşitli sınıflandırıcıların karşılaştırmalı bir analizine yer verilmesi, elde edilen bulguların genel geçerliliğini arttırabilir.

2.1.3. 2L-ZED-IDS: Çoklu Saldırı Türleri için İki Seviyeli Anomali Tespit Sistemi

Bu çalışmada (Catillo vd., 2020), önceki araştırmalarına dayalı olarak CICIDS2017 veri setini genişletmiş ve CSE-CIC-IDS2018 veri seti üzerinde çalışmıştır. Bu çalışmada, bir derin otoenkoder (Chen vd., 2019) CICIDS2017 ve CSE-CIC-IDS2018 veri setleri üzerinde eğitilmiştir. Ön işleme aşamasında, veri setlerindeki Flow_ID ve Timestamp özellikleri, çalışmaya uygun görülmediği için seçim dışı bırakılmıştır. Otoenkoder Python, Keras ve TensorFlow kullanılarak uygulanmış ve normal trafik ile DoS saldırı trafiği üzerinde eğitilmiştir. Eğitim ve test verileri oranı her iki veri seti için de %80 - 20 olarak belirlenmiştir. CSE-CIC-IDS2018 veri setinde, botnet saldırı türü için en yüksek doğruluk oranı (%99,2) elde edilmiştir. Bu doğruluk oranı, %95,0 kesinlik ve %98,9 geri çağırma değerleri ile ilişkilendirilmiştir. Çalışmanın tamamında en yüksek doğruluk oranı (%99,3) ise CICIDS2017 veri setinde, yine botnet saldırı türü için elde edilmiştir ve bu değer %94,8 kesinlik ve %98,6 geri çağırma ile desteklenmiştir. Çalışmanın önemli bir sınırlılığı, toplu saldırı türleri için bir doğruluk skorunun sağlanmamış olmasıdır. Bir diğer dezavantaj ise yalnızca bir sınıflandırıcının kullanılmış olmasıdır. Bu durum, sonuçların genelleştirilebilirliğini kısıtlamaktadır.

2.1.4. Çağdaş Sıralı Ağ Saldırıları Tahmini: Gizli Markov Modeli Kullanılarak Bir İnceleme

Bu çalışmada (Chadza vd., 2019), CSE-CIC-IDS2018 veri seti üzerinde iki geleneksel Gizli Markov Modeli (HMM) eğitim algoritması olan Baum Welch (Joshi vd., 2017) ve Viterbi (Lember vd., 2021) yöntemlerini MATLAB yazılımı aracılığıyla uygulamıştır. HMM, durumlar ve gözlemler üreten olasılıksal bir makine öğrenimi çerçevesidir. Bu çalışmada veri ön işleme ile ilgili bilgiler net bir şekilde sunulmamıştır. Ancak, Snort Saldırı Tespit Sistemi (Shah vd., 2017) tarafından belirlenen seçim kriterlerine dayanarak CSE-CIC-IDS2018 veri setinden yaklaşık 457.550 kayıt seçilmiştir. Bu kayıtların %70'i eğitim, geri kalanı ise test için ayrılmıştır. Yazarlar, Baum Welch ve Viterbi algoritmalarının her ikisiyle de yaklaşık %97 doğruluk oranına ulaşıldığını bulmuştur. Ancak, makalenin yalnızca üç sayfa uzunluğunda olması, gerçekleştirilen deneylerin detaylarına dair bilgi eksikliğini birincil dezavantaj olarak öne çıkarmaktadır. Bu durum, çalışmanın tekrarlanabilirliğini ve bulguların genelleştirilebilirliğini sınırlamaktadır (Chadza vd., 2019).

2.1.5. Tekrarlayan Sinir Ağlarına Dayalı Bilgisayar Trafiği Analizi Yöntemi

Bu çalışmada (Gupta vd., 2019), CSE-CIC-IDS2018 veri setini kullanarak bilgisayar ağ trafiğini analiz etmek için bir LSTM modelini öneren oldukça teorik bir çalışma sunmuşlardır. Journal of Physics Conference serisine sunulan bu çalışma, ampirik sonuçlara yer vermemekte ve detaylar açısından oldukça eksiktir. Yazarlar, sınıf dengesizliğini ele almak amacıyla Facebook AI araştırmaları tarafından geliştirilen Focal Loss fonksiyonunu (Pasupa vd., 2023) kullandıklarını belirtmişlerdir. Ancak, herhangi bir metrik kullanımı veya hesaplama ortamı hakkında bilgi verilmemiştir. Bu makalenin temel dezavantajı, ampirik verilerin eksikliği ve kullanılan yöntemlerin uygulama detaylarının sunulmamasıdır. Bu durum, çalışmanın pratik uygulanabilirliğini ve tekrarlanabilirliğini sınırlamaktadır (Gupta vd., 2019).

2.1.6. Denetimli Makine Öğrenimi Yöntemlerinin Veri Setleri Arası Genelleme Gücü: D'hooge ve Arkadaşlarının Yaklaşımı

Bu çalışmada (D'hooge vd., 2019), CICIDS2017 ve CSE-CIC-IDS2018 veri setlerini içeren çalışmalarıyla, saldırı tespitine yönelik bir veri setiyle eğitilen bir modelin diğer bir veri

setine ne kadar verimli şekilde genellenebileceğini araştırmıştır. Performans değerlendirmesi için farklı ailelerden 12 denetimli öğrenme algoritması kullanılmıştır DT (W. Chen vd., 2018), RF (Ahmad vd., 2018), DT tabanlı Bagging (Taşer vd., 2019), Gradient Boosting (Ayyadevara, 2018), ExtraTree (R. Wang vd., 2019), Adaboost (Baig vd., 2017), XGBoost (T. Chen & Guestrin, 2016), KNN (Vajda vd., 2017) , Ncentroid (Saikia vd., 2020), linearSVC (Sulaiman vd., 2020), RBFSVC (Rahman vd., 2019) ve Lojistik Regresyon (Rymarczyk vd., 2019). Modeller, Python çerçevesinde Scikit-learn ve XGBoost modülleri kullanılarak oluşturulmuştur. Özellik uzayını normalize etmek için özellik ölçeklendirme uygulanmıştır. Ağaç tabanlı sınıflandırıcılar en iyi performansı göstermiş ve bunlar arasında XGBoost birinci sırada yer almıştır. CSE-CIC-IDS2018 için en yüksek doğruluk, kesinlik ve geri çağırma değerleri sırasıyla %96, %99 ve %79 olarak elde edilmiştir. Çalışmada, bir veri seti (CICIDS2017) üzerinde eğitilen bir modelin diğer bir veri setine (CSE-CIC-IDS2018) genelleme yapamayacağı sonucuna varılmıştır. Ancak çalışmanın bir eksikliği, belirli kategorik özelliklerin (örneğin hedef port) iki veri seti için de aynı sayıda benzersiz değere sahip olduğu varsayımıdır. Örneğin, bir sınıflandırıcı bir veri setindeki ’X’ özelliği {araba, tekne} ile eğitilmişse, aynı özelliğin {araba, tekne, tren, uçak} değerlerini içeren bir başka veri setinde iyi bir şekilde genelleme yapması beklenmez. Bu durum, kategorik özelliklerin tutarsızlığının genelleme yeteneğini sınırlayabileceğini göstermektedir (D’hooge vd., 2019).

2.1.7. Siber Güvenlikte Derin Öğrenme ile Saldırı Tespiti: Yöntemler, Veri Setleri ve Karşılaştırmalı Bir Çalışma

Bu çalışmada (Ferrag vd., 2020), CSE-CIC-IDS2018 ve Bot-IoT veri setlerinde yedi farklı derin öğrenme modelini değerlendirmiştir. İncelenen modeller şunlardır: RNN, derin sinir ağları (Lin vd., 2019), sınırlı Boltzmann makineleri, derin inanç ağları (J. Li vd., 2018), CNN (Y. Zhao vd., 2019), derin Boltzmann makineleri (Taherkhani vd., 2018) ve derin otoenkoderler. Bot-IoT veri seti, New South Wales Üniversitesi (UNSW) tarafından 2018 yılında oluşturulmuş ve yaklaşık 72 milyon normal ve botnet IoT örneği ile 46 özellik içermektedir. Deneyler, Python ve TensorFlow kullanılarak GPU hızlandırmalı Google Colaboratory ortamında gerçekleştirilmiştir. Veri setinin yalnızca %5'i bu çalışmada kullanılmıştır (Koroniotis vd., 2019). Bot-IoT veri seti için en yüksek doğruluk oranı (%98,39) derin bir otoenkoder ile elde edilirken, CSE-CIC-IDS2018 için en yüksek doğruluk oranı (%97,38) bir RNN ile elde edilmiştir. Bot-IoT veri seti için en yüksek geri çağırma oranı (%97,01) bir CNN ile elde edilmişken, CSE-CIC-IDS2018 için en yüksek geri çağırma oranı

(%98,18) derin bir otoenkoder ile sağlanmıştır. Makalenin büyük bir bölümü, 35 siber güvenlik veri setinin sınıflandırılması ve yedi derin öğrenme modelinin açıklanmasına ayrılmıştır. Ancak, deneylere ilişkin ayrıntılar yalnızca son üç sayfada yer almakta ve detaylı bilgi eksikliği çalışmanın en önemli dezavantajı olarak öne çıkmaktadır (Ferrag vd., 2020).

2.1.8. Akıllı Tespit: DoS/DDoS Saldırılarının Makine Öğrenimi ile Çevrimiçi Tespiti

Bu çalışmada (Lima Filho vd., 2019), özel bir veri seti ile dört yaygın veri setini (CIC-DoS (Jazi vd., 2017), ISCX2012, CICIDS2017 ve CSE-CIC-IDS2018) kullanarak çevrimiçi rastgele ağ trafiği örneklerini toplamak ve bu trafiği DoS saldırıları veya normal trafik olarak sınıflandırmak üzere bir çalışma gerçekleştirmiştir. Her veri setinden elde edilen 33 özellik, kaynak ve hedef portlar, taşıma katmanı protokolü, IP paket boyutu ve TCP bayraklarından türetilmiştir. Bireysel veri setleri, aşağıdaki örnek türlerini içeren tek bir birim halinde birleştirilmiştir: normal trafik (23.088 örnek), TCP flood saldırıları (14.988 örnek), UDP flood (6.894 örnek), HTTP flood (347 örnek) ve HTTP slow (183 örnek). Birleştirilmiş veri setinde, ISCX2012 yalnızca normal trafik davranışını sağlamak amacıyla kullanılmıştır. Altı farklı algoritma (RF, DT, Lojistik Regresyon, SGDClassifier, Adaboost, MLP) üzerinde çapraz doğrulama ile Özyinelemeli Özellik Elemesi (Recursive Feature Elimination) yapılmıştır. Modeller, Scikit-learn kullanılarak oluşturulmuştur. Birleştirilmiş veri setinde, 20 özellik seçilerek oluşturulan RF algoritması en yüksek doğruluk oranına ulaşmıştır. CSE-CIC-IDS2018 veri seti için RF modeli, hem kesinlik hem de geri çağırma değerlerinde %100 başarı göstermiştir. Çalışmanın temel eksikliklerinden biri, birleştirilmiş veri setinde normal trafik davranışını sağlamak için ISCX2012'nin kullanılmasıdır. ISCX2012 güncel değildir ve yalnızca altı trafik protokolü ile sınırlı olması, çalışmanın genelleştirilebilirliğini kısıtlamaktadır (Lima Filho vd., 2019).

2.1.9. Anomali Tabanlı Saldırı Tespiti için Ensemble Model Yaklaşımı

Bu çalışmada (Fitni & Ramli, 2020) anomali tabanlı saldırı tespit sistemlerinde performansı artırmak amacıyla ensemble model yaklaşımını benimsemiş ve yedi farklı öğreniciyi (RF, Gaussian Naive Bayes, DT, Quadratic Discriminant Analysis, Gradient Boosting, Lojistik Regresyon) karşılaştırmıştır. Python ve Scikit-learn kullanılarak oluşturulan modeller, ön işleme aşamasında eksik veya sonsuz değerler ile tekrar eden başlık satırlarının temizlenmesinden sonra %80 test %20 eğitim setlerine bölünmüştür. Spearman'ın sıralama

korelasyon katsayısı ve Ki-kare testi ile yapılan özellik seçimi sonucunda 23 özellik belirlenmiştir. Gradient Boosting, Lojistik Regresyon ve DT ensemble modelde kullanılmak üzere en iyi performansı gösteren algoritmalar olarak seçilmiştir. Ensemble model, %98,80 doğruluk, %98,80 kesinlik, %97,10 geri çağırma ve 0.94 ROC, AUC değerleriyle güçlü bir performans sergilemiştir. Yazarlar, Gradient Boosting yerine CatBoost, LightGBM veya XGBoost gibi gelişmiş sınıflandırıcıların kullanılması durumunda modelin performansının daha da artırılabileceğini önermektedir (Fitni & Ramli, 2020).

2.1.10. Ağ Saldırı Tespitinde Derin Öğrenme Yöntemleri: Bir Anket ve Nesnel Karşılaştırma

Bu çalışmada **(Gamage & Samarabandu, 2020)**, ağ saldırı tespiti için kullanılan derin öğrenme modellerinin bir sınıflandırmasını tanıtmış ve ilgili araştırma makalelerini özetlemiştir. Bu çalışmada, dört farklı derin öğrenme modeli (ileri beslemeli sinir ağı, otoenkoder, derin inanç ağı, LSTM, KDD Cup 1999, NSL-KDD (Tavallaee vd., 2009), CICIDS2017 ve CSE-CIC-IDS2018 veri setleri üzerinde değerlendirilmiştir. KDD Cup 1999 veri seti, Defense Advanced Research Project Agency (DARPA) tarafından geliştirilmiş olup, DoS kategorisi de dahil olmak üzere dört saldırı türü içermektedir. Veri setlerinin ön işleme aşamasında geçersiz akış kayıtları (eksik değerler, string ifadeler vb.) kaldırılmış ve özellik ölçeklendirme yapılmıştır. Protokol türü, hizmet ve bayrak özellikleri gibi kategorik ve sayısal olmayan üç öznitelik için one-hot kodlama uygulanmıştır. Tam veri setleri eğitim ve test gruplarına bölünmüş, hiperparametre optimizasyonu gerçekleştirilmiştir. Sonuçlar, ileri beslemeli sinir ağlarının tüm veri setlerinde iyi performans gösterdiğini ortaya koymuştur. Bu sınıflandırıcı ile en yüksek doğruluk (%99,58), CICIDS2017 veri setinde elde edilmiş olup, bu değer %99,43 kesinlik ve %99,45 geri çağırma oranları ile ilişkilendirilmiştir. CSE-CIC-IDS2018 için ileri beslemeli ağ, en yüksek doğruluk oranı olarak %98,4'e ulaşmış ve bu sonuç %97,79 kesinlik ve %98,27 geri çağırma ile desteklenmiştir. Deneylerde kullanılan bazı bilgisayarlarda GPU hızlandırma kullanılmıştır. Bu çalışmanın temel eksikliklerinden biri, KDD Cup 1999 ve NSL-KDD gibi güncelliğini yitirmiş ve bilinen sorunlara sahip veri setlerinin kullanılmasıdır. KDD Cup 1999'un en büyük sorunu, önemli ölçüde tekrar eden kayıtlar içermesidir (Tavallaee vd., 2009). NSL-KDD, tekrar eden örnekler içermemesi nedeniyle iyileştirilmiş bir versiyon olsa da ideal olmaktan uzaktır. Örneğin, NSL-KDD test veri setinde bazı saldırı türlerine ait kayıtlar bulunmamaktadır **(Groff & Schwartz., 2019).**

2.1.11. Gömülü Özellik Seçimi ve LightGBM Kullanarak Verimli Trafik Sınıflandırma Yöntemi

(Hua, 2020), CSE-CIC-IDS2018 veri setindeki sınıf dengesizliğini çözmek için undersampling ve gömülü özellik seçimi yöntemini bir LightGBM sınıflandırıcı ile birleştirmiştir. LightGBM, veri setlerinde yüksek sayıda örnek ve özelliği ele almak için algoritmalar içermektedir. Undersampling yöntemi, dağılımı etkilemek amacıyla çoğunluk sınıfındaki örnekleri rastgele kaldırır. Bu çalışmada, veri temizleme aşamasında eksik değerler ve gereksiz özellikler çıkarılarak 77 özelliğe sahip bir veri seti oluşturulmuştur. String etiketler, önce tamsayı etiketlere dönüştürülmüş, ardından one-hot kodlama uygulanmıştır. Bu araştırmada altı farklı algoritma değerlendirilmiştir: SVM (Ahmad vd., 2018), RF, Adaboost, MLP, CNN ve Naive Bayes (Abdullah-Al-Kafi vd., 2022). Scikit-learn ve TensorFlow kullanılarak uygulanmıştır. Eğitim ve test veri setleri %70-30 oranında bölünmüş ve özellik seçimi XGBoost algoritması ile gerçekleştirilmiştir. LightGBM sınıflandırıcı, üç milyon örnekten oluşan veri setinde ve ilk on özellik seçildiğinde %98,37 doğruluk ile en iyi performansı göstermiştir. Bu doğruluk için kesinlik %98,14, geri çağırma ise %98,37 olarak hesaplanmıştır. Ayrıca, LightGBM sınıflandırıcı, öğreniciler arasında ikinci en hızlı eğitim süresine sahip olmuştur. Bu çalışma, diğer araştırmalardan daha fazla veri ön işleme bilgisi sağlamasına rağmen, veri temizleme işlemini yüzeysel bir şekilde ele almıştır.

2.1.12. Trafik Akışı Analitiği ile Botnet Tespit Araçlarının Değerlendirilmesi için Kıyaslama Tabanlı Referans Modeli

Bu çalışmada (Ramos vd., 2020) iki veri seti (CSE-CIC-IDS2018 ve ISOT HTTP Botnet) üzerinde beş farklı algoritmayı değerlendirerek en iyi botnet sınıflandırıcısını belirlemeyi hedeflemiştir. ISOT HTTP Botnet veri seti, Zararlı ve güvenli DNS trafiği örneklerini içermektedir. Çalışmada kullanılan algoritmalar şunlardır: RF, DT, KNN, Naive Bayes ve SVM. Özellik seçimi, RF'nin önem yöntemi **(Gupta & Bhavsar, 2017)** dahil olmak üzere çeşitli tekniklerle gerçekleştirilmiştir. Özellik seçimi sonrasında CSE-CIC-IDS2018 veri setinde 19, ISOT HTTP Botnet veri setinde ise 20 bağımsız özellik belirlenmiştir. Seçilen özellikler arasında hedef port numarası, kaynak port numarası ve taşıma protokolü yer almaktadır. Modeller, Python ve Scikit-learn kullanılarak uygulanmış ve beş katlı çapraz doğrulama ile değerlendirilmiştir. Eğitim seti, botnet örneklerinin %80'ini içerirken geri kalan %20'lik kısım test seti olarak ayrılmıştır. Model optimizasyonu için Grid Search algoritması

(Yuanyuan & Yongming, 2017) kullanılmıştır. CSE-CIC-IDS2018 veri seti için RF ve DT algoritmaları, %99,99 doğruluk oranı ile en iyi performansı göstermiştir. Bu doğruluk oranı, her iki algoritma için %100 kesinlik ve %99,99 geri çağırma değerleriyle ilişkilendirilmiştir. ISOT HTTP Botnet veri setinde de RF (%99,94) ve DT (%99,90) en yüksek doğruluk oranlarına ulaşmıştır. Çalışmanın temel eksikliği, veri ön işleme ile ilgili sağlanan bilgilerin yetersiz olmasıdır (Ramos vd., 2020).

2.1.13. CSE-CIC-IDS2018 Veri Seti Kullanılarak Bulut Bilişim ile Yapay Zekâ Tabanlı Ağ Saldırı Tespiti ve Hiperparametre Optimizasyonu

Bu çalışmada (Kanimozhi & Jacob, 2019), botnet saldırılarını tespit etmek amacıyla Python ve Scikit-learn kullanarak iki katmanlı bir MLP modeli eğitmişlerdir. Modelin hiperparametre optimizasyonu GridSearchCV yöntemiyle (Ranjan vd., 2019) gerçekleştirilmiş ve aşırı öğrenmeyi önlemek için L2 düzenlileştirme (Bilgic vd., 2014) kullanılmıştır. Aşırı öğrenme, bir modelin eğitim verilerini ezberlemesi ve genelleştirme yeteneğinin azalması olarak tanımlanır (Meyer vd., 2018). MLP sınıflandırıcı, CSE-CIC-IDS2018 veri setindeki yalnızca botnet örnekleri üzerinde eğitilmiş ve on katlı çapraz doğrulama yöntemi (Yadav & Shukla, 2016) uygulanmıştır. Bu çalışmada, ROC AUC değeri 1 olarak hesaplanmış ve bu mükemmel bir skor olarak kabul edilmiştir. İlgili doğruluk, kesinlik ve geri çağırma değerleri de %100 olarak elde edilmiştir. Ancak, bu makale yalnızca iki referansa yer vermiş ve detay açısından belirgin bir eksiklik göstermiştir. Çalışmanın bir diğer dezavantajı, performansı değerlendirmek için yalnızca bir sınıflandırıcının kullanılmış olmasıdır.

2.1.14. CSE-CIC-IDS2018 Veri Seti Kullanılarak Ağ Saldırı Tespit Sisteminde Farklı Optimizasyonlu Makine Öğrenimi Sınıflandırıcılarının Kalibrasyonu

Bu çalışmada (Kanimozhi & Jacob, 2019), ağ saldırı tespit sisteminde altı farklı sınıflandırıcıyı (MLP, RF, KNN, SVM, Adaboost, Naive Bayes) değerlendirerek en iyi performansı göstereni belirlemeyi hedeflemişlerdir. Modeller Python ve Scikit-learn kullanılarak geliştirilmiştir. Çalışmada, sınıflandırıcıların mükemmel bir şekilde kalibre edilmiş bir eğriden ne kadar sapma gösterdiğini görselleştiren bir kalibrasyon eğrisi kullanılmıştır. CSE-CIC-IDS2018 veri setindeki botnet örnekleri eğitim ve test gruplarına ayrılmış; ancak, bu ayrımın oranı hakkında bilgi verilmemiştir.MLP modeli, ROC AUC değeri 1 ile en iyi sınıflandırıcı olarak belirlenmiştir. Bu mükemmel AUC değerine karşılık gelen doğruluk,

kesinlik ve geri çağırma değerleri sırasıyla %99,97, %99,96 ve %100 olarak hesaplanmıştır. Ancak, MLP sınıflandırıcıya dair teknik detaylar verilmemiştir; bunun önceki çalışmada kullanılan iki katmanlı ağ olduğu tahmin edilmektedir (Karatas vd., 2020). Çalışmanın temel eksiklikleri arasında kullanılan yöntemlerin ve veri işleme adımlarının yetersiz detaylandırılması bulunmaktadır.

2.1.15. Dengesiz ve Güncel Veri Setinde Makine Öğrenimi Tabanlı IDS'lerin Performansını Artırma

Bu çalışmada (Karatas vd., 2020), sınıf dengesizliğini çözmek için SMOTE algoritmasını (Fernández vd., 2018) kullanarak CSE-CIC-IDS2018 veri setinde altı farklı öğrenicinin performansını değerlendirmiştir. Bu öğreniciler şunlardır: KNN, RF, Gradient Boosting, Adaboost, DT ve Doğrusal Ayrım Analizi (Negi vd., 2016). Modeller Python ortamında Keras, TensorFlow ve Scikit-learn kullanılarak geliştirilmiştir. CSE-CIC-IDS2018 veri setinin yaklaşık 5 milyon örnek içerdiği belirtilmiştir; ancak, veri setinin aslında yaklaşık 16 milyon örnek içerdiği göz önüne alındığında, çalışmada bir alt küme kullanıldığının açıkça belirtilmesi gerektiği ifade edilmiştir. Veri seti, eksik değerler ve "Infinity" gibi sorunları gidermek için ön işleme tabi tutulmuş, one-hot kodlama yapılmış ve rastgelelik sağlamak için satırlar karıştırılmıştır. Veri setinin %80'i eğitim, %20'si test grubu olarak ayrılmış ve beş katlı çapraz doğrulama uygulanmıştır. SMOTE uygulandıktan sonra toplam veri seti boyutu %17 oranında artmıştır. Bu çalışmada Adaboost öğrenicisi, %99,69 doğruluk, %99,70 kesinlik ve %99,69 geri çağırma değerleri ile en iyi performansı göstermiştir. Yazarlar veri temizleme sürecine diğer çalışmalara kıyasla daha fazla yer ayırmış olsa da, bu süreçle ilgili daha fazla ayrıntıya girilmesi gerektiği belirtilmiştir. Yine de, incelenen çalışmalar arasında veri temizleme konusunda en kapsamlı yaklaşımlardan birini sunduğu ifade edilmiştir.

2.1.16. DoS Saldırılarına Karşı CNN Tabanlı Ağ Saldırı Tespit Sistemi

Bu çalışmada (Kim vd., 2020), KDD Cup 1999 ve CSE-CIC-IDS2018 veri setlerindeki DoS saldırılarını tespit etmek amacıyla bir CNN eğitmişlerdir. Model, Python ve TensorFlow kullanılarak geliştirilmiştir. Her iki veri seti için eğitim ve test oranı %70–30 olarak belirlenmiştir. KDD veri setinde yaklaşık 283.000, CSE-CIC-IDS2018 veri setinde ise yaklaşık 11.000.000 örnek kullanılmıştır. Çalışmada, veri setlerinden görüntü veri kümeleri oluşturulmuş ve ikili ile çok sınıflı sınıflandırma gerçekleştirilmiştir. Yazarlar, her iki veri

setinde de ikili sınıflandırma için doğruluk oranının %99 olduğunu, buna karşılık kesinlik ve geri çağırma oranlarının sırasıyla %81,75 ve %82,25 olarak elde edildiğini ifade etmişlerdir. Çalışma kapsamında, karşılaştırma yapmak amacıyla bir RNN modeli de eklenmiştir. Çalışmanın temel eksikliği, KDD Cup 1999 veri setinin kullanımıdır. Bu veri seti, önceki çalışmalarda da belirtildiği üzere, güncelliğini yitirmiş ve çok sayıda yinelenen örnek içermektedir.

2.1.17. Rastgele Orman Özellik Seçimi Tabanlı Autoencoder Ağ Saldırı Tespit Sistemi

Bu çalışmada (Li vd., 2020), çevrimiçi gerçek zamanlı tespit amacıyla gözetimsiz kümeleme ve özellik seçimine dayalı bir model önermiştir. Ön işleme sırasında "Infinity" ve "NaN" değerleri 0 ile değiştirilmiş ve veriler seyrek (sparse) ve yoğun (dense) matrislere bölünerek L2 düzenlileştirme ile normalleştirilmiştir. Seyrek matrisler, çoğunlukla 0 değerine sahip elemanları, yoğun matrisler ise çoğunlukla sıfır olmayan elemanları barındırır. Model, Python ortamında geliştirilmiştir. En iyi özellikler RF yöntemiyle seçilmiş ve eğitim-veri seti oranı %85–15 olarak belirlenmiştir. Eğitim veri setinin %25'ine Affinity Propagation (AP) kümeleme algoritması (Mirsky vd., 2018) uygulanmış ve bu işlemle elde edilen özellik alt kümeleri autoencoder modeline aktarılmıştır. Önerilen modelin geri çağırma oranları, başka bir autoencoder modeli olan KitNet (Chorowski & Bahdanau, 2015) ile karşılaştırılmıştır. Her iki model için de birçok saldırı türünde geri çağırma oranı %100 olarak elde edilmiştir. Ancak, yalnızca önerilen model AUC metriğiyle değerlendirilmiş ve bazı saldırı türlerinde AUC değeri 1 olarak hesaplanmıştır. Tespit süresi sonuçlarına göre, önerilen modelin KitNet'ten daha hızlı olduğu gösterilmiştir. CSE-CIC-IDS2018 veri setine ait saldırı türleri için AUC ve geri çağırma performans skorları sağlanmıştır. Ancak, toplu saldırı türlerini kapsayan skorların olmaması, çalışmanın eksikliği olarak değerlendirilmiştir. Ayrıca, KitNet için AUC değerlerinin eksik olması da bir diğer dezavantajdır.

2.1.18. Derin Öğrenme Teknikleri Kullanılarak Dinamik Ağ Anomali Tespit Sistemi

Bu çalışmada (Lin vd., 2019), LSTM modellerinin performansını artırmak için Dikkat Mekanizması (AM) kullanarak dinamik ağ anomali tespit sistemlerini incelemiştir. Dikkat Mekanizması, insan beyninin odaklanma işlevini taklit ederek, hedefle en alakalı bilgileri otomatik bir ağırlıklandırma düzeniyle çıkarır ve temsil eder. Model, TensorFlow ile geliştirilmiş ve Adam Gradient Descent algoritması (Zhang, 2018) ile optimize edilmiştir. Bu

algoritma, Stokastik Gradient Descent'in (Sharma, 2018) bir yerine geçme algoritmasıdır. Ayrıca, DT, Gaussian Naive Bayes, RF, KNN, SVM, MLP ve Dikkat Mekanizması olmadan bir LSTM modeli gibi yedi algoritma daha değerlendirilmiştir. CSE-CIC-IDS2018 veri setinin yaklaşık %50'sini içeren bir alt küme, zaman damgası ve IP adresi özelliklerinin çıkarılmasıyla ön işleme tabi tutulmuştur. Veri seti, %90 eğitim, %9 test ve %1 doğrulama setine bölünmüştür. Normal trafik verisi rastgele eksiltme yöntemiyle 2.000.000 kayda indirgenmiş, Web ve sızma saldırıları ise sınıf dengesizliğini ele almak için SMOTE ile çoğaltılmıştır. Dikkat Mekanizması ile donatılmış LSTM modeli, %96.2 doğruluk ve %96 kesinlik ile geri çağırma oranlarıyla diğer öğrenicilerden üstün performans sergilemiştir. Bu çalışmanın katkıları faydalı olmakla birlikte, veri temizleme ile ilgili eksik bilgi sağlanması ve SMOTE ile uygulanan çoğaltma oranının belirtilmemesi gibi eksiklikler bulunmaktadır.

2.1.19. Yarı-Kendi Kendine Öğrenen Ağ Saldırı Tespit Sistemi

Bu çalışmada (F. Zhao vd., 2020), bir gürültü giderici autoencoder modelini, bulanık c-means algoritmasına dayanan sınıf ayrımı yöntemiyle birleştirerek ağ saldırılarını tespit etmeyi hedeflemiştir. Bu yöntem, eksik değerler ve gereksiz veriler gibi sorunları gidermek için benimsenmiş, ancak sınıf gürültüsü ile başa çıkmada etkisiz kalmıştır. Sınıf gürültüsü, aynı örnek için farklı sınıf etiketleri verilmesi veya yanlış sınıflandırılmış örneklerden kaynaklanabilir (Zhu vd., 2020). Model Python ve TensorFlow kullanılarak geliştirilmiştir. Veri seti, %70 eğitim, %15 doğrulama ve %15 test setine ayrılmıştır. Önerilen model, %97,9 doğruluk oranı ve %98,0 kesinlik ile geri çağırma oranları elde etmiştir. Ancak, çalışma deneysel ayrıntılar konusunda yetersiz kalmıştır.

2.1.20. Gerçek Zamanlı Ağ Güvenliği: Yapay Sinir Ağları ve Dinamik Grafik Tabanlı Kümelemenin Entegrasyonu

Bu çalışmada (Simhon vd., 2026), siber saldırıların artan karmaşıklığı ve sıklığı, Ağ Saldırı Tespit Sistemlerini (NIDS) modern siber güvenliğin vazgeçilmez bileşenlerinden biri hâline getirmiştir (Tarek Abdelaziz vd., 2025). Bu çalışmada, etiketli veri kümelerine ve önceden tanımlanmış saldırı imzalarına ihtiyaç duymadan çalışan, sıfır-atış öğrenme ve grafik tabanlı dinamik kümeleme yaklaşımlarını kullanan, gerçek zamanlı ve denetimsiz bir NIDS modeli olan D-MAGIC sunulmaktadır. D-MAGIC, ağ akışları arasındaki ilişkileri grafik yapısına gömerek benzer davranış örüntülerini kümelemekte ve hem bilinen hem de yeni ortaya çıkan

tehditleri tespit edebilmektedir (Keserwani vd., 2023). Ayrıca, kümeleme yöntemlerinin sınıflandıramadığı izole anormallikleri değerlendirmek amacıyla yaklaşık en yakın komşu (ANN) analizine dayalı SAGA yöntemi önerilmekte ve bu sayede daha ince ve gelişmekte olan saldırıların tespiti güçlendirilmektedir. CIC-IDS-2017 ve CSE-CIC-IDS-2018 veri kümeleri üzerinde yapılan deneysel sonuçlar, D-MAGIC'in mevcut yöntemlere kıyasla F1 skorunda %12'ye varan iyileşme sağladığını, yanlış pozitif oranlarını azalttığını ve düşük gecikme ile gerçek zamanlı tespit gerçekleştirdiğini; SAGA'nın ise CIC-IDS-2017 veri kümesinde doğruluk oranını %6'ya kadar artırdığını göstermektedir (Taherdoost, 2024).

Çalışmasında önerilen D-MAGIC ve SAGA modelleri, yenilikçi yaklaşımlar sunmalarına rağmen bazı sınırlamalara sahiptir. Sistem, gizlilik ve şifreli trafiği destekleme amacıyla yalnızca istatistiksel akış düzeyi özelliklerini kullanmakta ve Derin Paket İncelemesi (DPI) gerçekleştirmemektedir; bu durum, SQL Injection gibi içerik odaklı saldırıların veya akış düzeyinde normal trafiğe oldukça benzeyen düşük hacimli saldırıların tespitini zorlaştırmaktadır (Taherdoost, 2024). Deneysel değerlendirmeler, kullanılan veri setlerinin yapısı gereği ağırlıklı olarak sıralı saldırı senaryoları üzerinde gerçekleştirilmiş olup, eş zamanlı ve çok vektörlü saldırıların analizi çalışmanın kapsamı dışında bırakılmıştır. SAGA modeli, izole ve sinsi saldırıları tespit etmede etkili olmasına rağmen, D-MAGIC'e kıyasla belirgin derecede daha yüksek bir yanlış pozitif oranına sahiptir; örneğin CSE-CIC-IDS-2018 veri setinde D-MAGIC %4,03 FPR elde ederken, SAGA'nın yanlış pozitif oranı %27,05 seviyelerine çıkabilmektedir(M. Wang vd., 2021). Ayrıca, yüksek veri hacimlerinin işlenmesi amacıyla kullanılan kuyruk tabanlı düğüm kaldırma mekanizması bellek ve zaman verimliliğini artırsa da, tespit doğruluğunda sınırlı bir düşüşe yol açmaktadır. Deneysel sonuçlar incelendiğinde, özellikle Bot ve SQL Injection gibi bazı saldırı türlerinde modellerin F1-skoru ve geri çağırma oranlarının diğer saldırı tiplerine kıyasla düşük kaldığı görülmektedir. Bununla birlikte, mevcut sistemde anomali eşiklerinin istatistiksel yöntemlerle belirlenmesi, gerçek zamanlı geri bildirimlere dayalı daha uyarlanabilir ve hassas dinamik eşikleme mekanizmalarına duyulan ihtiyacı ortaya koymaktadır. Son olarak, çalışma gerçek zamanlı tespit yeteneğini göstermesine rağmen, tespit hattının her aşamasındaki depolama gereksinimleri, bellek kullanımı ve hesaplama gecikmelerine ilişkin daha ayrıntılı analizlerin gelecekte ele alınması gereken bir araştırma konusu olduğu belirtilmiştir (Zhu vd., 2020).

2.1.21. Grafik Sinir Ağları ve Transformer Mimarilerine Dayalı Hibrit Bir Yaklaşımla Ağ Saldırı Tespiti

Bu çalışmada (Zhang vd., 2024), Bulut bilişimin; sunucu, veritabanı ve depolama gibi hizmetleri internet üzerinden esnek ve ölçeklenebilir biçimde sunması, bu teknolojinin hızla yaygınlaşmasına olanak tanımaktadır. Ancak bu yaygınlaşma, geleneksel güvenlik mekanizmaları olan güvenlik duvarları gibi yöntemleri aşabilen daha karmaşık ve sofistike sızma saldırılarına karşı sistemlerin maruz kaldığı tehdit yüzeyini de genişletmektedir (Zhang vd., 2024). Bu bağlamda, makine öğrenmesi ve derin öğrenme teknikleriyle güçlendirilmiş Ağ Sızma Tespit Sistemleri (Network Intrusion Detection Systems – NIDS) giderek daha kritik bir rol üstlenmektedir (Ahmad vd., 2018). Mevcut literatürde önemli ilerlemeler kaydedilmiş olmasına rağmen, birçok yapay zekâ tabanlı saldırı tespit modeli; yüksek kaliteli ve kapsamlı etiketli veri kümelerine olan bağımlılığı ile dağıtık bulut ortamlarında gözlemlenen karmaşık ve dinamik saldırı örüntülerini yeterince temsil edememesi nedeniyle sınırlılıklar taşımaktadır. Bu çalışma, söz konusu problemlere çözüm sunmak amacıyla Grafik Evrişimsel Ağ (GCN) katmanları ile Transformer tabanlı kodlayıcı katmanlarını birleştiren ve GConvTrans olarak adlandırılan hibrit bir derin öğrenme mimarisini önermektedir (Devnath, 2023). Önerilen yaklaşımda, CIC-IDS 2018 veri setinde yer alan tablo biçimindeki ağ trafiği verileri hesaplamalı grafiklere dönüştürülerek ağ düğümleri arasındaki yerel yapısal ilişkilerin GCN katmanları aracılığıyla, küresel bağlam ve uzun menzilli bağımlılıkların ise çoklu başlıklı öz-dikkat (multi-head self-attention) mekanizmalarına sahip Transformer kodlayıcılarıyla öğrenilmesi sağlanmıştır (Abdi vd., 2024). Deneysel sonuçlar, GConvTrans modelinin eğitim, doğrulama ve test aşamalarında sırasıyla %84,7, %96,75 ve %96,94 doğruluk oranlarına ulaşarak yüksek bir tespit başarımı sunduğunu göstermektedir. Elde edilen bulgular, grafik tabanlı öğrenme yaklaşımlarının standart derin öğrenme yöntemleriyle bütünleştirilmesinin, karmaşık ağ saldırılarının tespitinde sağlam ve etkili bir çözüm sunduğunu ortaya koymaktadır (Vaswani vd., 2017). Bununla birlikte, çalışmanın CIC-IDS 2018 veri setinin belirli bir alt kümesiyle sınırlı olması, modelin farklı veri setleri ve ağ senaryoları üzerindeki genellenebilirliğinin gelecekteki çalışmalarda daha kapsamlı biçimde araştırılmasını gerekli kılmakta; ayrıca önerilen mimarinin hiperparametre optimizasyonu ve bağlantı tahmini (link prediction) gibi diğer grafik öğrenme görevlerine uyarlanması, önemli birer gelecek araştırma yönü olarak değerlendirilmektedir.

2.1.22. ATS-IDS: Artımlı öğrenmeyi kullanan, IoT'de hafif saldırı tespit sistemleri için uyarlanabilir bir öğretmen-öğrenci çerçevesi

Bu çalışma (Habibipour vd., 2026), kaynak kısıtlı IoT cihazları için yüksek doğruluklu ancak düşük hesaplama ve bant genişliği maliyetine sahip bir saldırı tespit sistemi sunmaktadır.

Çalışmanın mimarisinde bulut tarafında çalışan karmaşık bir öğretmen modeli ile uç cihazlarda konumlandırılan hafif bir öğrenci modeli yer almakta ve bu iki yapı arasında Bilgi Damıtma (Knowledge Distillation) yoluyla bilgi aktarımı gerçekleştirilmektedir (Habibipour vd., 2026). Sistem, bant genişliğini optimize etmek amacıyla yalnızca belirsizliği yüksek verilerin buluta gönderilmesini sağlayan Adaptive-Send algoritmasıyla veri iletimini ortalama %90 oranında azaltırken, model güncellemelerinde tüm ağırlıklar yerine yalnızca anlamlı ağırlık değişimlerini ileten Optimize-Weights algoritması sayesinde güncelleme maliyetini %85'e kadar düşürmektedir (Mendonça vd., 2022). UNSW-NB15, ToN-IoT ve CSE-CIC-IDS 2018 veri setleri üzerinde yapılan deneylerde model parametre sayısında %94,36 ve model boyutunda %91,65 oranında önemli bir küçülme elde edilmiş, Artımlı Öğrenme (Incremental Learning) yaklaşımıyla daha önce görülmemiş sıfırıncı gün saldırılarının tespit oranı ortalama %10,67 artırılmıştır. Bununla birlikte, ATS-IDS hâlen etiketli veri bağımlılığına sahip olup gerçek dünya senaryolarında kapsamlı etiketli veri elde etmenin zorluğu önemli bir sınırlama olarak öne çıkmakta, uzun süreli bağlantı kopukluklarının öğretmen-öğrenci senkronizasyonunu geciktirebilmesi sistemin hızlı gelişen tehditlere adaptasyonunu sınırlayabilmektedir. Ayrıca, her ne kadar simüle edilmiş ve açık veri setlerinde başarılı sonuçlar elde edilmiş olsa da, gerçek IoT ve endüstriyel ağ ortamlarında henüz test edilmemiş olması, çalışmanın pratik uygulanabilirliğinin doğrulanması açısından bir boşluk oluşturmaktadır (Abd Elaziz vd., 2023). Bu yönleriyle ATS-IDS, GConvTrans ve D-MAGIC/SAGA gibi diğer yaklaşımlarla birlikte değerlendirildiğinde, IoT tabanlı saldırı tespit sistemlerinde önemli bir ilerleme sunmakla beraber, gerçek ortam doğrulaması, veri bağımlılığı ve adaptasyon sürekliliği gibi konularda geliştirmeye açık alanlar barındırmaktadır (Alsharif & Rawat, 2023).

2.1.23. Makine Öğrenimine Dayalı Hibrit Özellik Seçimi Kullanarak MITM Saldırılarının Önlenmesine Yönelik

Bu çalışmada (Swain vd., 2025), Aradaki Adam (Man-in-the-Middle, MITM) saldırılarının etkin biçimde tespit edilebilmesi amacıyla makine öğrenmesi, boyut indirgeme ve hibrit özellik seçimi tekniklerini (Kosugi vd., 2024) bütünleştiren çok aşamalı bir tespit çerçevesi önerilmiştir. Çalışma, yüksek boyutlu ağ trafiği verilerinin neden olduğu hesaplama yükü ve gürültü problemlerini azaltmayı hedefleyerek UNSW-NB15, CSE-CIC-IDS2018 ve CIC-DDoS2019 veri setleri üzerinde kapsamlı deneyler gerçekleştirmiştir. Önerilen yöntemde, en bilgilendirici özniteliklerin belirlenmesi için Özyinelemeli Özellik Eleme (RFE) ile Karşılıklı Bilgi (MI) ölçütü birlikte kullanılarak hem doğrusal hem de doğrusal olmayan bağımlılıkların yakalanması sağlanmış (Natarajan, 2020); ardından Temel Bileşen Analizi (PCA) ile boyut

indirgeme uygulanarak gereksiz değişkenler elenmiş, hesaplama maliyeti düşürülmüş ve aşırı öğrenme riski azaltılmıştır. Elde edilen optimize edilmiş özellik kümeleri SVM, RF, KNN ve Naive Bayes (NB) sınıflandırıcıları ile değerlendirilmiş ve özellikle RF modelinin tüm veri setlerinde en yüksek doğruluk ve kararlılığı sağlayarak en başarılı yaklaşım olduğu gösterilmiştir. Bununla birlikte çalışma bazı önemli sınırlamalara sahiptir: algoritma performansının modele bağlı olarak değişkenlik göstermesi ve özellikle KNN'nin yüksek boyutluluk karşısında düşük verimlilik sergilemesi, SVM ve NB gibi yöntemlerin veri karmaşıklığı arttıkça esneklik kaybı yaşaması, çerçevenin yalnızca statik veri setleri üzerinde test edilmiş olması nedeniyle sıfırıncı gün veya dinamik tehditlere karşı gerçek zamanlı uyarlanabilirlik sunamaması, karar süreçlerinin yorumlanabilirliğini artıracak Açıklanabilir Yapay Zekâ bileşenlerinin entegre edilmemiş olması, derin öğrenme tabanlı yöntemlerin (ör. CNN veya RNN) yüksek boyutlu ve yapılandırılmamış trafik verilerindeki potansiyelinin değerlendirilmemiş olması ve hibrit bulut ortamlarında gerçek zamanlı uygulama performansının henüz doğrulanmamış olması bu eksiklikler arasında yer almaktadır. Sonuç olarak çalışma, hibrit özellik seçimi ve boyut indirgeme ile MITM saldırı tespitinde doğruluk ve verimlilik açısından etkili bir temel sunmakta (Ma vd., 2020); ancak adaptif, açıklanabilir ve gerçek zamanlı ölçeklenebilir güvenlik sistemleri geliştirmek için ileri düzey yöntemlerle desteklenmesi gerekmektedir.

2.1.24. Siber Tehditleri Tespit Etmek İçin Hibrit Aşırı Gradyan Artırma ve Uzun Kısa Süreli Bellek Algoritması

Bu çalışmada (Amin vd., 2023), Geleneksel siber saldırı tespit sistemlerinin düşük tespit oranları, ölçeklenebilirlik problemleri ve yüksek boyutlu ağ verilerinden kaynaklanan performans kayıplarını gidermek amacıyla HXG-BLSTM adı verilen hibrit bir derin öğrenme mimarisi önerilmiştir. Çalışmada beş aşamalı bir iş akışı benimsenmiş; güncel ve çeşitli saldırı senaryolarını içeren CSE-CIC-IDS2018 veri seti üzerinde ön işleme adımı olarak yeniden örnekleme ile sınıf dengesizliği azaltılmış ve tüm özellikler 0–1 aralığına normalize edilmiştir. Özellik seçimi aşamasında başlangıçta ZOA (Trojovska vd., 2022), Parçacık Sürü Optimizasyonu (PSO) (Ragab, 2024) ve Genetik Algoritma (GA) (Sohail, 2023) gibi meta-sezgisel yöntemler değerlendirilmiş; ancak en yüksek performansı sunması nedeniyle Extreme Gradient Boosting (XGBoost) tabanlı önem derecelendirme yaklaşımı tercih edilerek 79 özellikten 22'si seçilmiş ve bellek kullanımı önemli ölçüde azaltılmıştır. Seçilen bu optimize edilmiş özellikler, zaman bağımlı örüntüleri yakalama ve gradyan kaybolması sorununu azaltma avantajına sahip LSTM tabanlı derin bir ağ ile sınıflandırılmıştır (Al Razib vd., 2022).

Deneysel sonuçlar, modelin ikili sınıflandırmada çok yüksek doğruluk, F1-skoru ve AUC değerleri elde ettiğini, çoklu sınıflandırmada ise farklı saldırı türlerinin neredeyse tamamını yüksek başarıyla ayırt edebildiğini göstermektedir. Bununla birlikte çalışma bazı önemli sınırlamalara sahiptir: özellik seçimi sonrasında belirli saldırı türlerinde beklenmedik performans düşüşleri gözlenmesi, hiperparametrelerin otomatik optimizasyon yerine manuel ve deneme-yanılma yöntemiyle ayarlanması, modelin yalnızca tek bir veri seti üzerinde doğrulanmış olması nedeniyle genellenebilirliğinin sınırlı kalması ve farklı veri dağılımlarında ya da sıfırıncı gün saldırılarında dayanıklılığının henüz test edilmemesi bu eksiklikler arasındadır. Ayrıca çok sayıda meta-sezgisel algoritmanın ayrı ayrı denenmesi yüksek hesaplama maliyeti doğurmakta ve özellik optimizasyonu sürecini zaman alıcı hâle getirmektedir. Sonuç olarak HXG-BLSTM yaklaşımı, özellik seçimi ile derin öğrenmeyi birleştirerek yüksek doğruluk ve verimlilik sunmakta; ancak otomatik hiperparametre ayarlama, çoklu veri seti doğrulaması ve hesaplama maliyetinin azaltılması gibi konularda ileri geliştirmelere ihtiyaç duymaktadır.

2.1.25. Tehdit Tespiti ve Risk Azaltma için Makine Öğrenimine Dayalı Hibrit Özellik Seçimi Kullanarak Yeni Nesil Güvenlik Duvarlarının Değerlendirilmesi

Bu çalışmada (Panse & Bandhu, 2025), yeni nesil güvenlik duvarlarının (NGFW) yalnızca kural tabanlı filtreleme yapan pasif yapılar olmaktan çıkarılarak, dinamik ve ölçeklenebilir biçimde tehdit algılayabilen akıllı savunma bileşenlerine dönüştürülmesi amacıyla makine öğrenmesi tabanlı hibrit bir özellik seçimi ve sınıflandırma çerçevesi önerilmiştir (Hamilton vd., 2020). Sistem, hesaplama maliyetini azaltırken gerçek zamanlı çalışabilirliği korumayı hedefleyen iki aşamalı bir özellik seçimi boru hattı üzerine inşa edilmiştir. İlk aşamada filtre tabanlı Bilgi Kazancı Oranı (Information Gain Ratio) ile özellikler önem derecesine göre sıralanmış, ikinci aşamada ise sarmalayıcı (wrapper) yaklaşım kapsamında k-means kümeleme ile sınıflar arası ayrılabilirlik doğrulanarak 41 özellikten en ayırt edici 18'lik bir alt küme elde edilmiştir. Bu optimize edilmiş özellikler, gürültülü ve dengesiz veri yapısına karşı dayanıklılığı ile bilinen topluluk öğrenmesi temelli RF sınıflandırıcısına beslenmiş ve model CSE-CIC-IDS2018 ile KDD Cup 1999 veri setleri üzerinde normal trafik ve DoS, Probe, R2L, U2R gibi farklı saldırı kategorileri için değerlendirilmiştir. Deneysel sonuçlar, yüksek doğruluk ve F1-skoru değerleriyle birlikte özellikle veri setinde az temsil edilen azınlık sınıflarındaki saldırılarda dahi güçlü tespit başarısı sağlandığını, modelin kısa eğitim süresi sayesinde yüksek hacimli ağlarda hızlı güncellenebildiğini ve düşük yanlış pozitif oranlarıyla gereksiz alarm üretimini minimize ettiğini göstermektedir (Mhawi vd., 2022). Bununla birlikte çalışma bazı

önemli sınırlamalara sahiptir: yaklaşımın büyük ölçüde çevrimdışı ve statik veri setleri üzerinde doğrulanmış olması nedeniyle canlı trafik akışlarına tam entegre gerçek zamanlı müdahale kapasitesinin henüz kanıtlanmamış olması, tehditlerin sürekli evrildiği ortamlarda modeli güncel tutacak çevrimiçi öğrenme mekanizmalarının bulunmaması, çok kiracılı bulut altyapılarında veri gizliliğini koruyarak eğitim yapılabilmesi için federe öğrenme gibi dağıtık yaklaşımların entegre edilmemiş olması ve RF tabanlı yapının karmaşık anomali örüntüleri veya sıfırıncı gün saldırılarında sınırlı kalabilmesi nedeniyle derin öğrenme tabanlı autoencoder ya da RNN mimarileriyle hibritleştirilme gereksinimi bu eksiklikler arasında yer almaktadır. Sonuç olarak önerilen çerçeve, NGFW sistemleri için yüksek doğruluklu ve hesaplama açısından verimli bir tehdit tespit altyapısı sunmakta; ancak gerçek zamanlı adaptasyon, gizlilik ve ileri seviye anomali algılama yetenekleri açısından ek geliştirmelere ihtiyaç duymaktadır (Panse & Bandhu, 2025).

2.1.26. IDS-NGNN: Azaltma ve Birleştirme Yöntemine Dayalı İç İçe Geçmiş Grafik Sinir Ağları Tabanlı Akıllı NIC Tabanlı Saldırı Tespit Sistemi

Bu çalışmada (Abu Bakar vd., 2025), geleneksel saldırı tespit sistemlerinin (IDS) ağ trafiğindeki karmaşık, çok seviyeli ilişkileri yeterince temsil edememesi ve düz (flat) grafik sinir ağlarının derin yapılarda aşırı pürüzsüzleşme (over-smoothing) problemi yaşaması temel çıkış noktası olarak ele alınmış; bu sorunları aşmak amacıyla yuvalanmış grafik sinir ağları (Nested GNN – NGNN) ile donanım hızlandırmayı birleştiren yenilikçi bir IDS mimarisi önerilmiştir. Önerilen IDS-NGNN yapısı. (Zhang vd., 2024), üç katmanlı hiyerarşik bir grafik öğrenme yaklaşımına dayanmaktadır: dahili GNN katmanı **(George & Thampi, 2018)** ana bilgisayar düzeyindeki süreç ve akış bazlı alt grafikleri analiz ederek ince taneli yerel davranışları yakalamakta, yuvalanmış grafik katmanı bu yerel temsilleri azaltma (reduce) ve birleştirme (merge) işlemleriyle hiyerarşik biçimde özetlemekte (Khraisat vd., 2024), harici GNN katmanı ise ana bilgisayarlar arası iletişim örüntülerini ve küresel ağ anomalilerini modelleyerek yerel ve küresel bağımlılıkların eş zamanlı öğrenilmesini sağlamaktadır. Mimari, yüksek bant genişliğine sahip ağlarda gerçek zamanlı çalışabilmek için SmartNIC tabanlı donanım hızlandırma (NVIDIA A30x/A100) ile desteklenmiş; bu sayede çıkarım süresi grafik başına milisaniye altına indirilerek 100 Gbps seviyesindeki ağlarda dahi düşük gecikmeli saldırı tespiti mümkün kılınmıştır. CIC-IDS-2017, CSE-CIC-IDS2018 ve ToN-IoT gibi çok sayıda kamuya açık veri seti üzerinde yapılan deneyler, modelin yüksek doğruluk ve F1-skoru sunduğunu, ayrıca grafik yapısında kasıtlı bozulmalar uygulanmasına rağmen performans kaybının sınırlı kaldığını göstermiştir **(Faheem & Al-Khasawneh, 2024)**. Bununla birlikte

çalışma bazı önemli sınırlamalara da sahiptir. Çok katmanlı ve hiyerarşik NGNN tasarımı, düz GNN modellerine kıyasla hesaplama yükünü belirgin ölçüde artırmaktadır. Kısa ve sığ işlem senaryolarında GPU kaynaklarının tam verimle kullanılamaması donanım altında kullanım sorununa yol açabilmektedir. Çok kiracılı bulut ortamlarında performans, izolasyon ve adalet (fairness) özellikleri henüz değerlendirilmemiştir. Kendini uyarlayabilen gelişmiş saldırı senaryolarına karşı modelin dayanıklılığı ileri çalışmaların konusu olarak bırakılmış ayrıca veri seti yanlılığı ve alan adaptasyonu problemleri, modelin farklı ağ ortamlarına genellenebilirliği açısından potansiyel bir kısıt oluşturmaktadır. Sonuç olarak IDS-NGNN, grafik tabanlı saldırı tespitinde hem mimari hem de sistem düzeyinde önemli bir ilerleme sunmakta ancak hesaplama maliyeti, donanım verimliliği ve adaptif tehditlere karşı dayanıklılık gibi başlıklarda ek geliştirmelere ihtiyaç duymaktadır.

2.1.27. Dengesiz Ağ Verilerinde Güvenilir Yerel Açıklamalar için 2B Kopulalar ile CoLIME

Bu çalışmada (Bacevicius vd., 2026), karmaşık makine öğrenmesi ve derin öğrenme modellerinin karar süreçlerini yerel düzeyde açıklamak için yaygın olarak kullanılan LIME yönteminin, öznitelikleri bağımsız ve normal dağılımlı varsayması nedeniyle ağ trafiği gibi yüksek derecede bağımlı ve dengesiz veri yapılarında güvenilir açıklamalar üretememesi problemine çözüm olarak CoLIME adı verilen kopula tabanlı yeni bir açıklanabilirlik çerçevesi önerilmiştir. Standart LIME'daki Gauss temelli sentetik örnekleme yaklaşımı yerine (Tan vd., 2023), öznitelikler arasındaki doğrusal olmayan bağımlılıkları korumak amacıyla Gaussian, Clayton, Gumbel ve Frank kopulaları kullanılarak çift değişkenli (2D) bağımlılık modellemesi yapılmış (Pawlicki vd., 2020); ayrıca her bir öznitelik için normal dağılım varsayımı terk edilerek lognormal, Gamma ve Weibull gibi veriyle daha uyumlu marjinal dağılımlar Kolmogorov–Smirnov testi ile seçilmiştir (Ayantayo vd., 2023). Yerel açıklamaların gerçek örnek çevresini daha iyi temsil edebilmesi için pertürbasyonlar belirli bir yarıçapla sınırlandırılmış veya bazı kritik özellikler sabitlenmiş; elde edilen sentetik örnekler üzerinden ise yerel davranışı taklit eden Ridge Regresyon tabanlı bir vekil model kullanılarak açıklamalar üretilmiştir (Nebaba vd., 2021). CIC-IDS2017, CSE-CIC-IDS2018 ve UNSW-NB15 veri setleri üzerinde RF ve XGBoost modelleriyle yapılan deneyler **(Pavithra & Venkata Vikas, 2024)**, CoLIME'ın vekil model sadakatini (fidelity) ve açıklama kararlılığını anlamlı ölçüde artırdığını, ayrıca saldırı davranışıyla doğrudan ilişkili protokol temelli özelliklere daha tutarlı ağırlıklar atayarak anlamsal doğruluğu güçlendirdiğini göstermiştir (Mirsadeghi vd., 2023). Bununla birlikte yöntem bazı sınırlamalara sahiptir: kopula temelli bağımlılık tahminleri seyrek

veri bölgelerinde kararsızlaşabilmekte, öznitelik bağımlılığının zayıf olduğu durumlarda sağlanan performans kazanımı sınırlı kalmakta, yüksek sadakat sağlayan KNN tabanlı vekil modeller önemli ölçüde hesaplama maliyeti doğurmakta, nadir ve uç olaylardaki (tail dependence) bağımlılıkların sentetik örneklerde olduğundan düşük tahmin edilmesi risk oluşturmakta ve her ne kadar klasik LIME'a kıyasla daha kararlı sonuçlar üretse de tamamen deterministik açıklama yöntemlerinin istikrar seviyesine henüz ulaşamamaktadır. Sonuç olarak CoLIME, dengesiz ve bağımlı yapıya sahip ağ güvenliği verilerinde daha güvenilir ve tutarlı yerel açıklamalar sunarak açıklanabilir yapay zekâ alanına önemli katkı sağlamakta (Knab vd., 2026); ancak ölçeklenebilirlik, uç durum modelleme ve deterministik kararlılık açısından ileri geliştirmelere ihtiyaç duymaktadır.

2.1.28. SZOA: Mikro şebeke planlaması ve yönetimi için geliştirilmiş bir sinerjik Zebra optimizasyon algoritması

Bu çalışmada (Cao & Wei, 2025), yeni nesil güvenlik duvarlarının (NGFW) yalnızca kural tabanlı filtreleme yapan pasif yapılar olmaktan çıkarılarak (Mei vd., 2022), dinamik ve ölçeklenebilir biçimde tehdit algılayabilen akıllı savunma bileşenlerine dönüştürülmesi amacıyla makine öğrenmesi tabanlı hibrit bir özellik seçimi ve sınıflandırma çerçevesi önerilmiştir. Sistem, hesaplama maliyetini azaltırken gerçek zamanlı çalışabilirliği korumayı hedefleyen iki aşamalı bir özellik seçimi boru hattı üzerine inşa edilmiştir: ilk aşamada filtre tabanlı Bilgi Kazancı Oranı (Information Gain Ratio) ile özellikler önem derecesine göre sıralanmış (Nguyen vd., 2016), ikinci aşamada ise sarmalayıcı (wrapper) yaklaşım kapsamında k-means kümeleme ile sınıflar arası ayrılabilirlik doğrulanarak 41 özellikten en ayırt edici 18'lik bir alt küme elde edilmiştir (Krishnan vd., 2018). Bu optimize edilmiş özellikler, gürültülü ve dengesiz veri yapısına karşı dayanıklılığı ile bilinen topluluk öğrenmesi temelli RF sınıflandırıcısına beslenmiş ve model CSE-CIC-IDS2018 ile KDD Cup 1999 veri setleri üzerinde normal trafik ve DoS, Probe, R2L, U2R gibi farklı saldırı kategorileri için değerlendirilmiştir. Deneysel sonuçlar, yüksek doğruluk ve F1-skoru değerleriyle birlikte özellikle veri setinde az temsil edilen azınlık sınıflarındaki saldırılarda dahi güçlü tespit başarısı sağlandığını, modelin kısa eğitim süresi sayesinde yüksek hacimli ağlarda hızlı güncellenebildiğini ve düşük yanlış pozitif oranlarıyla gereksiz alarm üretimini minimize ettiğini göstermektedir (Heidari vd., 2019). Bununla birlikte çalışma bazı önemli sınırlamalara sahiptir: yaklaşımın büyük ölçüde çevrimdışı ve statik veri setleri üzerinde doğrulanmış olması nedeniyle canlı trafik akışlarına tam entegre gerçek zamanlı müdahale kapasitesinin henüz kanıtlanmamış olması, tehditlerin sürekli evrildiği ortamlarda modeli güncel tutacak çevrimiçi

öğrenme mekanizmalarının bulunmaması, çok kiracılı bulut altyapılarında veri gizliliğini koruyarak eğitim yapılabilmesi için federe öğrenme gibi dağıtık yaklaşımların entegre edilmemiş olması ve RF tabanlı yapının karmaşık anomali örüntüleri veya sıfırıncı gün saldırılarında sınırlı kalabilmesi nedeniyle derin öğrenme tabanlı autoencoder ya da RNN mimarileriyle hibritleştirilme gereksinimi bu eksiklikler arasında yer almaktadır. Sonuç olarak önerilen çerçeve, NGFW sistemleri için yüksek doğruluklu ve hesaplama açısından verimli bir tehdit tespit altyapısı sunmakta; ancak gerçek zamanlı adaptasyon, gizlilik ve ileri seviye anomali algılama yetenekleri açısından ek geliştirmelere ihtiyaç duymaktadır (Cao & Wei, 2025).

2.1.29. Birleştirilmiş uzamsal-zamansal dönüştürücü ve nedensel evrimsel ağlar aracılığıyla çok boyutlu gizli trafik saldırısı tespiti

Bu çalışmada (Chi, 2025), düşük hızlı DDoS (ör. Slowloris), gelişmiş kalıcı tehdit (APT) sızmaları (Somani vd., 2017) ve ağ steganografisi gibi geleneksel imza veya istatistik tabanlı yöntemlerden kaçabilen gizli (covert) saldırıların tespitine yönelik (Faria vd., 2020) olarak ST-CCNet (Stealth-Targeted Criss-Cross Network) adı verilen çok boyutlu ve hibrit bir derin öğrenme mimarisi önerilmiştir. Model, trafik verilerini zaman serileri, protokol başlık alanları (TCP bayrakları vb.) ve istatistiksel akış metrikleri olmak üzere üç heterojen boyutta ele alarak öğrenilebilir bir gömme katmanı aracılığıyla ortak bir temsil uzayına taşımaktadır (Huang vd., 2022). Yerel ve ani davranış değişimlerini yakalamak için nedensel ve genişletilmiş (dilated) konvolüsyonlardan oluşan Causal Convolution modülü kullanılırken, uzun vadeli küresel bağımlılıkların modellenmesi amacıyla çok başlı öz-dikkat mekanizmasına sahip Spatio-Temporal Transformer bileşeni entegre edilmiştir; ayrıca geçitli dinamik birleştirme (gated fusion) yapısı (Zhang vd., 2024) ile yerel ve küresel temsillerin katkısı trafik türüne göre uyarlanarak hem kısa süreli mikro anomaliler hem de saatler süren komuta-kontrol örüntüleri eş zamanlı biçimde yakalanabilmektedir. CSE-CIC-IDS2018 üzerinde gerçekleştirilen deneyler, modelin yüksek F1 ve geri çağırma değerleriyle mevcut yaklaşımları geride bıraktığını, özellikle Slowloris gibi düşük hızlı saldırılarda belirgin performans artışı sağladığını ve görece düşük parametre sayısı ile milisaniye mertebesinde çıkarım süresi sunarak gerçek zamanlı uygulamalara uygunluk gösterdiğini ortaya koymuştur. Bununla birlikte çalışma bazı sınırlamalar içermektedir: modelin farklı protokol türleri arasında genelleme yeteneğinin henüz tam olarak doğrulanmamış olması, karar süreçlerini açıklayacak açıklanabilir yapay zekâ (XAI) mekanizmalarının bulunmaması, farklı ağ ortamlarına uyum sağlayacak alan adaptasyonu tekniklerinin eksikliği, dizi uzunluğu arttıkça nedensel

konvolüsyon ve dikkat mekanizmalarının hesaplama karmaşıklığının yükselmesi nedeniyle uzun süreli APT trafiğinde maliyet artışı yaşanabilmesi ve her ne kadar saha testleri yapılmış olsa da evrensel ve ölçeklenebilir bir gerçek zamanlı savunma çerçevesine tam entegrasyonun henüz tamamlanmamış olması bu kısıtlar arasında yer almaktadır. Sonuç olarak ST-CCNet, gizli ve düşük yoğunluklu saldırıların çok boyutlu analizi için güçlü ve pratik bir çözüm sunmakta; ancak genellenebilirlik, açıklanabilirlik ve hesaplama verimliliği açısından ileri geliştirmelere ihtiyaç duymaktadır.

2.1.30. CSE-CIC-IDS2018 Veri Kümesinde Sınıflandırma Performansını İyileştirmek için Ağırlıklı Topluluk Sıralaması ile Özellik Seçimi

Bu çalışmada **(Göcs & Johanyák, 2023)**, yüksek boyutlu ağ trafiği verilerinde gereksiz ve gürültülü özelliklerin neden olduğu model karmaşıklığını azaltmak ve sınıflandırma başarımını artırmak amacıyla, çoklu istatistiksel özellik seçimi yöntemlerini birleştiren ağırlıklı topluluk (ensemble) tabanlı bir özellik sıralama yaklaşımı önerilmiştir. Çalışma, tek bir filtre yöntemine bağımlı kalmanın yanlı veya dengesiz sonuçlar üretebileceği varsayımından hareketle Bilgi Kazancı (IG), Kazanç Oranı (GR), Ki-kare (χ^2), Simetrik Belirsizlik (SU), Relief ve ANOVA gibi altı farklı ölçütü birlikte kullanmakta ve her bir yöntemin ürettiği skorları sabit ortalama yerine ağırlıklı ortalama ile birleştirerek nihai bir önem puanı hesaplamaktadır. Bu ağırlıkların optimal değerleri ise çok sayıda deneme gerektirmeden deney tasarımı yaklaşımı sunan Taguchi yöntemi ile belirlenmiş, böylece hem hesaplama maliyeti kontrol altında tutulmuş hem de en etkili kombinasyon sistematik biçimde elde edilmiştir (Hashemi vd., 2022). Elde edilen özellik alt kümeleri, CSE-CIC-IDS2018 üzerinde DT, RF ve Destek Vektör Makineleri (SVM) sınıflandırıcılarıyla değerlendirilmiştir. Sonuçlar, özellikle belirli saldırı senaryolarında özellik sayısının dramatik biçimde azaltılmasına rağmen sınıflandırma doğruluğu, kesinlik, geri çağırma ve F1 skorlarında anlamlı iyileşmeler sağlandığını ve eğitim süresinin kısalarak aşırı öğrenme riskinin azaldığını göstermiştir **(Sumant & Patil, 2022)**. Bununla birlikte çalışma bazı sınırlamalara sahiptir. Çoklu yöntemlerin birlikte çalıştırılmasını gerektiren topluluk tabanlı yapı yüksek hesaplama maliyeti doğurmakta ve doğruluk–zaman dengesi kurulmasını zorlaştırmakta, deneyler veri setindeki tüm saldırı türlerini kapsamak yerine belirli alt kategorilerle sınırlandırılmış bulunmakta, yöntemin dinamik ve sıfırıncı gün tehditlere karşı çevrimiçi adaptasyon kabiliyeti henüz değerlendirilmemiş olmakta ve ağırlık optimizasyonunda bulanık mantık ya da adaptif öğrenme gibi daha gelişmiş tekniklerin entegrasyonu gelecek çalışmalar için açık bir araştırma alanı olarak bırakılmaktadır (Hoque vd., 2018). Sonuç olarak önerilen ağırlıklı topluluk yaklaşımı, özellik seçimi yoluyla IDS

performansını artırmada etkili ve pratik bir çözüm sunmakta; ancak ölçeklenebilirlik, genellenebilirlik ve adaptif tehditlere karşı dayanıklılık açısından ek geliştirmelere ihtiyaç duymaktadır.

2.1.31. Hızlı ve Otomatik Olay Müdahalesi için Akıllı Bir Dijital Adli Değerlendirme Aracı

Bu çalışmada (Jagathbala vd., 2025), artan siber saldırı hacmi ve karmaşıklığı karşısında geleneksel dijital adli bilişim araçlarının yavaş, manuel ve parçalı yapısının olay müdahale süreçlerini geciktirdiği probleminden hareketle, disk, dosya ve ağ analizini tek bir otomatikleştirilmiş platformda birleştiren yapay zekâ destekli bir dijital adli tıp triyaj sistemi önerilmiştir. ForenXplorer mimarisi, çok katmanlı ve bütünleşik bir analiz yaklaşımı benimsemektedir: ağ analizi modülü, CSE-CIC-IDS2018 veri seti ile eğitilmiş Destek Vektör Makinesi (SVM) temelli bir model aracılığıyla PCAP dosyalarından elde edilen trafik akışlarında anormallikleri tespit etmekte, daha karmaşık örüntüler için ise mekânsal ve zamansal bağımlılıkları modelleyebilen CNN–LSTM hibrit mimarisini desteklemektedir; dosya analizi katmanı meta veri çıkarımı, imza doğrulama ve SHA-256 kriptografik özetleme ile kanıt bütünlüğünü sağlamaktayken (Chan vd., 2013); disk görüntüleme bileşeni imaj alma ve doğrulama işlemlerini otomatikleştirerek özellikle FAT dosya sistemlerinde derinlemesine inceleme gerçekleştirmektedir. Sistem, React.js tabanlı kullanıcı arayüzü ve Flask arka ucu ile geliştirilmiş, Docker konteynerleri ile ölçeklenebilir ve taşınabilir bir yapı sunmaktadır. Deneysel sonuçlar, manuel analiz süreçlerine kıyasla inceleme süresinde belirgin azalma sağlandığını, ağ analizi modülünün yüksek doğruluk ve düşük yanlış pozitif oranı elde ettiğini, dosya bütünlüğü doğrulamasında yüksek başarı sağlandığını ve otomatik raporlamayla hızlı karar destek çıktıları üretilebildiğini göstermektedir. Bununla birlikte sistem bazı önemli sınırlamalara sahiptir. Denetimli öğrenmeye dayalı yapısı nedeniyle sıfırıncı gün saldırılarına karşı genelleme kapasitesinin sınırlı olması, yalnızca FAT dosya sistemlerini desteklemesi ve modern NTFS veya ext4 gibi sistemlerle uyumsuzluğu, büyük boyutlu disk imajlarında SHA-256 hesaplamasının önemli bir performans darboğazı oluşturması (Coşar vd., 2024), harici tehdit istihbaratı servisleriyle tam entegrasyonun bulunmaması ve uçucu tehditlerin tespiti için kritik olan bellek (RAM) adli bilişimi bileşeninin henüz sisteme dahil edilmemiş olması bu eksiklikler arasında yer almaktadır. Sonuç olarak ForenXplorer, hızlı ve otomatik olay müdahalesi için bütünleşik ve pratik bir adli bilişim triyaj platformu sunmakta; ancak ölçeklenebilirlik, genellenebilirlik ve kapsamlı tehdit kapsaması açısından ek geliştirmelere ihtiyaç duymaktadır.

2.1.32. Hibrit grafik tabanlı evrimsel ağ ve transformatör mimarisi kullanan ağa izinsiz giriş tespiti

Bu çalışmada (Appiahene vd., 2026), bulut bilişim ve IoT ortamlarında giderek karmaşıklaşan ağ trafiği örüntülerinin klasik makine öğrenmesi ve düz derin öğrenme modelleriyle yeterince temsil edilememesi problemine çözüm olarak GConvTrans adı verilen hibrit bir grafik–transformer mimarisi önerilmiştir. Çalışma, geleneksel tablo yapısındaki ağ akış verilerini, her örneğin en yakın komşularıyla ilişkilendirildiği hesaplamalı bir grafik temsile dönüştürerek hem yerel hem de küresel bağımlılıkların birlikte modellenmesini hedeflemektedir. Bu kapsamda Grafik Konvolüsyonel Ağ (GCN) katmanı, düğüm komşuluklarını normalize ederek yerel yapısal ilişkileri ve özellik etkileşimlerini gizli temsillere aktarırken; Transformer kodlayıcı katmanı, çok başlı öz-dikkat mekanizması aracılığıyla uzak düğümler arasındaki uzun menzilli bağımlılıkları öğrenerek küresel bağlamı yakalamaktadır. Veri ön işleme aşamasında CSE-CIC-IDS2018 veri setindeki dengesiz sınıf dağılımını azaltmak için SMOTE uygulanmış ve her giriş, belirli sayıda komşuyla bağlanarak grafik yapısına entegre edilmiştir. Deneysel sonuçlar, modelin özellikle doğrulama ve test aşamalarında yüksek doğruluk oranları elde ettiğini, t-SNE görselleştirmelerinin sınıflar arasında belirgin ayrışma sağladığını ve nispeten sığ mimarisi sayesinde hesaplama ve bellek gereksinimlerinin düşük kalması nedeniyle gerçek zamanlı izleme senaryolarına uygunluk gösterdiğini ortaya koymuştur. Bununla birlikte çalışma bazı sınırlamalara sahiptir: deneylerin yalnızca veri setinin belirli bir alt kümesiyle sınırlandırılması genellenebilirliği kısıtlamakta, GCN-only veya Transformer-only gibi temel modellerle ayrıntılı ablasyon ve karşılaştırmalı analizlerin eksikliği bileşenlerin katkısının net biçimde değerlendirilmesini engellemekte, modelin farklı grafik görevleri veya daha çeşitli saldırı türlerinde henüz test edilmemiş olması uygulama kapsamını daraltmakta ve Transformer katmanındaki bellek tüketimini azaltmak için kullanılan optimizasyonlara rağmen çok daha büyük ölçekli grafiklerde ölçeklenebilirlik hâlâ araştırma gerektirmektedir. Sonuç olarak GConvTrans, grafik tabanlı yerel yapı öğrenimi ile küresel dikkat mekanizmalarını birleştirerek ağ saldırı tespitinde etkili ve verimli bir yaklaşım sunmakta; ancak genellenebilirlik, kapsamlı karşılaştırmalar ve büyük ölçekli ortamlarda performans açısından ek geliştirmelere ihtiyaç duymaktadır.

2.1.33. Modern ve Dengesiz Bir Veri Kümesinde Makine Öğrenimine Dayalı Saldırı Tespit Sistemlerinin Güvenilirliğini En Üst Düzeye Çıkarma

Bu çalışmada (Ponkumar vd., 2025), güncel ve gerçekçi saldırı senaryoları içeren CSE-CIC-IDS2018 veri seti üzerinde, özellikle sınıf dengesizliğinin saldırı tespit sistemlerinin güvenilirliği üzerindeki olumsuz etkilerini azaltmaya odaklanan kapsamlı bir makine öğrenmesi tabanlı IDS çerçevesi önerilmiştir **(Fitni & Ramli, 2020)**. Çalışma, veri temizleme ve normalizasyon adımlarıyla başlayan çok aşamalı bir boru hattı tasarlamış; nadir görülen saldırı sınıflarının öğrenimini güçlendirmek amacıyla SMOTE yöntemi ile sentetik azınlık örnekleri üretmiş, ardından boyut indirgeme ve hız optimizasyonu için Temel Bileşen Analizi (PCA) ile RF önem derecelendirmesini birleştiren hibrit bir özellik seçimi yaklaşımı kullanarak en bilgilendirici 11 özelliği belirlemiştir. Bu optimize edilmiş özellik kümesi üzerinde AdaBoost, RF, KNN, DT ve Gradient Boosting gibi farklı sınıflandırıcılar değerlendirilmiş ve özellikle RF ile KNN modellerinin yüksek genel doğruluk oranları ve düşük hesaplama maliyeti sunduğu gözlemlenmiştir. Ayrıca normal trafik ile Botnet, Brute-force ve DDoS gibi yaygın saldırı türlerinin yüksek hassasiyetle ayrıştırılabildiği ve özellik seçimi sayesinde CPU kullanım süresi ile eğitim maliyetinin önemli ölçüde azaldığı rapor edilmiştir. Bununla birlikte çalışma bazı kritik sınırlamalar içermektedir: sistemin Sızma (Infiltration) ve Web Saldırıs (Web Attack) gibi daha sinsi ve düşük yoğunluklu saldırı türlerini tespit etmekte başarısız kalması ve bu sınıfların büyük bölümünü yanlışlıkla normal trafik olarak etiketlemesi, deneylerin çoğunlukla kontrollü laboratuvar ortamlarında gerçekleştirilmiş olması nedeniyle gerçek ağlardaki dinamik trafik değişimlerine karşı dayanıklılığın henüz doğrulanmamış olması, modellerin gerçek zamanlı dağıtıma hazır bulunmaması ve ek optimizasyon gerektirmesi ile SMOTE'un hatalı uygulanması durumunda karar sınırlarını bozabilecek yapay örnekler üretme riski bu eksiklikler arasında yer almaktadır. Sonuç olarak çalışma, veri dengesizliğine karşı hibrit özellik seçimi ve örnekleme stratejilerinin IDS güvenilirliğini artırabileceğini göstermekte; ancak nadir ve karmaşık saldırı türlerinin tespiti ile gerçek zamanlı uygulanabilirlik açısından daha gelişmiş ve adaptif yaklaşımlara ihtiyaç duyulduğunu ortaya koymaktadır.

2.2 Zebra Optimizasyon Algorıtması (ZOA) ile yapılmış çalışmalar

Literatürde zebralardan ilham alarak oluşturulmuş ZOA meta sezgisel algoritması ile bazı problemlere çözüm öneren çalışmalar mevcuttur.

Çizelge 2.2 ZOA ile yapılan çalışmalar

Yazarlar	Algoritma	Uygulama Alanı	Temel Sonuçlar ve Katkılar

(Trojovska vd., 2022)	ZOA	Genel Optimizasyon ve Mühendislik Tasarımı	Algoritmanın temeli atıldı; 68 benchmark fonksiyonu ve mühendislik problemlerinde yüksek başarı gösterdi
(Elymany vd., 2024)	ZOA-ANFIS	Yenilenebilir Enerji (PV ve Rüzgar)	ZOA, ANFIS parametrelerini optimize ederek hesaplama süresini PV için %26,17, rüzgar için %35,5 azalttı.
(Abdelmalek vd., 2024b)	ZOA	Fotovoltaik (PV) Sistemler (MPPT)	Kısmi gölgeleme koşullarında %99,95 izleme verimliliğine ulaşıldı ve rakip yöntemlerden %10-15 daha hızlı yakınsama sağlandı.
(Arkwazee & Abttan, 2025)	IMZOA (Geliştirilmiş ZOA)	Güç Sistemleri (Ekonomik ve Emisyon Dağıtımı)	Gauss dağılımı ve dinamik savunma ile maliyetlerde %0,80 azalma ve yüksek çözüm kararlılığı sağlandı.
(Ozbay, 2025)	CLESQ-ZOA	Özellik Seçimi ve Karaciğer Kanseri Vakası	Logaritmik sarmal ve kaotik haritalama ile optimal özellik alt kümeleri yüksek doğrulukla belirlendi
(Faizan & Mohamad, 2026)	ZOA + KNN	Makine Öğrenmesi (Özellik Seçimi)	KNN ile entegre edilen ZOA, özellik sayısını %91'e varan oranlarda azaltırken sınıflandırma doğruluğunu artırdı
(Z. Wang vd., 2025)	MZOA (Çoklu Strateji Füzyonlu)	Robotik (Yol Planlama)	Levy uçuşu ve lens tersine öğrenme ile yol uzunluğu temel ZOA'ya göre %8,7 kısaltıldı
(Huang vd., 2022)	HZOA (Hibrit ZOA - GWO)	3D İHA Yol Planlama	Gri Kurt Optimize Edici (GWO) hiyerarşisi ZOA'ya eklenerek karmaşık uzayda daha pürüzsüz ve düşük maliyetli rotalar oluşturuldu.
(Balci & Duman, 2025)	FFDBZOA-OBL	Güç Sistemleri Planlaması (TEP)	Bulanık mantık ve karşıt tabanlı öğrenme (OBL) ile yatırım maliyetlerinde %82,6 başarı oranı sağlandı
(C. Li vd., 2025a)	MI-ZOA (Çoklu Strateji Entegre)	Ağ Güvenliği (DDoS Tespiti)	MSCNN-BiGRU-SHA modelinin hiperparametreleri MI-ZOA ile optimize edilerek %99,71 doğruluk sağlandı.
(Punia vd., 2025)	EZOA (Levy Uçuşu + LOBL)	Güvenilirlik ve Mühendislik Optimizasyonu	Mercek Karşıt Tabanlı Öğrenme (LOBL) ile global yakınsama doğruluğu artırıldı ve yerel optimumdan kaçış sağlandı
(Shang-Guan vd., 2025)	c5SinZOA	Endüstriyel Yumuşak Sensör Modelleme	Petrol üretim süreçlerinde MLP model parametreleri optimize edilerek en düşük tahmin hata oranlarına ulaşıldı
(Stephan vd., 2025)	HIC-ZO, AEE-ZO (5 Farklı Varyant)	Yüksek Boyutlu Veri Madenciliği	Hibrit eylemsizlik kontrolü ve adaptif keşif stratejileri ile özellik azaltma verimliliği GA ve PSO'nun üzerine çıktı
(Ali, 2026)	GWO-ZOA (Hibrit)	Karmaşık Optimizasyon Problemleri	GWO'nun sömürü ve ZOA'nın keşif yetenekleri birleştirilerek CEC2017/2020 testlerinde en iyi performans elde edildi
(Devprabhakar vd., 2025)	EMTSZOA (Üçlü Strateji)	Kablosuz Sensör Ağları (WSN)	Sinüs Kosinüs Stratejisi (SCS) ve dinamik anahtarlama ile ağ ömrü %93,80 oranına çıkarıldı

2.2.1. Zebra Optimizasyon Algoritması: Optimizasyon Problemlerini Çözmek İçin Yeni Bir Biyolojik Esinlenmeli Optimizasyon Algoritması

Bu çalışmada (Trojovska vd., 2022), ZOA zebraların doğadaki sosyal etkileşimleri ve hayatta kalma stratejilerinden esinlenen yeni bir biyolojik temelli meta-sezgisel optimizasyon algoritmasıdır. Algoritmanın temel biyolojik dayanağı, zebraların beslenme arayışı sırasında

sergiledikleri sürü davranışı ile yırtıcılara karşı geliştirdikleri savunma mekanizmalarının matematiksel olarak modellenmesine dayanmaktadır. Bu bağlamda, sürüye rehberlik eden "öncü zebra" davranışı global arama (keşif) sürecini temsil ederken, aslan gibi yırtıcılara karşı sergilenen zikzak kaçış ve toplu savunma tepkileri yerel arama (sömürü) mekanizması olarak kurgulanmıştır (Trojovska vd., 2022). ZOA'nın performansı; unimodal, yüksek boyutlu multimodal ve sabit boyutlu multimodal fonksiyonların yanı sıra CEC2015 ve CEC2017 test setlerini kapsayan toplam 68 benchmark fonksiyonu üzerinde değerlendirilmiş ve elde edilen sonuçlar GWO, PSO, GA, TSA, WOA, GSA, TLBO, MPA ve QANA gibi dokuz yaygın meta-sezgisel algoritma ile karşılaştırılmıştır. Deneysel ve istatistiksel analizler, ZOA'nın keşif ve sömürü aşamaları arasında hassas bir denge kurarak birçok test fonksiyonunda rakip algoritmalara kıyasla istatistiksel olarak anlamlı bir üstünlük sağladığını ortaya koymuştur. Ayrıca çalışma kapsamında gerilme/sıkıştırma yayı, kaynaklı kiriş, hız düşürücü ve basınçlı kap tasarımı gibi dört farklı mühendislik problemi ele alınarak, ZOA'nın gerçek dünya mühendislik uygulamalarındaki etkinliği doğrulanmıştır. Bununla birlikte, çalışmada algoritmanın ikili/binary optimizasyon problemleri, çok amaçlı senaryolar ve veri yoğun makine öğrenmesi uygulamaları üzerindeki performansı kapsam dışı bırakılmış; parametre adaptasyonu, erken yakınsama riski ve gerçek zamanlı sistemlerde hesaplama maliyeti gibi konular gelecekteki çalışmalar için açık araştırma alanları olarak belirtilmiştir (Trojovska vd., 2022).

2.2.2. Hibrit Mikroşebekeler İçin ZOA ve Yapay Gorilla Sürü Optimizasyonu ile Hibrit Optimizasyon-ANFIS Tabanlı MPPT

Bu çalışmada (Elymany vd., 2024), yenilenebilir enerji sistemlerinde güç üretim verimliliğini artırmak amacıyla ZOA tabanlı bir ANFIS modeli önerilmiştir. Çalışmada fotovoltaik (PV) ve rüzgâr enerjisi dönüşüm sistemlerinden (WECS) oluşan hibrit bir mikroşebeke yapısı ele alınmış; sistem batarya enerji depolama sistemi (BESS), hidrojen enerji depolama sistemi (HESS) ve süperkapasitör enerji depolama sistemi (SESS) ile desteklenerek yenilenebilir kaynakların kesintili doğasından kaynaklanan dalgalanmaların dengelenmesi hedeflenmiştir. Yazarlar, ZOA algoritmasını ilk kez hem PV hem de rüzgâr enerjisi sistemlerinde Maksimum Güç Noktası Takibi (MPPT) sürecinde ANFIS parametrelerini optimize etmek amacıyla kullanmış ve yöntemi Artificial Gorilla Troops Optimizer (GTO) tabanlı bir yaklaşımla karşılaştırmıştır (Elymany vd., 2024). MATLAB/Simulink ortamında gerçekleştirilen benzetim sonuçları, ZOA tabanlı yaklaşımın yüksek yakınsama hızı ve izleme

doğruluğu sayesinde sistem performansını artırdığını; ayrıca hesaplama süresi açısından fotovoltaik sistemlerde %26,17, rüzgâr enerjisi sistemlerinde ise %35,5 oranında iyileşme sağladığını ortaya koymuştur. Bu sonuçlar, ZOA'nın özellikle zaman verimliliği gerektiren enerji yönetimi uygulamalarında etkili bir optimizasyon aracı olduğunu göstermektedir. Bununla birlikte, çalışma ağırlıklı olarak benzetim temelli analizlerle sınırlı kalmıştır. Uzun dönemli gerçek zamanlı saha verileri, farklı iklim koşulları ve ani arıza senaryoları altında sistemin davranışı ayrıntılı biçimde incelenmemiştir. Ayrıca önerilen yaklaşımın kontrol karmaşıklığı, gömülü sistemlerde gerçek zamanlı uygulanabilirliği ve siber-fiziksel güvenlik boyutu gibi konular açık araştırma alanları olarak kalmaktadır.

2.2.3. Kısmi Gölgeleme Koşulları Altında Fotovoltaik Sistemlerde Etkili Zebra Optimizasyon Algoritması Tabanlı MPPT'nin Deneysel Doğrulaması

Bu çalışmada (Abdelmalek vd., 2024b) fotovoltaik (PV) sistemlerde kısmi ve dinamik gölgeleme koşulları altında Küresel Maksimum Güç Noktasının (GMPP) etkin biçimde izlenebilmesi için ZOA tabanlı bir MPPT yöntemi önermiştir. Çalışmada ZOA'nın keşif ve sömürü dengesini başarılı şekilde kurarak yerel maksimum güç noktaları (LMPP) ile GMPP arasındaki ayrımı güvenilir biçimde gerçekleştirdiği gösterilmiş; deneysel sonuçlar ZOA'nın %99,95 izleme verimliliği ve 0,875 saniyelik ortalama izleme süresi ile GWO, PSO, WOA ve FPA gibi yöntemleri geride bıraktığını ortaya koymuştur. Ayrıca ZOA'nın maksimum güç noktasına %10–15 daha hızlı ulaştığı ve işletme noktası dalgalanmalarını azaltarak enerji kayıplarını minimize ettiği raporlanmıştır. Bu bulgular, ZOA'nın PV sistemlerinde zorlu çevresel koşullar altında dahi sağlam ve güvenilir bir MPPT çözümü sunduğunu göstermektedir (Abdelmalek vd., 2024b).

2.2.4. Çoklu Strateji Entegre Zebra Optimizasyon Algoritmasına Dayalı DDoS Saldırısı Otonom Algılama Modeli

Bu çalışmada (C. Li vd., 2025), bilinmeyen ve karmaşık DDoS saldırılarının tespiti için Multi-Scale CNN–BiGRU–Single Headed Attention (MSCNN-BiGRU-SHA) tabanlı otonom bir derin öğrenme modeli önermiştir. Çalışmada, modelin Conv_filter ve GRU_unit hiperparametreleri (R. Wang vd., 2019), klasik ZOA'nın geliştirilmiş bir sürümü olan Multi-Strategy Integrated ZOA (MI-ZOA) ile optimize edilmiştir. Önerilen yaklaşım, CICDDoS2019 veri seti üzerinde eğitilmiş ve daha önce görülmemiş saldırıları değerlendirmek amacıyla oluşturulan (Stephan vd., 2025) GINKS2023 veri seti üzerinde test edilmiştir. Deneysel

sonuçlar, MI-ZOA ile optimize edilen MSCNN-BiGRU-SHA modelinin CICDDoS2019 üzerinde %99,71 doğruluk, GINKS2023 üzerinde ise %93,86 doğruluk elde ettiğini ve klasik ZOA, Grid Search, Random Search ve Bayesian Optimization yöntemlerine kıyasla doğruluk, geri çağırma ve F1-skoru açısından belirgin üstünlük sağladığını göstermiştir. Bununla birlikte, çalışma ağırlıklı olarak DDoS saldırılarına odaklanmış olup, modelin farklı saldırı türlerine genellenebilirliği, gerçek zamanlı ağ ortamlarında gecikme ve kaynak tüketimi gibi pratik kısıtlar ile ayrıntılı biçimde ele alınmamıştır. Ayrıca veri seti oluşturma sürecinin kontrollü bir ortamda gerçekleştirilmiş olması, modelin gerçek dünya ağ trafiğindeki adaptasyon yeteneği açısından gelecekte ek doğrulamalara ihtiyaç duyulduğunu göstermektedir (C. Li vd., 2025).

2.2.5. Optimizasyon Problemlerini Çözmek için Hibrit Gri Kurt Optimizasyon Algoritması – Zebra Optimizasyon Algoritması

Bu çalışmada (Ali, 2026), Gri Kurt Optimizasyon Algoritması'nın sömürü yeteneği ile ZOA'nın keşif kapasitesini bir araya getiren sıralı bir hibrit yöntem (GWO-ZOA) önerilmiştir. Bu hibrit yapı, Gri Kurt Optimizasyon Algoritması'nda görülen çeşitlilik kaybı ile ZOA'nın görece yavaş yakınsama davranışını dengelemeyi amaçlamaktadır. Standart karşılaştırma fonksiyonları ve gerçek dünya mühendislik tasarım problemleri üzerinde gerçekleştirilen deneysel değerlendirmeler, GWO-ZOA yönteminin hem yakınsama doğruluğu hem de küresel arama başarımı açısından karşılaştırılan yöntemlere kıyasla istatistiksel olarak anlamlı bir üstünlük sağladığını göstermiştir. Elde edilen sonuçlar, önerilen hibrit yaklaşımın karmaşık optimizasyon problemlerinde kararlı ve güvenilir bir performans sunduğunu ortaya koymaktadır (Ali, 2026).

2.2.6. PLSR için çok stratejili füzyon zebra optimizasyon algoritmasına dayalı karakteristik dalga boyu seçimi

Bu çalışmada (Liu vd., 2025), yakın kızılötesi (NIR) spektral verilerdeki yüksek boyutluluk ve karmaşıklığı azaltmak amacıyla Çok Stratejili Birleştirilmiş ZOA (MFZOA) önerilmiştir. MFZOA yaklaşımı; başlangıç popülasyonunun daha dengeli oluşturulması için iyi nokta kümesi yöntemini, keşif ve sömürü dengesinin sağlanması için doğrusal olmayan yakınsama faktörünü ve yerel en iyi çözümlerden kaçınmak amacıyla uyarlamalı Gauss mutasyon stratejisini bir arada kullanmaktadır. Beş farklı NIR veri kümesi üzerinde gerçekleştirilen deneysel değerlendirmeler, MFZOA'nın daha az sayıda dalga boyu seçmesine rağmen daha yüksek tahmin doğruluğu elde ettiğini ve geleneksel dalga boyu seçim yöntemlerine kıyasla daha iyi bir genelleme başarımı sunduğunu göstermiştir. Elde edilen

sonuçlar, MFZOA yaklaşımının spektral değişken seçimi problemlerinde etkili ve güvenilir bir optimizasyon yöntemi olduğunu ortaya koymaktadır (Liu vd., 2025).

2.2.7. Enerji sistemlerinde sürdürülebilir birleşik ekonomik ve emisyon dağıtımı için geliştirilmiş zebra optimizasyon algoritması

Bu çalışmada (Arkwazee & Abttan, 2025), güç sistemlerinde birleşik ekonomik ve emisyon yük dağıtımı (CEED) problemini çözmek amacıyla Geliştirilmiş ZOA (IMZOA) önerilmiştir. Önerilen yöntemde, standart ZOA'nın keşif yeteneğini artırmak için beslenme aşamasına Gauss dağılımı entegre edilmiş, arama stratejileri ise iterasyon sürecine bağlı uyarlamalı bir olasılık mekanizması ile yönetilmiştir. Ayrıca savunma tepkilerinde uygulanan dinamik ölçeklendirme, erken aşamalarda geniş arama yapılmasını, ilerleyen aşamalarda ise çözümün hassas biçimde iyileştirilmesini sağlamıştır. IMZOA, 6, 31 ve 110 üniteli farklı ölçeklerdeki güç sistemleri üzerinde test edilmiş; elde edilen sonuçlar, yöntemin yakıt maliyeti ve emisyon değerlerini rakip algoritmalara kıyasla düşürdüğünü ve özellikle büyük ölçekli sistemlerde yüksek çözüm kararlılığı ve sağlamlık sunduğunu göstermiştir (Arkwazee & Abttan, 2025).

2.2.8. Küresel Optimizasyon ve Özellik Seçimi Problemleri için Çoklu Stratejilere Sahip Geliştirilmiş Zebra Optimizasyon Algoritması: Hepatocellüler Karsinom Vaka Çalışması

Bu çalışmada (Ozbay, 2025), ZOA keşif ve sömürü dengesinde yaşadığı sınırlamaları gidermek amacıyla Çok Stratejili Geliştirilmiş ZOA (CLESQ-ZOA) önerilmiştir. Önerilen yaklaşımda, başlangıç popülasyonunun çeşitliliğini artırmak için kaotik haritalama, arama uzayının daha etkin taranabilmesi için logaritmik sarmal stratejisi ve yerel en iyi çözümlere erken yakınsamayı önlemek amacıyla geliştirilmiş arama kalitesi (ESQ) tekniği birlikte kullanılmıştır. Standart karşılaştırma fonksiyonları ve ikili özellik seçimi problemleri üzerinde gerçekleştirilen deneysel ve istatistiksel analizler, CLESQ-ZOA'nın hem çözüm doğruluğu hem de kararlılık açısından klasik ZOA ve diğer sezgisel algoritmalara kıyasla üstün performans sergilediğini göstermiştir. Ayrıca Hepatosellüler Karsinom veri seti üzerinde yapılan uygulama sonuçları, önerilen yöntemin gerçek dünya özellik seçimi problemlerinde etkili, güvenilir ve uygulanabilir bir çözüm sunduğunu ortaya koymuştur (Ozbay, 2025).

2.2.9. Çoklu Strateji Füzyonu ile Geliştirilmiş Zebra Optimizasyon Algoritması ve Robot Yol Planlamasında Uygulaması

Bu çalışmada (Z. Wang vd., 2025), ZOA erken yakınsama ve yerel en iyi çözümlere takılma sorunlarını gidermek amacıyla Çok Stratejili Geliştirilmiş ZOA (MZOA) önerilmiştir. Önerilen yöntemde, popülasyon çeşitliliğini artırmak için mercek görüntüleme tabanlı tersine öğrenme, keşif ve sömürü dengesini sağlamak amacıyla üçgen yürüme operatörleri ve çözüm uzayının daha etkin taranması için Levy uçuşu mekanizmaları birlikte kullanılmıştır. CEC2005 ve CEC2017 karşılaştırma fonksiyonları üzerinde yapılan deneysel analizler, MZOA'nın temel ZOA'ya kıyasla yaklaşık %15,8 oranında performans artışı sağladığını ve daha kararlı çözümler ürettiğini göstermiştir. Ayrıca robot yol planlama uygulamalarında MZOA, hareket yolunu %8,7 oranında kısaltarak hem basit hem de engelli ortamlarda pürüzsüz ve geometrik olarak uygun rotalar oluşturabildiğini ortaya koymuştur. Elde edilen bulgular, çoklu strateji füzyonunun ZOA'nın küresel arama yeteneğini güçlendirdiğini ve karmaşık mühendislik problemlerinde güvenilir bir çözüm sunduğunu doğrulamaktadır (Z. Wang vd., 2025).

2.2.10. Zebra Optimizasyon Algoritması ve Kayma Modu Teorisini Entegre Eden Fırçasız DC Motor Sürücü Sistemi Kontrol Cihazının Tasarımı

Bu çalışmada (Chao vd., 2025) tarafından gerçekleştirilen çalışmada, fırçasız doğru akım motorlarının (BLDCM) hız izleme performansını artırmak amacıyla ZOA ile Üstel Erişim Yasası Tabanlı Kayan Mod Denetleyici (ERLSMC) bütünleştirilmiştir. Önerilen yöntemde ZOA, kayan mod denetleyicinin üç temel kazanç parametresini çevrim içi olarak optimize ederek geçici rejim davranışını ve dinamik yanıtı iyileştirmektedir. Benzetim ve deneysel sonuçlar, ZOA ile ayarlanan denetleyicinin farklı yük ve hız koşulları altında aşımı önemli ölçüde azalttığını, daha hızlı yerleşme süresi sağladığını ve yüksek hassasiyetli hız takibi gerçekleştirdiğini ortaya koymuştur. Elde edilen bulgular, ZOA tabanlı denetim yaklaşımının BLDC motor sürücü sistemlerinde kararlı ve güvenilir bir hız denetimi sunduğunu doğrulamaktadır (Chao vd., 2025).

2.2.11. Modern Enerji Sistemi Planlama Probleminin Çözümü için Tasarım Operatörleri Aracılığıyla Modifiye Edilmiş Zebra Optimizasyon Algoritması

Bu çalışmada **(Balci & Duman, 2025)** tarafından gerçekleştirilen çalışmada, modern güç sistemlerinde İletim Genişleme Planlaması (TEP) problemlerinin çözümü için Bulanık Uygunluk–Mesafe Dengesi ve Karşıt Tabanlı Öğrenme yaklaşımlarıyla güçlendirilmiş FFDBZOA-OBL algoritması önerilmiştir. Önerilen yöntemde, uygunluk ve mesafe ölçütleri arasındaki denge bulanık mantık yardımıyla uyarlamalı biçimde sağlanırken, karşıt tabanlı öğrenme stratejileriyle başlangıç popülasyonunun çeşitliliği artırılmış ve erken yakınsama

sorunu azaltılmıştır. CEC2020 ve CEC2022 karşılaştırma fonksiyonları ile gerçek dünya TEP senaryoları üzerinde yapılan deneysel ve istatistiksel analizler, FFDBZOA-OBL algoritmasının temel ZOA ve diğer sezgisel yöntemlere kıyasla daha hızlı yakınsama, yüksek çözüm kararlılığı ve daha düşük yatırım maliyetleri sağladığını göstermiştir. Elde edilen bulgular, önerilen yaklaşımın yüksek boyutlu ve karmaşık güç sistemi planlama problemlerinde ölçeklenebilir, sağlam ve etkili bir optimizasyon yöntemi olduğunu ortaya koymaktadır **(Balci & Duman, 2025)**.

2.2.12. Bezier eğrileri ve hibrit zebra optimizasyon algoritmasına dayalı üç boyutlu karmaşık uzayda yol planlama yöntemi

Bu çalışmada (Huang vd., 2022) tarafından gerçekleştirilen bu çalışmada, insansız hava araçlarının (İHA) engellerle dolu üç boyutlu ortamlarda güvenli ve verimli biçimde yol planlaması yapabilmesi amacıyla Bézier eğrileri ile Hibrit ZOA (HZOA) temelli bir yaklaşım önerilmiştir. Yöntemde, uçuş yolu doğrudan parametrik eğrilerle temsil edilerek kontrol noktaları karar değişkeni hâline getirilmiş; toplam uçuş mesafesi, yol pürüzsüzlüğü ve uçuş irtifası çok amaçlı bir maliyet fonksiyonu altında optimize edilmiştir. Temel ZOA'nın erken yakınsama eğilimini azaltmak amacıyla Gri Kurt Optimizasyon Algoritması'nın sosyal hiyerarşik yapısı ZOA'ya entegre edilmiş ve üç öncü çözüm üzerinden yönlendirilen hibrit bir arama stratejisi geliştirilmiştir. CEC2019 test fonksiyonları ve farklı karmaşıklık seviyelerindeki üç boyutlu simülasyon ortamlarında elde edilen sonuçlar, önerilen yöntemin uçuş maliyeti, irtifa kontrolü ve yol pürüzsüzlüğü açısından ZOA, GWO, DE, NGO ve SMA gibi yöntemlere kıyasla istatistiksel olarak anlamlı bir üstünlük sağladığını göstermiştir. Çalışma, HZOA ve Bézier eğrisi bütünleşmesinin, karmaşık görev sahalarında İHA'lar için geometrik olarak uygulanabilir, pürüzsüz ve güvenilir uçuş rotaları üretmede etkili bir çözüm sunduğunu ortaya koymaktadır (Huang vd., 2022).

2.2.13. Güvenilirlik yedekliliği tahsisi ve mühendislik optimizasyon problemleri için geliştirilmiş zebra optimizasyon algoritması

Bu çalışmada (Punia vd., 2025), güvenilirlik–yedeklilik atama ve mühendislik tasarım problemlerinde erken yakınsama ve yerel en iyide takılma sorunlarını aşmak amacıyla Geliştirilmiş ZOA (EZOA) önerilmiştir. Önerilen yaklaşımda, arama uzayının başlangıç aşamalarında daha etkin taranabilmesi için Levy uçuşu stratejisi, popülasyon çeşitliliğinin artırılması ve çözüm kalitesinin yükseltilmesi için ise Mercek Karşıt Tabanlı Öğrenme (LOBL) mekanizması ZOA yapısına entegre edilmiştir. EZOA; 23 klasik test fonksiyonu, CEC-2019 benchmark seti, dört karma tamsayılı güvenilirlik problemi ve dört kısıtlı mühendislik tasarım

problemi üzerinde değerlendirilmiş; elde edilen sonuçlar, önerilen yöntemin yakınsama hızı, çözüm kararlılığı ve istatistiksel başarı açısından temel ZOA ve diğer meta-sezgisel algoritmalara kıyasla anlamlı bir üstünlük sağladığını ortaya koymuştur. Ayrıca Friedman sıralama testi ile Wilcoxon ve Mann–Whitney istatistiksel analizleri, EZOA'nın özellikle yüksek boyutlu ve kısıtlı problemlerde sağlam ve güvenilir bir optimizasyon aracı olduğunu doğrulamıştır (Punia vd., 2025).

2.2.14. Debutanizer kolonunun MLP yumuşak sensör modeli için kaos yakınsama faktörlü Zebra optimizasyon algoritması ve Gauss mutasyonu

Bu çalışmada (Shang-Guan vd., 2025), endüstriyel süreç izleme problemleri kapsamında debutanizer sütununun yumuşak algılayıcı (soft sensör) modeli için kaotik yakınsama faktörü ve Gauss mutasyonu ile güçlendirilmiş bir ZOA (c5SinZOA) önermiştir. Önerilen yöntemde, zebraların beslenme aşamasına entegre edilen kaotik yakınsama faktörü sayesinde küresel arama kabiliyeti artırılırken, her yineleme sonunda uygulanan Gauss mutasyonu ile popülasyon çeşitliliği korunmuştur. CEC2022 test fonksiyonları ve gerçek endüstriyel debutanizer verileri üzerinde gerçekleştirilen deneyler, c5SinZOA'nın ortalama hata (MSE), kök ortalama kare hata (RMSE) ve mutlak hata (MAE) ölçütlerinde rakip algoritmalara kıyasla daha düşük değerler ürettiğini göstermiştir. Elde edilen bulgular, önerilen yaklaşımın hem tahmin doğruluğu hem de model uyumu açısından endüstriyel süreç izleme uygulamalarında etkili ve güvenilir bir çözüm sunduğunu ortaya koymuştur (Shang-Guan vd., 2025).

2.2.15. Yüksek boyutlu uzaylarda verimli özellik seçimi için geliştirilmiş Zebra optimizasyon stratejileri

Bu çalışmada (Stephan vd., 2025), yüksek boyutlu veri kümelerinde özellik seçimi problemini çözmek amacıyla ZOA beş farklı varyantı (O-ZO, AEE-ZO, RNM-ZO, HIC-ZO ve PF-ZO) kapsamlı biçimde incelenmiştir. Çalışmada, biyoloji, sağlık ve metin alanlarına ait 18 farklı veri seti kullanılarak önerilen varyantlar; sınıflandırma doğruluğu, özellik azaltma oranı, uygunluk değeri ve hesaplama süresi açısından Genetik Algoritma, Parçacık Sürü Optimizasyonu ve Gri Kurt Optimizasyonu ile karşılaştırılmıştır. Deneysel bulgular, özellikle HIC-ZO ve AEE-ZO varyantlarının keşif ve sömürü dengesini daha etkin kurarak hem sınıflandırma başarımını artırdığını hem de daha az sayıda özellikle kararlı çözümler ürettiğini göstermiştir. İstatistiksel analizler, gelişmiş ZOA varyantlarının geleneksel yöntemlere kıyasla anlamlı bir üstünlük sağladığını doğrulamakta olup, bu algoritmaların yüksek boyutlu veri

kümelerinde güvenilir ve uygulanabilir bir özellik seçimi yaklaşımı sunduğunu ortaya koymaktadır (Stephan vd., 2025).

2.2.16. Kablosuz sensör ağlarında kümeleme ve yönlendirme için enerjiye duyarlı çok amaçlı üçlü strateji tabanlı zebra optimizasyon algoritması

Bu çalışmada (Devprabhakar vd., 2025), Kablosuz Sensör Ağlarında (KSA) enerji verimliliğini ve ağ ömrünü artırmak amacıyla EMTSZOA adı verilen çok amaçlı, üçlü strateji tabanlı ZOA önerilmiştir. Önerilen yöntem; Sinüs–Kosinüs Stratejisi, Dinamik Anahtarlama Olasılığı ve Dinamik Uyarlanabilir Ağırlık mekanizmalarını bir araya getirerek küme başı seçimi ve yönlendirme sürecini optimize etmektedir. Düğüm–küme başı mesafesi, küme başı–baz istasyonu mesafesi, konum faktörü, yük dengeleme faktörü ve ortalama düğüm enerjisi gibi ölçütler birlikte değerlendirilmiştir. Deneysel sonuçlar, EMTSZOA'nın enerji tüketimini azalttığını, paket iletim oranını ve ağ yaşam süresini artırdığını; özellikle 500 düğümlü ağ yapısında %93,80 yaşam beklentisi ile mevcut yöntemlere kıyasla belirgin bir üstünlük sağladığını göstermiştir (Devprabhakar vd., 2025).

2.2.17. Özellik Seçimi için Zebra Optimizasyon Algoritması

Bu çalışmada (Faizan & Mohamad, 2026) tarafından gerçekleştirilen bu çalışmada, makine öğrenmesi modellerinde sınıflandırma doğruluğunu artırmak ve hesaplama maliyetlerini azaltmak amacıyla ZOA, özellik seçimi sürecinde KNN sınıflandırıcısı ile birlikte kullanılmıştır. UCI Makine Öğrenmesi Veri Havuzu'ndan alınan farklı kıyaslama veri kümeleri üzerinde yapılan deneyler, ZOA'nın gereksiz ve tekrarlı özellikleri etkili biçimde eleyerek %91'e varan özellik azaltma oranına ulaşabildiğini göstermiştir. Deneysel sonuçlar, seçilen daha küçük özellik alt kümeleriyle sınıflandırma doğruluğunun korunduğunu hatta artırıldığını ortaya koymuş; böylece ZOA'nın yüksek boyutlu ve karmaşık veri yapılarında hesaplama yükünü düşüren, kararlı ve güvenilir bir özellik seçimi yöntemi sunduğu doğrulanmıştır (Faizan & Mohamad, 2026).

3. MATERYAL VE YÖNTEM

3.1 Zebra Optimizasyon Algoritmasının Matematiksel Modeli

Bu bölüm, ZOA modellemesini oluşturmak amacıyla, doğal zebra davranışlarının matematiksel simülasyonlarını sunmaktadır**.**

3.1.1. İlham Kaynağı

Zebralar, atgiller familyasına ait olup Doğu ve Güney Afrika kökenlidir. Bu hayvanlar, siyah-beyaz çizgili postlarıyla tanınmaktadır. Bu çizgiler genellikle boyun ve gövde üzerinde dikey olarak düzenlenmiş olup, yalnızca yırtıcılardan gizlenmek için değil, aynı zamanda ısırıcı böceklere karşı caydırıcı bir işlev görmektedir. Zebraların temel özellikleri ve boyutları arasında 210 ila 300 cm arasında değişen bir vücut uzunluğu, 38 ila 75 cm arasında değişen bir kuyruk uzunluğu, 110 ila 160 cm arasında değişen omuz yüksekliği ve 175 ila 450 kg arasında değişen bir ağırlık yer alır. Zebralar, gerektiğinde yüksek hızlarda koşabilmelerini sağlayan uzun ve ince bacaklara sahip dayanıklı hayvanlardır. Diğer yabani atgiller gibi, zebraların her ayağında tek bir toynak, uzun bir boyun ve otları kolayca otlamak için uyarlanmış bir kafa yapısı bulunmaktadır (Trojovska vd., 2022). Doğal yaşam alanlarında zebralar, iki önemli sosyal davranış sergiler: avlanma ve yırtıcılara karşı savunma stratejileri. Avlanma sürecinde, bir lider zebra, diğerlerinin otlayabilecekleri bölgelere ilerlemesini sağlamak için öncülük eder. Bu süreçte, sürünün geri kalanı bu lider zebrayı düzlüklerde takip eder. Zebralar, yırtıcılardan kaçarken zikzak bir kaçış deseni uygular. Ancak bazen, yırtıcıyı şaşırtmak veya korkutmak amacıyla bir araya gelebilirler. Zebraların bu zeki davranışlarının matematiksel modellemesi, önerilen ZOA tasarımına ilham kaynağı oluşturmaktadır (Trojovska vd., 2022).

3.1.2. Popülasyonun Başlatılması

ZOA, nüfus tabanlı bir meta-sezgisel optimizasyon çerçevesi olarak kavramsallaştırılmış olup, nüfus birimlerinin zebra olarak metaforik bir temsili, algoritmanın merkezinde yer almaktadır. Matematiksel olarak, her zebra, optimizasyon sürecinde bir potansiyel çözüm vektörü olarak tanımlanmaktadır. Bu zebraların yaşadığı habitat ya da düzlük, ele alınan problemin çok boyutlu arama uzayındaki belirli bir lokasyonu temsil etmektedir. Arama uzayındaki her bir zebranın pozisyonu, ilgili karar değişkenlerine atanan değerleri doğrudan belirlemektedir. Bu bağlamda, ZOA paradigmasında her zebra, değişken değerlerini

bileşenleri olarak kapsayan bir aday çözümün vektörel bir temsili niteliğindedir. Zebraların bütünü, yani nüfus, matematiksel olarak bir matris formunda yapılandırılmış olup, birden fazla aday çözümün eşzamanlı olarak temsil edilmesini ve manipüle edilmesini sağlamaktadır. ZOA nüfus matrisindeki zebra pozisyonlarının başlangıç dağılımı, rastgele süreçler aracılığıyla belirlenmekte ve böylece ZOA nüfus matrisinin ilk yapılandırması ortaya çıkmaktadır (Trojovska vd., 2022). Bunun yanı sıra, her bir zebra ya da optimizasyon ajanı için karar verme parametreleri, arama alanındaki uzaysal koordinatlarına doğrudan bağlıdır. Dolayısıyla, ZOA çerçevesinde her zebra yalnızca metaforik bir varlık değil, aynı zamanda çözüm hipotezlerinin vektörel bir temsilcisidir. Bu vektörün her bir bileşeni, optimizasyon problemiyle ilgili bir karar değişkenini yansıtmaktadır. Vektör, aslında, ZOA'nın çözüm uzayında gezinmesini sağlayan bir araçtır. Zebra topluluğunun ya da nüfus matrisinin mimari temelleri, bir matris temsili aracılığıyla oluşturulmakta ve birden fazla çözüm vektörünün paralel olarak simülasyonu ve değerlendirilmesi için yapılandırılmış bir yaklaşım sunmaktadır. Zebraların arama alanındaki başlangıç dağılımı, rastgele mekanizmalarla yönetilmekte olup, ZOA'nın keşif dinamiklerine zemin hazırlamaktadır. Denklem (1), ZOA nüfus matrisinin başlangıç parametrelerini ve yapısını tanımlayarak optimizasyon sürecinin başlangıcını işaret etmektedir (Ghadi vd., 2023).

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N\times m}, \qquad (1)$$

Burada X zebra popülasyonunu temsil etmektedir. X_i, i. zebrayı ; $x_{i,j}$ i. zebranın j. problem değişkeni için önerdiği değeri ifade etmektedir. N, popülasyon büyüklüğünü (zebra sayısını) m ise karar değişkenlerinin sayısını temsil etmektedir. Her zebra, optimizasyon problemine yönelik bir aday çözümü ifade etmektedir. Dolayısıyla, her bir zebranın problem değişkenleri için önerdiği değerlere bağlı olarak, amaç fonksiyonu değerlendirilebilir. Amaç fonksiyonu için elde edilen değerler, Denklem (2)'de belirtildiği üzere bir vektör olarak ifade edilmektedir (Trojovska vd., 2022).

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N\times 1}, \qquad (2)$$

Vektör F, amaç fonksiyonuna ait değerleri temsil ederken, F_i, i. zebra için belirlenen spesifik amaç fonksiyonu değerini ifade etmektedir. Amaç fonksiyonu değerlerinin etkili bir şekilde karşılaştırılması, ilgili aday çözümlerin kalitesinin kapsamlı bir analizini kolaylaştırmakta ve böylece ele alınan problem için en uygun çözümün belirlenmesini sağlamaktadır. Minimizasyon problemleri bağlamında, amaç fonksiyonunun en düşük değerine sahip olan zebra, üstün aday çözüm olarak kabul edilmektedir. Buna karşılık, maksimizasyon problemleri kapsamında, amaç fonksiyonunun en yüksek değerine sahip olan zebra, üstün aday çözüm olarak değerlendirilir. Zebraların pozisyonları ve dolayısıyla amaç fonksiyonu değerleri her iterasyonda güncellendiğinden, her iterasyonda en iyi aday çözümün yeniden değerlendirilmesi ve belirlenmesi zorunludur (Trojovska vd., 2022).

Vahşi zebraların doğal davranışları, örneğin otlama ve yırtıcılara karşı savunma, ZOA üyelerinin güncellenmesi için bir temel olarak kullanılmıştır. Bu çerçevede, her iterasyonda ZOA popülasyonunun üyeleri iki ayrı aşamada güncellenmektedir: Otlama ve Yırtıcılara karşı savunma stratejileri.

3.1.3. Aşama 1: Beslenme Davranışı

İlk aşamada, popülasyon üyeleri, zebralarda gözlemlenen otlama davranışlarının simülasyonlarına dayanarak güncellenir. Zebralar, temel besin kaynakları olarak öncelikli olarak ot ve sazlıklarla beslenirler. Ancak, favori yiyeceklerinin kıt olduğu durumlarda, tomurcuk, meyve, kabuk, kök ve yaprakları da tüketebilirler. Zebralar, bitki örtüsünün kalitesine ve mevcudiyetine bağlı olarak zamanlarının % 60-80'ini otlayarak geçirebilirler. Zebralar arasında, üstteki daha az besleyici otları tüketerek, alt kısımda bulunan kısa ve daha besleyici otlar için diğer türlere elverişli koşullar sağlayan türler, düz zebralar olarak adlandırılır. ZOA'da, popülasyonun en iyi üyesi lider zebra olarak kabul edilir ve otlama aşamasında diğer popülasyon üyelerini arama uzayındaki konumuna doğru yönlendirir. Bu nedenle, zebraların otlama aşamasında konumlarının güncellenmesinin matematiksel modellemesi, Denklem (3) ve Denklem (4) ile ifade edilebilir (Trojovska vd., 2022).

$$x_{i,j}^{\text{new},P1} = x_{i,j} + r \cdot \left(PZ_j - I \cdot x_{i,j}\right) \qquad (3)$$

$$X_i = \begin{cases} X_i^{\text{new},P1}, & F_i^{\text{new},P1} < F_i \\ X_i, & \text{else}, \end{cases} \qquad (4)$$

Burada $x_i^{\text{new},P1}$ birinci aşamada i. zebranın yeni durumunu, $x_{i,j}^{\text{new},P1}$ ise j. boyuttaki değerini, $F_i^{\text{new},P1}$, amaç fonksiyonu değerini ifade etmektedir. PZ , lider zebrayı PZ_j lider zebranın j. boyutunu temsil eder. r değeri [0, 1] aralığında rastgele bir sayı olup, $I = round\ (1 + rand)$, [0, 1] aralığında rastgele sayı olan $rand$'e bağlıdır. Dolayısıyla $I \in 1{,}2$, olur ve eğer I parametresi 2'ye eşit ise, bu aşamada popülasyon hareketinde çok daha fazla çeşitlilik gözlemlenir (Trojovska vd., 2022).

3.1.4. Aşama 2: Yırtıcılara Karşı Savunma Stratejileri

İkinci aşamada, zebraların yırtıcı saldırılarına karşı uyguladıkları savunma stratejilerinin simülasyonu, ZOA popülasyonu üyelerinin arama uzayındaki konumlarını güncellemek için kullanılır. Aslanlar, zebralar açısından temel yırtıcılar arasında yer alırken, çitalar, leoparlar, yabani köpekler ve kahverengi sırtlanlar gibi diğer tehlikeli fauna üyeleri de potansiyel tehdit oluşturabilir. Zebraların bu tehditlere karşı kullandıkları savunma stratejileri, karşılaşılan yırtıcıya bağlı olarak çeşitli şekillerde ortaya çıkar (Trojovska vd., 2022b). Zebraların aslan saldırıları karşısında sergilediği savunma mekanizması, zikzak bir kaçış deseni uygulamak ve rastgele yanal dönüşler yapmak gibi kaçınma manevralarını içerir. Daha küçük yırtıcılar, örneğin sırtlanlar ve köpekler karşısında zebralar daha iddialı bir davranış eğilimi gösterirler. Bu durumlarda, saldırganı şaşırtmak ve korkutmak amacıyla etkili bir savunma stratejisi oluşturmak için bir araya gelirler (Trojovska vd., 2022). ZOA tasarımının kavramsal çerçevesinde, iki senaryonun eşit olasılıkla gerçekleştiği varsayılmaktadır:

- Bir zebra, bir aslan saldırısına maruz kalır ve bu durum zebranın kaçış stratejisi benimsemesine neden olur.
- Diğer yırtıcılar bir zebrayı hedef alır ve bu durum zebranın agresif bir strateji tercih etmesini sağlar.

İlk senaryoda, zebralar aslan saldırılarına maruz kaldığında, mevcut durumlarındaki saldırı bölgesinden hızla uzaklaşırlar. Bu strateji, matematiksel olarak Denklem (5) kullanılarak modellenebilir. İkinci senaryoda ise, diğer yırtıcılar bir zebrayı hedef aldığında, sürüdeki diğer zebralar saldırıya uğrayan zebranın etrafında toplanır, savunma pozisyonu alır ve saldırganı şaşırtmaya çalışır. Bu stratejinin matematiksel formülasyonu da yine Denklem (5) ile ifade edilmektedir.

Zebraların konumlarının güncellenmesi sürecinde, bir zebranın yeni bir pozisyonda amaç fonksiyonu için daha iyi bir değer elde etmesi durumunda, bu yeni pozisyon kabul edilir. Bu güncelleme koşulu, Denklem (6) ile resmileştirilmiştir (Trojovska vd., 2022).

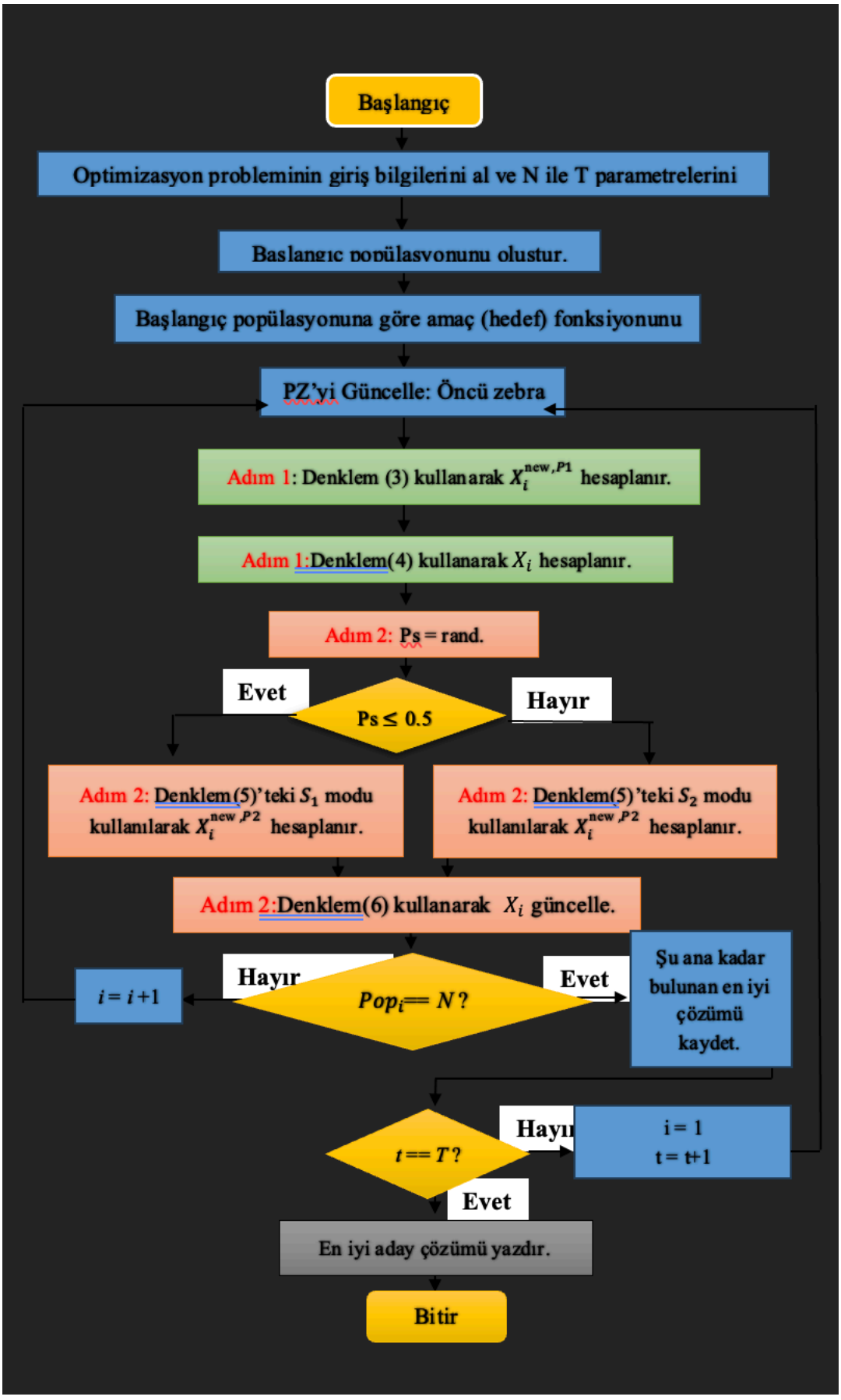
$$x_{i,j}^{\text{new},P2} = \begin{cases} S_1: x_{i,j} + R \cdot (2r-1) \\ \quad \cdot \left(1 - \dfrac{t}{T}\right) \cdot x_{i,j}, & P_s \leq 0.5 \\ S_2: x_{i,j} + r \cdot \left(AZ_j - I \cdot x_{i,j}\right), & \text{else} \end{cases} \tag{5}$$

$$X_i = \begin{cases} X_i^{\text{new},P2}, & F_i^{\text{new},P2} < F_i \\ X_i, & \text{else}, \end{cases} \tag{6}$$

Burada, $x_{i,j}^{\text{new},P2}$ ikinci aşamada i. zebranın j. boyuttaki değerini, $F_i^{\text{new},P2}$ amaç fonksiyonu değerini ifade etmektedir. t değişkeni iterasyon indeksini T değişkeni maksimum iterasyon sayısını, R ise 0.01'e eşit bir sabiti sayıyı, P_s ise [0,1] aralığında rastgele iki stratejiden birinin seçilme olasılığını, AZ ise saldırıya uğrayan zebranın durumunu ve AZ_j değişkeni bu zebranın j. boyuttaki değerini temsil etmektedir (Trojovska vd., 2022).

3.2 ZOA'nın Sözde Kodu ve Akış Şeması

Şekil 3.1, ZOA'nın akış şemasını, Şekil 3.2 ise ZOA'nın sözde kodunu göstermektedir (Trojovska vd., 2022).



Şekil 3.1. ZOA'nın Akış Şeması (Trojovska vd., 2022)

ZOA Başla:
Giriş: Optimizasyon problemine ait bilgiler.
İterasyon sayısını (T) ve zebra popülasyonunun sayısını (N) belirle.
Zebraların başlangıç konumlarını başlat ve uygunluk (fitness) fonksiyonlarını değerlendir..
For t = 1: *T*
  Lider zebrayı güncelle(*ZL*).
  For i = 1: *N*
    Adım I: Beslenme(Yiyecek arama) faaliyeti
    i^{th} zebranın yeni durumunu Denklem(3) kullanarak hesapla.
    i^{th} zebrayı Denklem(4) kullanarak güncelle.
    Adım II: Avcılara karşı savunma teknikleri.
    Ps = rand
    If Ps <0.5 Then
      Sömürü (Aslana karşı savunma tekniği)
      i^{th} zebranın yeni durumunu Denklem (5) kullanarak hesapla.
    Else
    Keşif (Diğer avcılara karşı savunma teknikleri)
      i^{th} zebranın yeni durumunu Denklem(5) kullanarak hesapla.
    End If
    i^{th} zebrayı Denklem(6) kullanarak güncelle.
  End For
  Şu ana kadar bulunan en iyi aday çözümü kaydet.
End For
Çıktı: Verilen optimizasyon problemi için ZOA tarafından elde edilen en iyi çözüm.
End ZOA.

Şekil 3.2. ZOA sözde kodu (Trojovska vd., 2022b)

3.3 Önerilen Optimizasyon Algoritması

Bu bölümde, ZOA'nın ikili bir versiyonu olan İkili (Binary) Zebra Optimizasyon Algoritması (BinZOA) tasarlanmıştır. ZOA' da elde edilen sürekli değerler transfer fonksiyonları (S-Şekilli) kullanılarak ikili forma eşlenmiştir.

3.4 İkili (Binary) Zebra Optimizasyon Algoritması (BinZOA)

Bu bölümde, Zebra Optimizasyon Algoritmasının ikili bir versiyonu olan İkili Zebra Optimizasyon Algoritması (BinZOA), özellik seçimini problemini ele almak üzere tasarlanmıştır. ZOA sürekli optimizasyon zorluklarında iyi olsa da, başlangıçta özellik seçimi problemi gibi ayrık problemleri çözme yeteneğinden yoksundur. Sonuç olarak, bu çalışmada ZOA özellik şeçimi problemlerini çözecek şekilde yeniden yapılandırılmıştır. ZOA'yı ikili bir versiyona dönüştürmek için transfer fonksiyonları kullanılmıştır. Bu fonksiyonlar standart algoritmada sürekli olarak üretilen değerlerin 0 veya 1 değerlerinden oluşan ikili değerlerine dönüştürülmesini sağlarlar. Bu bağlamda 0 değeri bir data setinde herhangi bir özelliğinin seçilmediğini 1 değeri ise özelliğin seçildiğini gösterir. Sürekli arama uzayının ikili arama

uzayına dönüştürmek için literatürde birçok transfer fonksyionu kullanılmaktadır (Örneğin S-şekilli, V-şekilli vb. (Ervural & Hakli, 2023). Bunlardan en yaygın kullanılanları 'S' formundaki transfer fonksiyonlarıdır. Bu çalışmada tek tip S-şekilli transfer fonksiyonu kullanılmıştır. Bu transfer fonksiyonu Denklem (7)'de detaylı şekilde gösterilmiştir. Bu fonksiyon kullanılarak sürekli değerler ikili forma dönüştürülmüştür. Denklem (8)'de sürekli bir arama uzayındaki aday çözümlerin ikili arama uzayına haritalanması süreci gösterilmiştir.

$$transfer(x) = \frac{1}{1+e^{-2*x_{i,j}}} \qquad (7)$$

$$X_{bin(i,:)} = transfer_{func}(X(i,:), i, m, transfer_{number}) \qquad (8)$$

Yukarıdaki denklemlere göre, $x_{bin(i,:)}$ X matrisinin *i.* satırının ikili(binary) versiyonunu temsil etmektedir. $transfer_{func}(X(i,:), i, m, transfer_{number})$ fonksiyon çağrısı, vektörün *i.* indeksindeki değeri, boyut (m) ve belirtilen transfer fonksiyonunu ($transfer_{func}$) kullanarak dönüştürme sürecine işaret etmektedir. $X_{bin(i,:)}$ matrisi, aşağıdaki denklemde ifade edildiği şekilde gösterilebilir.

$$X_{bin(i,:)} = \begin{bmatrix} X\_Bin_1 \\ \vdots \\ X\_Bin_i \\ \vdots \\ X_{\_binN} \end{bmatrix}_{N \times m} \qquad (9)$$

Çalışmada, popülasyonun tüm elemanları Denklem (8)'de gösterildiği şekilde ayrık (discrete) ifadelere dönüştürülmüştür ve bu dönüşüm tüm popülasyona uygulanmıştır. Şekil 3.3, BinZOA'nın sözde kodunu göstermektedir.

Start BinZOA:
Input: Optimizasyon problemine ait bilgiler alınır.
Maksimum iterasyon sayısı (T) ve zebra popülasyon büyüklüğü (N) belirlenir.
Zebraların başlangıç konumları rastgele oluşturulur ve uygunluk (fitness) değerleri hesaplanır..
Zebraların konumları, Denklem (7)'de verilen transfer fonksiyonu kullanılarak ikili (binary) forma dönüştürülür.
For t = 1: *T*
Zebra liderini (ZL) güncelleyin.
Denklem (7) kullanılarak transfer fonksiyonu ile ZL'yi ikili forma dönüştürün.
For *i* = 1: *N*
Adım 1: Beslenme (Yiyecek Arama) faaliyeti
i^{th} zebranın yeni durumunu Denklem (3) kullanarak hesapla.
Denklem (7) kullanarak transfer fonksiyonu ile i^{th} zebrayı güncelle.
i^{th} zebrayı güncelle (Denklem(4) kullanarak.
Adım II: Avcılara karşı savunma taktikleri.
Ps= rand
If *Ps* <0.5 Then
Sömürü (Aslana karşı savunma)
Denklem(5) kullanarak i^{th} zebranın konumunu güncelle.
Else
Keşif (Diğer yırtıcılara karşı savunma teknikleri)
Denklem (5) kullanarak i^{th} zebranın yeni durumunu hesaplayın.
Denklem (7) kullanılarak transfer fonksiyonu ile ith zebrayı ikili forma dönüştürün.
End If
i^{th} zebrayı Denklem (6) kullanarak güncelleyin.
End For
Şu ana kadar en iyi aday çözümü kaydedin.
End For
Çıktı: Verilen optimizasyon problemi için BinZOA tarafından elde edilen en iyi çözüm.
End BinZOA.

Şekil 3.3 BinZOA sözde kodu (Trojovska vd., 2022)

3.5 Özellik Seçimi Problemi İçin Uygunluk (Fitness) Fonksiyon Tanımı

Bu bölümde, İkili ZOA (BinZOA) CSE-CIC-IDS2018 veri seti üzerinde özellik seçimi problemi için gösterdiği performans değerlendirilmiştir. BinZOA algoritması, yüksek boyutlu ağ trafiği verileri içerisinden en ayırt edici özellik alt kümelerinin belirlenmesi amacıyla kullanılmıştır. Elde edilen özellik alt kümeleri, ilk aşamada KNN, RF, DT, MLP gibi makine öğrenmesi tabanlı sınıflandırıcılar ile değerlendirilmiş ve en yüksek başarımı sağlayan özellik kombinasyonları belirlenmiştir. İkinci aşamada ise, BinZOA tarafından seçilen en iyi özellik alt kümeleri kullanılarak LSTM, RNN, CNN gibi derin öğrenme modelleri eğitilmiştir. Bu aşama, saldırı tespit performansının daha ileri bir seviyeye taşınmasını ve zamansal bağımlılıkların daha etkin şekilde modellenmesini amaçlamaktadır. Özellikle LSTM ve RNN modelleri, ağ trafiği verilerindeki zamansal ilişkileri yakalayarak saldırı türlerinin daha doğru biçimde sınıflandırılmasını sağlamıştır. CNN modeli ise veri içerisindeki yerel örüntüleri öğrenerek sınıflandırma başarımını artırmıştır. Deneylerde kullanılan CSE-CIC-IDS2018 veri seti, farklı siber saldırı türlerini içeren kapsamlı ve dengeli bir veri setidir. Veri setinin %80'i

eğitim, %20'si test amacıyla rastgele ayrılmıştır. Elde edilen sonuçlar, BinZOA tabanlı özellik seçimi ile derin öğrenme modellerinin birlikte kullanılmasının, siber saldırı tespitinde hem doğruluk oranlarını artırdığını hem de hesaplama maliyetlerini düşürdüğünü ortaya koymaktadır.

Özellik seçimi problemi, sınıflandırma doğruluğu hatası ile seçilen özellik sayısını dengeleyen bir minimizasyon fonksiyonu ile çözülmüştür. Uygunluk fonksiyonu Denklem (10)'da tanımlanmıştır.

$$Fitness_i = minimize\left(\alpha \times Erorr_i + \beta \times \left(\frac{d_i}{D}\right)\right) \quad (10)$$

Sınıflandırma hatası (1-doğruluk) $Erorr_i$ olarak ifade edilmiştir ve hata değeri α katsayısı (α=0,7) ile ağırlıklandırılmıştır. Toplam özellik sayısı *D* değişkeni ile, seçilen özellik alt kümesi ise *d* değişkeni ile gösterilmiştir. Seçilen özellik oranı, 1-α olarak hesaplanan β katsayısı (β=0,3) ile ağırlıklandırılmıştır. Bu uygunluk fonksiyonu, en az hata ile en az sayıda özelliğin seçilmesini hedeflemektedir. α ve β katsayısı değerleri, bu iki faktör arasında denge sağlamak için dikkatle seçilmiştir.

Bu çalışmada, KNN, RF, DT, MLP sınıflandırıcıları kullanılarak gerçekleştirilen deneyler aracılığıyla ZOA'nın özellik seçimi sürecindeki etkinliği değerlendirilmiştir. Daha sonra en iyi özellikler belirlenerek derin öğrenme modelleri kullanılmıştır. KNN algoritmasının performansı, komşu sayısı k = 3 olarak belirlenerek ve Manhattan mesafe metriği kullanılarak analiz edilmiştir. Mesafeye dayalı ağırlıklandırma yaklaşımı sayesinde sınıflandırma doğruluğunun arttığı gözlemlenmiştir. RF algoritması, birden fazla karar ağacından oluşan topluluk yapısı ile modellenmiş; her bir ağacın sınıf dengesizliğine karşı ağırlıklandırıldığı ve minimum örnek sayısı kısıtlarıyla ağaç derinliğinin sınırlandırıldığı görülmüştür. DT algoritması, bilgi kazancı ölçütüne dayalı olarak karar vermekte ve her düğümde toplam özellik sayısının karekökü kadar özelliği dikkate alan sqrt yöntemi kullanmaktadır. MLP modeli ise iki gizli katmandan oluşan bir sinir ağı yapısına sahip olup ReLU aktivasyon fonksiyonu ile çalışmaktadır. Eğitim sürecinde Adam optimizasyon algoritması, mini-batch öğrenme stratejisi ve aşırı öğrenmeyi önlemek amacıyla erken durdurma (early stopping) mekanizması kullanılmış; doğrulama için verinin % 10'u ayrılmıştır. ZOA ile belirlenen en ayırt edici özellik alt kümeleri kullanılarak elde edilen sonuçlar, yalnızca klasik makine öğrenmesi

algoritmalarıyla sınırlı kalmamış; ikinci aşamada derin öğrenme tabanlı modeller ile saldırı tespit performansı daha ileri bir seviyeye taşınmıştır. Bu kapsamda, LSTM, RNN ve CNN modelleri, ZOA tarafından seçilen özellikler üzerinden eğitilerek değerlendirilmiştir. LSTM ve RNN modelleri, ağ trafiği verilerindeki zamansal bağımlılıkları etkin biçimde öğrenerek saldırı türlerinin daha doğru tespit edilmesini sağlamıştır. CNN modeli ise veri içerisindeki yerel örüntüleri ve yapısal ilişkileri yakalayarak sınıflandırma başarımını artırmıştır. Tüm algoritmalar, CSE-CIC-IDS2018 veri seti üzerinde siber saldırı tespiti amacıyla uygulanmış ve elde edilen sonuçlar; ZOA tabanlı özellik seçiminin, hem klasik hem de derin öğrenme modellerinde doğruluk, kesinlik, duyarlılık ve F1-skoru gibi performans metriklerini anlamlı düzeyde iyileştirdiğini göstermiştir. Özellikle derin öğrenme modelleri ile elde edilen sonuçlar, ZOA'nın saldırı tespit performansını üst seviyeye taşıyan etkili bir optimizasyon yöntemi olduğunu ortaya koymaktadır.

3.6 Performans Metrikleri

Bu çalışmada sağlanan metrikleri açıklayabilmek için temel metriklerle başlamak ve ardından bu temeller üzerine inşa etmek gereklidir. Uygulanabilir performans metrikleri listemiz şu şekildedir:

- **True Positive (TP)**: Pozitif olan durumların doğru bir şekilde pozitif olarak tanımlandığı sayı.
- **True Negative (TN)**: Negatif olan durumların doğru bir şekilde negatif olarak tanımlandığı sayı.
- **False Positive (FP)** (Tip I Hata olarak da bilinir): Negatif olan durumların yanlış bir şekilde pozitif olarak tanımlandığı sayı.
- **False Negative (FN)** (Tip II Hata olarak da bilinir): Pozitif olan durumların yanlış bir şekilde negatif olarak tanımlandığı sayı.

Bu temel metriklere dayanarak, diğer performans metrikleri şu şekilde türetilir:

- **Recall** (Duyarlılık): TP/(TP + FN) formülü ile hesaplanır.
- **Precision** (Kesinlik): TP/(TP + FP) formülü ile hesaplanır.
- **Accuracy** (Doğruluk): (TP + TN)/(TP + TN + FP + FN) formülü ile hesaplanır.

3.7 Çalışmada Kullanılan Bilgisayar Özellikleri

Uygulamalar, Çizelge 3.1'de belirtilen donanım ve yazılım özelliklerine sahip bir bilgisayarda gerçekleştirilmiştir.

Çizelge 3.1. Bilgisayar Özellikleri

İsim	Ayarlar
Donanım	
CPU	Apple M2 Pro
Frekans	3.49
RAM	16
Yazılım	
İşletim Sistemi	macOS Sonoma
Yazılım Dili	Python

3.8 Parametre Ayarları

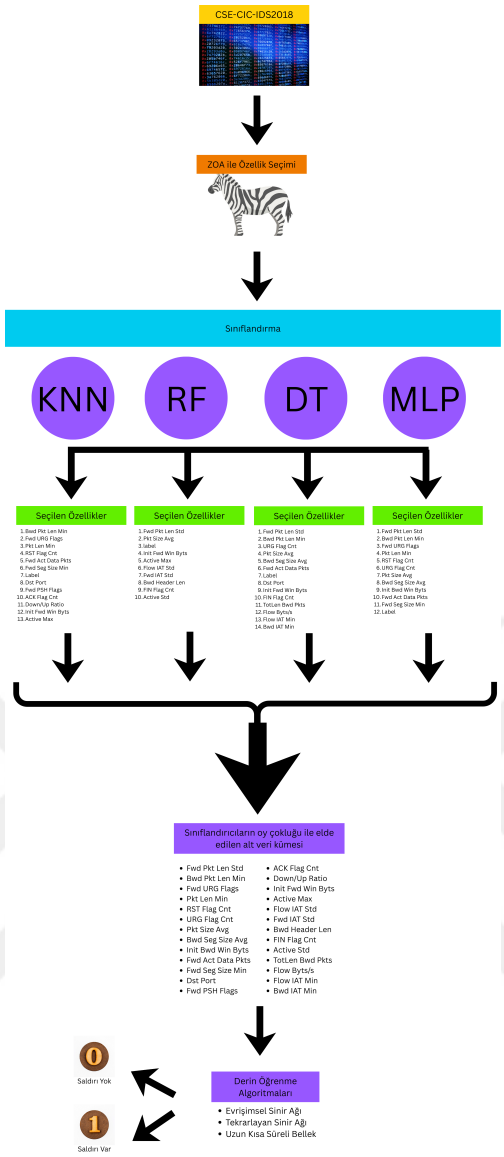
Çizelge 3.2'te uygulamalar için kullanılan ortak parametre ayarları gösterilmiştir.

Çizelge 3.2. Parametre Ayarları

Parametreler	Değerler
Popülasyon Boyutu (*N*)	40
Max. İterasyon sayısı (*T*)	100
Veri Seti	CIC-IDS-2018
Bağımsız Çalışma Sayısı (*Run*)	20

4. SONUÇLAR VE TARTIŞMA

Bu bölümde, Binary ZOA (BinZOA) kullanılarak CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneyler detaylı bir şekilde analiz edilmiştir. BinZOA, transfer fonksiyonu kullanılarak optimize edilmiş ve veri setindeki gereksiz özellikler elenerek önemli özelliklerin seçilmesi sağlanmıştır. Özellik seçimi sonrasında, farklı sınıflandırıcıların performansı derinlemesine değerlendirilmiştir. Bu kapsamda, KNN, MLP, RF, DT algoritmaları, CNN, RRN ve LSTM ile gerçekleştirilen sınıflandırma deneyleri ele alınmıştır. BinZOA algoritması ile optimize edilmiş özellikler, her bir sınıflandırıcı için en iyi fitness değeri, en iyi doğruluk, en iyi süre ve en iyi seçilen özelik metrikleri ile değerlendirilmiştir. Özellikle fitness değerleri, doğruluk oranları ve işlem süreleri gibi performans ölçütleri, algoritmaların etkinliğini ve verimliliğini karşılaştırmalı bir şekilde ortaya koymuştur. Her bir sınıflandırıcı, rastgele başlangıç noktalarının oluşturduğu olası değişiklikleri ve sonuçların güvenilirliğini değerlendirmek amacıyla 20 farklı tekrar ile çalıştırılmıştır. Bu tekrarlar, hem algoritmaların istikrarını hem de genel performansını test etmek için kullanılmıştır. Elde edilen sonuçlar, transfer fonksiyonunun ve BinZOA'nın, özellik seçimi sürecinde yüksek doğruluk oranları sağlarken aynı zamanda işlem süresini önemli ölçüde azalttığını göstermiştir. Tüm bu analizlerin özetlendiği sonuçlar Çizelge 4.9'da detaylı bir şekilde sunulmuştur. Bu kapsamlı değerlendirme, BinZOA ile optimize edilmiş özellik seçiminin, farklı sınıflandırıcılar üzerinde ne kadar etkili olduğunu açıkça ortaya koymuştur. Elde edilen en iyi özellik kümeleri, ikinci aşamada CNN, LSTM, RNN modellerine girdi olarak verilerek saldırı tespit performansı daha ileri bir seviyeye taşınmıştır. Şekil 4.1 de çalışmanın genel modeli sunulmuştur.



Şekil 4.1. Önerilen BinZOA tabanlı özellik seçimi ve sınıflandırma sürecinin genel modeli

4.1 KNN Algoritmasının Performans Değerlendirmesi

Bu bölümde, KNN algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. KNN, geniş kullanım alanına sahip basit ve etkili bir algoritma olarak sınıflandırma problemlerinde sıklıkla tercih edilmektedir. Günümüzde, birçok problem sınıflandırma formatına dönüştürülerek KNN gibi algoritmalarla etkili bir şekilde çözülebilmektedir. Bu çalışmada, CSE-CIC-IDS2018 veri seti kullanılarak siber saldırıların tahmin edilmesi amacıyla KNN algoritması uygulanmıştır. Algoritmanın tasarımı sırasında, en yakın komşu sayısını ifade eden *k* parametresi, veri setinin yapısına ve boyutuna uygun olarak 3 olarak belirlenmiştir. Bu parametre, algoritmanın doğruluk oranlarını optimize etmek için dikkatle seçilmiştir. Bu değerlendirme kapsamında, KNN algoritmasının performansı doğruluk (Accuracy), kesinlik

(Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel sınıflandırma metrikleri üzerinden analiz edilmiştir. Bu bölümde sunulan sonuçlar, BinZOA algoritması dâhil edilmeden elde edilmiş olup, algoritmanın ham performansını yansıtmaktadır. Bu durum, BinZOA algoritmasıyla optimize edilmiş sonuçlarla karşılaştırma yapılabilmesi için bir temel oluşturmayı amaçlamaktadır. Sonuçlar, KNN algoritmasının CSE-CIC-IDS2018 veri setinde yüksek doğruluk oranları ve dengeli sınıflandırma performansı sunduğunu göstermektedir. Ancak, daha ileri analizlerde BinZOA algoritmasıyla yapılan optimizasyonların, doğruluk oranı ve diğer metriklere olan etkisinin incelenmesi planlanmaktadır. Detaylı sonuçlar Çizelge 4.1'de sunulmuştur.

Çizelge 4.1. *k*-NN Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9159
Kesinlik (Precision)	0,9147
Duyarlılık (Recall)	0,9159
F1 skoru (F1-Score)	0,9153
Süre (Time)	4,4807

Çalışmanın devamında KNN algoritması BinZOA ile gerçekleştirilen özellik seçimi sonrasında, CSE-CIC-IDS2018 veri seti üzerinde 20 farklı tekrar ile test edilmiştir. Bu tekrarlar sırasında her iterasyon için aşağıdaki metrikler analiz edilmiştir:

İterasyon Sayısı (Iteration): BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngü sayısını ifade eder.

Fitness Değeri (Fitness): Her iterasyonda elde edilen en iyi fitness değerini gösterir. Bu değer, BinZOA'nın veri setindeki önemli özellikleri ne derece etkili seçtiğini ortaya koyar.

Süre (Time): Her bir iterasyonda *k*-NN algoritmasının eğitim ve test sürecinde harcadığı toplam süreyi ifade eder.

Doğruluk (Accuracy): İlgili iterasyonda *k*-NN tarafından sınıflandırılan verinin doğruluk oranını ifade eder.

20 farklı çalıştırma sonucunda elde edilen bu metrikler, KNN algoritmasının performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk oranları ve işlem süreleri arasındaki ilişki dikkatle incelenmiş ve her iterasyon için elde edilen sonuçlar Çizelge 4.2‘de gösterilmiştir. İlgili çizelgede 0. iterasyon KNN özellik seçimi olmadan ham olarak elde edilen sonuçlarıdır.

Çizelge 4.2. BinZOA ile KNN Sonuçları

Iterasyon	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Zaman	Özellik Sayısı	Fitness
0	0,9159	0,9147	0,9159	0,9153	4,4807	*k*-NN Ham Sonuçları	
1	0,957	0,9565	0,957	0,9566	0,1739	15	0,1536
2	0,972	0,9719	0,972	0,972	0,2034	13	0,126
3	0,9595	0,9594	0,9595	0,9594	0,2503	15	0,1531
4	0,9522	0,9518	0,9522	0,9519	0,2026	12	0,1339
5	0,9466	0,9459	0,9466	0,9462	0,2051	14	0,1536
6	0,9487	0,9483	0,9487	0,9484	0,1804	12	0,1351
7	0,9203	0,9191	0,9203	0,9196	0,2235	13	0,1643
8	0,9317	0,93	0,9317	0,9305	0,1458	9	0,1235
9	0,9573	0,957	0,9573	0,9571	0,234	13	0,1376
10	0,926	0,9244	0,926	0,925	0,1858	15	0,1724
11	0,9239	0,9239	0,9239	0,9239	0,2233	12	0,1549
12	0,9332	0,9319	0,9332	0,9324	0,1844	13	0,1529
13	0,9571	0,9568	0,9571	0,9569	0,1681	13	0,1368
14	0,9589	0,9587	0,9589	0,9588	0,2422	14	0,1455
15	0,925	0,9236	0,925	0,9242	0,1416	11	0,1434
16	0,9467	0,9458	0,9467	0,9461	0,2279	14	0,1517
17	0,9511	0,9508	0,9511	0,9509	0,2571	15	0,1573
18	0,9495	0,9494	0,9495	0,9494	0,1866	13	0,1424
19	0,9187	0,9166	0,9187	0,9174	0,1299	9	0,1349
20	0,9427	0,9419	0,9427	0,9422	0,1802	10	0,1231

Çizelge 4.2‘de, her bir iterasyonda elde edilen metrikleri özetlemektedir. Fitness değerleri, BinZOA'nın performansını ve optimize edilen çözümün kalitesini gösterirken, doğruluk oranları, KNN'nin sınıflandırma başarısını temsil eder. Süre, BinZOA ile optimize edilmiş özellikler ile çalışan KNN algoritmasının işlem verimliliğini değerlendirmek için kritik bir ölçüttür.

4.2 RF Algoritmasının Performans Değerlendirmesi

Bu bölümde, RF algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. Random Forest, geniş kullanım alanına sahip, esnek ve güçlü bir makine öğrenimi algoritmasıdır. Özellikle sınıflandırma problemlerinde yüksek doğruluk oranları ve dayanıklılığı ile yaygın olarak tercih edilmektedir. Günümüzde birçok problem, sınıflandırma formatında yeniden tasarlanarak RF gibi algoritmalarla başarılı bir şekilde çözülmektedir. Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, Random Forest algoritması siber güvenlik saldırılarını tahmin etmek ve sınıflandırmak amacıyla uygulanmıştır. Algoritma, veri setindeki karmaşık ilişkileri analiz ederek saldırı türlerini başarılı bir şekilde sınıflandırmıştır. Elde edilen sonuçlar, algoritmanın doğruluk (Accuracy), kesinlik (Precision),

duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel performans metrikleri üzerinden değerlendirilmiştir. Bu metrikler, Random Forest algoritmasının CSE-CIC-IDS2018 veri setinde genel olarak yüksek doğruluk oranı ve dengeli bir performans sunduğunu göstermektedir. Accuracy değeri, modelin genel sınıflandırma doğruluğunu ifade ederken, Precision ve Recall değerleri, modelin pozitif sınıfları doğru bir şekilde tahmin etme kapasitesini ortaya koymaktadır. F1-Score, doğruluk ve duyarlılığı dengeleyerek modelin genel performansını tek bir metrikte özetlemektedir. Elde edilen bu sonuçlar, Random Forest algoritmasının, özellikle siber saldırı tespiti gibi kritik sınıflandırma problemlerinde güçlü bir aday olduğunu açıkça ortaya koymaktadır. Elde edilen sonuçlar Çizelge 4.3'de gösterilmiştir.

Çizelge 4.3. Random Forest Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9998
Kesinlik (Precision)	0,9998
Duyarlılık (Recall)	0,9998
F1 skoru (F1-Score)	0,9998
Süre (Time)	2,3912

Çalışmanın devamında, RF algoritması, BinZOA ile gerçekleştirilen özellik seçimi sonrasında CSE-CIC-IDS2018 veri seti üzerinde 20 farklı tekrar ile test edilmiştir. Bu testlerde, her iterasyon için aşağıdaki metrikler analiz edilmiştir:

- *İterasyon Sayısı (Iteration):* ZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngü sayısını ifade eder.
- *Fitness Değeri (Fitness):* Her iterasyonda elde edilen en iyi fitness değerini gösterir. Bu değer, BinZOA'nın veri setindeki önemli özellikleri ne kadar etkili seçtiğini ortaya koyar.
- *Seçilen Özellik Sayısı:* BinZOA tarafından seçilen ve modelin eğitimi sırasında kullanılan özelliklerin sayısını ifade eder.
- *Süre (Time):* Her bir iterasyonda Random Forest algoritmasının eğitim ve test sürecinde harcadığı toplam süreyi ifade eder.
- *Doğruluk (Accuracy):* İlgili iterasyonda Random Forest algoritması tarafından sınıflandırılan verinin doğruluk oranını ifade eder.
- *Precision, Recall ve F1-Score:* Algoritmanın sınıflandırma başarısını değerlendirmek için kullanılan ek metriklerdir.

20 farklı çalıştırma sonucunda elde edilen bu metrikler, Random Forest algoritmasının performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk

oranları, seçilen özellik sayıları ve işlem süreleri arasındaki ilişki dikkatle incelenmiş ve her iterasyon için elde edilen sonuçlar Çizelge 4.4.'te özetlenmiştir. İlgili çizelgede, 0. iterasyonlar, RF algoritmasının BinZOA özellik seçimi olmadan ham olarak elde edilen sonuçlarını temsil etmektedir.

Çizelge 4.4. BinZOA ile Random Forest Sonuçları

Iterasyon	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Zaman	Özellik Sayısı	Fitness
0	0,9998	0,9998	0,9998	0,9998	2,3912	RF Ham Sonuçları	
1	0,9999	0,9999	0,9999	0,9999	1,4497	10	0,0814
2	0,9993	0,9993	0,9993	0,9993	1,4594	12	0,0976
3	0,9997	0,9997	0,9997	0,9997	1,6411	12	0,0975
4	0,9996	0,9996	0,9996	0,9996	1,2109	13	0,1057
5	0,9987	0,9987	0,9987	0,9987	1,608	12	0,0976
6	0,9995	0,9995	0,9995	0,9995	1,2956	12	0,0975
7	0,9995	0,9995	0,9995	0,9995	1,7496	10	0,0814
8	0,9998	0,9998	0,9998	0,9998	1,3436	10	0,0813
9	0,9996	0,9996	0,9996	0,9996	1,6244	12	0,0976
10	0,9997	0,9997	0,9997	0,9997	1,6179	11	0,0896
11	0,9986	0,9986	0,9986	0,9986	1,9347	11	0,0899
12	0,9996	0,9996	0,9996	0,9996	1,9324	12	0,0976
13	0,9995	0,9995	0,9995	0,9995	1,1662	11	0,0895
14	0,9991	0,9991	0,9991	0,9991	2,0571	12	0,0977
15	0,9995	0,9995	0,9995	0,9995	1,7307	12	0,0976
16	0,9989	0,9989	0,9989	0,9989	0,9737	12	0,0975
17	0,9997	0,9997	0,9997	0,9997	1,5211	11	0,0894
18	0,9994	0,9994	0,9994	0,9994	1,9046	13	0,1057
19	0,9999	0,9999	0,9999	0,9999	1,1715	12	0,0974
20	0,9987	0,9987	0,9987	0,9987	1,1175	14	0,1139

Elde edilen sonuçlar, RF algoritmasının BinZOA ile optimize edilmiş özellik seçimi sonrası performansının yüksek olduğunu ve doğru özellik seçiminin algoritmanın doğruluk oranını etkili bir şekilde desteklediğini göstermektedir. Ayrıca, BinZOA kullanılmadan yapılan 0. iterasyondaki sonuçlar ile karşılaştırıldığında, optimize edilmiş özellik seçiminin sınıflandırma başarısını artırdığı açıkça görülmektedir. Bu sonuçlar, özellikle büyük veri setlerinde BinZOA ve RF algoritmasının birlikte kullanımının verimli olduğunu ortaya koymaktadır.

4.3 MLP Algoritmasının Performans Değerlendirmesi

Bu bölümde, MLP algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. MLP, geniş kullanım alanına sahip, güçlü bir yapay sinir ağı algoritmasıdır. Özellikle

karmaşık sınıflandırma problemlerinde yüksek doğruluk oranları ve esnek yapısıyla tercih edilmektedir. Günümüzde birçok problem, MLP gibi algoritmalarla başarılı bir şekilde çözülmektedir. Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, MLP algoritması siber güvenlik saldırılarını tahmin etmek ve sınıflandırmak amacıyla uygulanmıştır. Algoritma, veri setindeki karmaşık ilişkileri analiz ederek saldırı türlerini başarılı bir şekilde sınıflandırmıştır.

Elde edilen sonuçlar, algoritmanın doğruluk oranı (Accuracy), kesinlik (Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel performans metrikleri üzerinden değerlendirilmiştir. Bu metrikler, ileri beslemeli bir yapay sinir ağı (feedforward MLP) olan MLP algoritmasının CSE-CIC-IDS2018 veri setinde genel olarak yüksek doğruluk oranı ve dengeli bir performans sunduğunu göstermektedir. Kullanılan MLP modeli, iki gizli katman içerir ve bu katmanlarda sırasıyla 100 ve 50 nöron bulunmaktadır. Bu yapı, karmaşık ilişkileri öğrenmek için yeterli kapasite sağlar ve ReLU aktivasyon fonksiyonu sayesinde doğrusal olmayan ilişkileri etkili bir şekilde modelleyebilir. Model, Adam optimizasyon algoritması ile eğitilmiş ve mini-batch boyutu 32 olarak belirlenmiştir. Erken durdurma mekanizması kullanılarak aşırı öğrenme engellenmiştir. Accuracy değeri, modelin genel sınıflandırma doğruluğunu ifade ederken; Precision ve Recall değerleri, modelin pozitif sınıfları doğru bir şekilde tahmin etme kapasitesini ortaya koymaktadır. F1-Score, doğruluk ve duyarlılığı dengeleyerek modelin genel performansını tek bir metrikte özetlemektedir. Elde edilen bu sonuçlar, MLP algoritmasının, özellikle siber saldırı tespiti gibi kritik sınıflandırma problemlerinde güçlü bir aday olduğunu açıkça ortaya koymaktadır. Elde edilen ham sonuçlar, Çizelge 4.5'de gösterilmiştir.

Çizelge 4.5. MLP Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9985
Kesinlik (Precision)	0,9985
Duyarlılık (Recall)	0,9985
F1 skoru (F1-Score)	0,9985
Süre (Time)	15,8696

Çalışmanın devamında, MLP algoritması, BinZOA ile gerçekleştirilen özellik seçimi sonrasında CSE-CIC-IDS2018 veri seti üzerinde 20 farklı tekrar ile test edilmiştir. Bu testlerde, her iterasyon için aşağıdaki metrikler analiz edilmiştir:

- *İterasyon Sayısı (Iteration):* BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngü sayısını ifade eder.

- *Fitness Değeri (Fitness):* Her iterasyonda elde edilen en iyi fitness değerini gösterir. Bu değer, BinZOA'nın veri setindeki önemli özellikleri ne kadar etkili seçtiğini ortaya koyar.
- *Seçilen Özellik Sayısı:* BinZOA tarafından seçilen ve modelin eğitimi sırasında kullanılan özelliklerin sayısını ifade eder.
- *Süre (Time):* Her bir iterasyonda MLP algoritmasının eğitim ve test sürecinde harcadığı toplam süreyi ifade eder.
- *Doğruluk (Accuracy):* İlgili iterasyonda MLP algoritması tarafından sınıflandırılan verinin doğruluk oranını ifade eder.
- *Precision, Recall ve F1-Score:* Algoritmanın sınıflandırma başarısını değerlendirmek için kullanılan ek metriklerdir.

20 farklı çalıştırma sonucunda elde edilen bu metrikler, MLP algoritmasının performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk oranları, seçilen özellik sayıları ve işlem süreleri arasındaki ilişki dikkatle incelenmiş ve her iterasyon için elde edilen sonuçlar Çizelge 4.6'da özetlenmiştir. İlgili çizelgede, 0. iterasyonlar, MLP algoritmasının BinZOA özellik seçimi olmadan ham olarak elde edilen sonuçlarını temsil etmektedir.

Çizelge 4.6. BinZOA ile MLP Sonuçları

Iterasyon	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Zaman	Özellik Sayısı	Fitness
0	0,9985	0,9985	0,9985	0,9985	15,8696	MLP Ham Sonuçları	
1	1	1	1	1	19,9484	12	0,1795
2	0,9983	0,9983	0,9983	0,9983	17,2838	13	0,3917
3	0,9993	0,9993	0,9993	0,9993	19,5049	17	0,4187
4	0,9986	0,9986	0,9986	0,9986	18,2514	13	0,3506
5	0,9984	0,9984	0,9984	0,9984	16,8444	16	0,3877
6	0,999	0,999	0,999	0,999	22,3695	17	0,3777
7	0,999	0,999	0,999	0,999	20,2446	17	0,363
8	0,9993	0,9993	0,9993	0,9993	25,1505	17	0,4037
9	0,9985	0,9986	0,9985	0,9985	20,591	19	0,3694
10	0,9991	0,9991	0,9991	0,9991	15,9318	21	0,406
11	0,9987	0,9987	0,9987	0,9987	20,4036	21	0,395
12	0,9981	0,9981	0,9981	0,9981	15,352	18	0,4029
13	0,9982	0,9982	0,9982	0,9982	15,0478	17	0,3559
14	0,9988	0,9988	0,9988	0,9988	17,7876	16	0,3862
15	0,9006	0,8291	0,9006	0,8614	3839,5361	16	0,4142
16	0,9993	0,9993	0,9993	0,9993	20,5639	17	0,3536
17	0,9984	0,9984	0,9984	0,9984	18,693	16	0,3741
18	0,9989	0,9989	0,9989	0,9989	15,8158	13	0,3689

19	0,9983	0,9983	0,9983	0,9983	21,7487	24	0,4138
20	0,9994	0,9994	0,9994	0,9994	19,3795	17	0,3612

Elde edilen sonuçlar, MLP algoritmasının BinZOA ile optimize edilmiş özellik seçimi sonrası performansının yüksek olduğunu ve doğru özellik seçiminin algoritmanın doğruluk oranını etkili bir şekilde desteklediğini göstermektedir. Ayrıca, BinZOA kullanılmadan yapılan 0. iterasyondaki sonuçlar ile karşılaştırıldığında, optimize edilmiş özellik seçiminin sınıflandırma başarısını artırdığı açıkça görülmektedir. Bu sonuçlar, özellikle büyük veri setlerinde BinZOA ve MLP algoritmasının birlikte kullanımının verimli olduğunu ortaya koymaktadır.

4.4 DT Algoritmasının Performans Değerlendirmesi

Bu bölümde, DT algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. Decision Tree, basit ve anlaşılabilir yapısı sayesinde sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılan bir algoritmadır. Özellikle karar verme süreçlerini görselleştirebilmesi ve yorumlanabilir sonuçlar sunması, algoritmanın tercih edilme nedenlerindendir. Günümüzde birçok problem, Decision Tree algoritmasıyla etkili bir şekilde çözülmektedir.

Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, DT algoritması siber güvenlik saldırılarını tahmin etmek ve sınıflandırmak amacıyla uygulanmıştır. Algoritma, veri setindeki özellikleri değerlendirerek saldırı türlerini doğru bir şekilde sınıflandırmıştır. Performans değerlendirmesi sırasında doğruluk oranı (Accuracy), kesinlik (Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel metrikler incelenmiştir. Bu metrikler, DT algoritmasının CSE-CIC-IDS2018 veri setinde genel olarak yüksek doğruluk oranı sunduğunu göstermektedir. Özellikle Accuracy değeri, modelin genel sınıflandırma başarısını ifade ederken, diğer metrikler, pozitif sınıfları doğru bir şekilde tahmin etme kapasitesini göstermektedir. Elde edilen ham sonuçlar, Çizelge 4.7'de gösterilmiştir.

Çizelge 4.7. Desicion Tree Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9985
Kesinlik (Precision)	0,9985
Duyarlılık (Recall)	0,9985
F1 skoru (F1-Score)	0,9985
Süre (Time)	15,8696

Çalışmanın devamında, DT algoritması, BinZOA ile gerçekleştirilen özellik seçimi sonrasında CSE-CIC-IDS2018 veri seti üzerinde 20 farklı tekrar ile test edilmiştir. Bu testlerde, her iterasyon için aşağıdaki metrikler analiz edilmiştir:

- *İterasyon Sayısı (Iteration):* BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngü sayısını ifade eder.
- *Fitness Değeri (Fitness):* Her iterasyonda elde edilen en iyi fitness değerini gösterir. Bu değer, BinZOA'nın veri setindeki önemli özellikleri ne kadar etkili seçtiğini ortaya koyar.
- *Seçilen Özellik Sayısı:* BinZOA tarafından seçilen ve modelin eğitimi sırasında kullanılan özelliklerin sayısını ifade eder.
- *Süre (Time):* Her bir iterasyonda Decision Tree algoritmasının eğitim ve test sürecinde harcadığı toplam süreyi ifade eder.
- *Doğruluk (Accuracy):* İlgili iterasyonda Decision Tree algoritması tarafından sınıflandırılan verinin doğruluk oranını ifade eder.
- *Precision, Recall ve F1-Score:* Algoritmanın sınıflandırma başarısını değerlendirmek için kullanılan ek metriklerdir.

20 farklı çalıştırma sonucunda elde edilen bu metrikler, Decision Tree algoritmasının performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk oranları, seçilen özellik sayıları ve işlem süreleri arasındaki ilişki dikkatle incelenmiş ve her iterasyon için elde edilen sonuçlar Çizelge 4.8'de özetlenmiştir. İlgili çizelgede, 0. iterasyonlar, Decision Tree algoritmasının BinZOA özellik seçimi olmadan elde edilen sonuçlarını temsil etmektedir.

Çizelge 4.8. BinZOA DT Sonuçları

Iterasyon	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Zaman	Özellik Sayısı	Fitness
0	0,8677	0,8831	0,8677	0,8562	0,0718	Desicion Tree Ham Sonuçları	
1	0,8632	0,8586	0,8632	0,8518	0,0422	12	0,0973
2	0,7528	0,8343	0,7528	0,751	0,0425	13	0,1054
3	0,8637	0,8446	0,8637	0,8447	0,045	12	0,0973
4	0,69	0,7991	0,69	0,6923	0,0442	13	0,1054
5	0,7752	0,8159	0,7752	0,7616	0,0405	13	0,1054
6	0,7835	0,8319	0,7835	0,7791	0,0314	13	0,1054
7	0,7994	0,8163	0,7994	0,7872	0,0499	11	0,0892
8	0,6392	0,7395	0,6392	0,628	0,0348	12	0,0973
9	0,6753	0,6923	0,6753	0,6463	0,0465	11	0,0892
10	0,873	0,9041	0,873	0,8658	0,0447	12	0,0973
11	0,7822	0,8161	0,7822	0,7671	0,0452	13	0,1054
12	0,7752	0,8826	0,7752	0,7535	0,0505	12	0,0973

13	0,9524	0,964	0,9524	0,9501	0,0393	14	0,1135
14	0,7143	0,8619	0,7143	0,7211	0,0308	11	0,0892
15	0,8906	0,9127	0,8906	0,8837	0,0371	12	0,0973
16	0,7732	0,8289	0,7732	0,7652	0,0399	12	0,0973
17	0,6379	0,7296	0,6379	0,6166	0,0498	12	0,0973
18	0,8432	0,87	0,8432	0,8378	0,0615	13	0,1054
19	0,6805	0,663	0,6805	0,6379	0,0576	12	0,0973
20	0,6804	0,7561	0,6804	0,6742	0,0314	10	0,0811

Elde edilen sonuçlar, DT algoritmasının BinZOA ile optimize edilmiş özellik seçimi sonrası performansının yüksek olduğunu ve doğru özellik seçiminin algoritmanın doğruluk oranını etkili bir şekilde desteklediğini göstermektedir. Ayrıca, BinZOA kullanılmadan yapılan 0. iterasyondaki sonuçlar ile karşılaştırıldığında, optimize edilmiş özellik seçiminin sınıflandırma başarısını artırdığı açıkça görülmektedir. Bu sonuçlar, özellikle büyük veri setlerinde BinZOA ve DT algoritmasının birlikte kullanımının verimli olduğunu ortaya koymaktadır.

4.5 Makine Öğrenmesi Sınıflandırıcı Algoritmalarının Sonuçlarının Karşılaştırılması

Bu bölümde, KNN, MLP, RF, DT algoritmalarının BinZOA ile optimize edilmiş sonuçları karşılaştırılmış ve en iyi performansı sunan algoritmalar belirlenmiştir. Karşılaştırma, doğruluk oranı (Accuracy), fitness değeri (Fitness), işlem süresi (Time), kesinlik (Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel metrikler üzerinden gerçekleştirilmiştir. Sonuçlar, Çizelge 4.9'da gösterilmiştir.

Çizelge 4.9. Genel Algoritma Karşılaştırılması

Algoritma	Iterasyon	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Zaman	Özellik Sayısı
MLP	2	1	1	1	1	19,9484	
KNN	3	0,972	0,9719	0,972	0,972	0,2034	13
RF	2	0,9999	0,9999	0,9999	0,9999	1,4497	10
RF	20	0,9999	0,9999	0,9999	0,9999	1,1715	12
MLP	2	1	1	1	1	19,9484	12
KNN	3	0,972	0,9719	0,972	0,972	0,2034	13
RF	2	0,9999	0,9999	0,9999	0,9999	1,4497	10

ZOA'nın özellik seçimi perfonmansını inceleyecek olursak; özellik seçimi, verinin boyutunu azaltarak modellerin performansını artırmayı ve işlem sürelerini optimize etmeyi hedeflemektedir. BinZOA algoritmasının farklı sınıflandırıcılarla (KNN, MLP, RF, DT)

entegre edilmesiyle seçilen özelliklerin doğruluk, fitness değeri, işlem süresi ve diğer başarı ölçütleri üzerindeki etkisi karşılaştırmalı olarak değerlendirilmiştir. Özellik seçiminin etkisi, sadece model performansını artırmakla kalmamış, aynı zamanda gereksiz verilerin azaltılması yoluyla daha hızlı ve verimli bir işlem süreci sağlamıştır. Aşağıdaki Çizelge 4.10 ve Çizelge 4.11, BinZOA tarafından seçilen özelliklerin her bir algoritma üzerindeki etkisi detaylı olarak sunulmaktadır.

Çizelge 4.10. BinZOA İle Bazlı Özellik Seçim Değerleri-1

Özellikler	1	2	3	4	5	6	7	8	9	10
Dst Port	DT	KNN	DT, MLP	MLP	MLP	KNN, MLP	RF, DT, MLP	RF	RF, MLP	KNN, RF, MLP
Protocol	KNN	MLP	-	KNN, DT, MLP	RF	MLP	-	KNN	DT	-
TotLen Bwd Pkts	KNN	DT	MLP	DT	KNN	KNN, RF, DT	KNN, DT, MLP	-	KNN, RF, DT	KNN, MLP
Fwd Pkt Len Std	RF, MLP	RF, MLP	RF, MLP	KNN	KNN, RF	DT, MLP	KNN, MLP	-	RF, MLP	MLP
Bwd Pkt Len Min	KNN, DT, MLP	KNN, RF	MLP	KNN, DT, MLP	KNN, DT	RF, MLP	-	DT, MLP	DT, MLP	-
Flow Byts/s	-	DT	-	DT	DT, MLP	-	KNN, RF, DT, MLP	MLP	MLP	-
Flow IAT Mean	DT	-	DT	-	-	RF	RF	KNN	-	KNN, RF, DT
Flow IAT Std	KNN, RF, DT	-	RF	-	DT	-	MLP	MLP	MLP	KNN, DT, MLP
Flow IAT Min	-	RF	MLP	-	-	-	DT	-	RF, DT	-
Fwd IAT Std	RF	DT	RF, DT	KNN, RF, DT	KNN, RF	DT, MLP	-	RF, MLP	DT, MLP	RF, MLP
Bwd IAT Std	KNN, DT	-	KNN	-	RF, MLP	-	RF, MLP	KNN	RF, MLP	KNN, RF, MLP
Bwd IAT Max	KNN	RF, DT	KNN, DT	KNN	MLP	-	KNN, MLP	RF, DT	-	DT, MLP
Bwd IAT Min	-	RF, MLP	RF	-	RF, MLP	-	-	KNN	KNN, DT, MLP	DT
Fwd PSH Flags	-	KNN	KNN, MLP	KNN, RF	KNN, RF, DT	RF	-	MLP	MLP	DT

Fwd URG Flags	DT, MLP	KNN, MLP	DT, MLP	RF	KNN	DT, MLP	KNN	MLP	KNN	KNN
Bwd Header Len	RF, DT	MLP	KNN, RF, DT	KNN, RF, DT, MLP	KNN, DT, MLP	MLP	MLP	MLP	-	KNN, MLP
Bwd Pkts/s	-	-	RF	RF, MLP	RF	KNN	DT	DT	RF	RF, MLP
Pkt Len Min	KNN, MLP	KNN	RF, DT, MLP	DT	KNN, MLP	KNN, RF, MLP	RF	RF, DT	RF, DT, MLP	KNN, RF
FIN Flag Cnt	KNN, RF	-	MLP	DT	-	KNN	DT, MLP	MLP	MLP	RF, MLP
RST Flag Cnt	MLP	KNN, DT, MLP	KNN, MLP	RF, DT, MLP	KNN, MLP	DT, MLP	DT	DT	-	KNN
PSH Flag Cnt	DT	DT, MLP	KNN, MLP	RF	MLP	DT	MLP	MLP	-	DT, MLP
ACK Flag Cnt	KNN	KNN, MLP	KNN	RF	KNN, DT	RF, DT	KNN, MLP	DT	KNN, MLP	KNN, RF, MLP
URG Flag Cnt	MLP	RF, DT	DT, MLP	MLP	MLP	KNN, RF, DT, MLP	-	DT	RF, MLP	-
Down/Up Ratio	DT	KNN, DT	-	RF	KNN, RF, DT, MLP	-	KNN, RF	MLP	KNN, DT, MLP	DT
Pkt Size Avg	KNN, RF, MLP	RF	DT, MLP	DT, MLP	DT, MLP	-	RF	DT, MLP	KNN	MLP
Fwd Seg Size Avg	KNN, DT	MLP	KNN, RF, MLP	MLP	-	KNN, DT	KNN	KNN, RF	-	-
Bwd Seg Size Avg	KNN, MLP	DT	-	KNN	RF, MLP	KNN, RF, MLP	KNN	RF, DT	KNN	DT
Init Fwd Win Byts	KNN, RF	KNN, MLP	KNN, RF	KNN, MLP	DT	KNN	KNN, DT, MLP	KNN, RF	-	KNN, DT, MLP
Init Bwd Win Byts	MLP	RF, DT	KNN, RF, DT	RF, MLP	KNN, MLP	MLP	RF, DT	MLP	-	KNN
Fwd Act Data Pkts	DT, MLP	KNN, MLP	-	-	KNN, RF, DT	RF	KNN, DT	KNN, RF, DT	KNN, DT	KNN, DT, MLP
Fwd Seg Size Min	MLP	KNN, MLP	DT, MLP	RF, MLP	MLP	KNN, RF, DT, MLP	MLP	-	-	RF, MLP
Active Std	RF	RF, DT	KNN	-	RF	MLP	-	RF, MLP	KNN	MLP

Active Max	RF	KNN, RF, DT	KNN	DT	DT	KNN, RF, DT, MLP	RF, MLP	-	MLP	MLP
Active Min	KNN	-	KNN, MLP	-	-	-	MLP	KNN, DT, MLP	KNN, RF, DT, MLP	-
Idle Std	-	RF	-	KNN, RF, DT	DT	-	-	MLP	KNN, RF, MLP	RF, MLP
Idle Min	-	-	KNN, RF	KNN	-	DT, MLP	KNN, MLP	MLP	KNN, RF, MLP	KNN, DT, MLP
Label	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP

Çizelge 4.11. BinZOA İle Bazlı Özellik Seçim Değerleri-2

Özellikler	11	12	13	14	15	16	17	18	19	20
Dst Port	KNN	DT	KNN, DT, MLP	DT, MLP	KNN, MLP	KNN, DT, MLP	MLP	KNN, RF, MLP	KNN, DT, MLP	KNN, DT
Protocol	DT, MLP	DT	KNN, RF	-	MLP	RF, DT	KNN, DT	-	MLP	KNN, RF
TotLen Bwd Pkts	DT	-	KNN, DT, MLP	MLP	DT, MLP	KNN, RF, MLP	RF, MLP	-	-	RF, MLP
Fwd Pkt Len Std	MLP	DT, MLP	DT, MLP	KNN, DT, MLP	RF, DT, MLP	MLP	DT	RF, DT, MLP	MLP	-
Bwd Pkt Len Min	DT, MLP	RF	DT	-	-	MLP	KNN	-	MLP	RF, DT, MLP
Flow Byts/s	KNN, MLP	RF, MLP	DT	RF	MLP	-	-	DT, MLP	DT, MLP	-
Flow IAT Mean	RF	KNN, RF, DT	-	RF	RF	-	RF	RF	-	-
Flow IAT Std	RF, DT, MLP	KNN, DT, MLP	RF, MLP	RF, MLP	DT	-	DT	DT	KNN	MLP
Flow IAT Min	RF	DT	DT	-	KNN, DT, MLP	-	MLP	RF	DT	-
Fwd IAT Std	-	KNN, RF	-	RF, MLP	-	-	MLP	DT	MLP	DT
Bwd IAT Std	KNN, MLP	-	-	MLP	KNN, RF	MLP	-	-	RF, MLP	RF, MLP
Bwd IAT Max	KNN	RF, MLP	KNN, MLP	-	RF	KNN	KNN, RF	RF	DT, MLP	KNN, RF
Bwd IAT Min	RF, MLP	RF, MLP	KNN, DT	KNN, DT	KNN, RF, DT, MLP	RF	KNN, DT, MLP	KNN, RF, MLP	KNN, DT, MLP	KNN, MLP
Fwd PSH Flags	-	MLP	KNN, RF	MLP	KNN, RF, DT	KNN, RF, MLP	KNN, DT, MLP	KNN, DT	-	KNN, RF
Fwd URG Flags	MLP	-	RF, MLP	KNN	RF, MLP	RF, MLP	KNN, RF	-	MLP	RF, MLP
Bwd Header Len	KNN, MLP	DT, MLP	-	RF, DT, MLP	MLP	MLP	KNN, RF, DT	KNN, DT	MLP	KNN
Bwd Pkts/s	DT	-	-	RF, DT	RF	DT	DT	KNN, RF, DT, MLP	KNN, RF	-
Pkt Len Min	KNN, RF	KNN, MLP	KNN, RF, MLP	KNN, DT	MLP	RF	DT, MLP	KNN	-	RF
FIN Flag Cnt	KNN, MLP	-	DT	KNN, MLP	RF, MLP	KNN, DT	MLP	DT	KNN, RF, MLP	DT

RST Flag Cnt	MLP	-	-	KNN	-	MLP	RF	RF, MLP	KNN, RF	DT
PSH Flag Cnt	DT, MLP	KNN, RF	KNN	RF, MLP	-	KNN, RF, MLP	RF, MLP	KNN, MLP	KNN, RF	KNN, RF, DT, MLP
ACK Flag Cnt	MLP	KNN	MLP	KNN	-	KNN, DT	-	KNN	KNN	MLP
URG Flag Cnt	RF	KNN, DT, MLP	RF, DT, MLP	KNN, DT, MLP	MLP	DT	KNN, RF, DT	-	RF	RF, MLP
Down/Up Ratio	KNN, DT	KNN, RF, MLP	KNN, RF, MLP	KNN	RF, DT	KNN, MLP	DT	KNN	DT, MLP	DT
Pkt Size Avg	RF, DT, MLP	MLP	DT	-	KNN, DT	-	KNN, RF, MLP	-	-	KNN, MLP
Fwd Seg Size Avg	RF, MLP	RF, DT, MLP	KNN, RF	RF	MLP	KNN	KNN, RF, MLP	-	MLP	MLP
Bwd Seg Size Avg	KNN, MLP	MLP	DT, MLP	KNN, RF, MLP	KNN	MLP	-	RF	MLP	MLP
Init Fwd Win Byts	DT	KNN, RF	KNN, DT	KNN, DT	KNN, RF, MLP	DT	KNN, MLP	KNN, MLP	RF	-
Init Bwd Win Byts	-	KNN	-	MLP	MLP	KNN, MLP	MLP	MLP	MLP	-
Fwd Act Data Pkts	KNN, MLP	-	RF, DT, MLP	RF, MLP	KNN	MLP	KNN, DT	RF, DT	RF, DT, MLP	MLP
Fwd Seg Size Min	DT, MLP	-	KNN, MLP	-	KNN, DT	KNN, RF, MLP	KNN, MLP	MLP	RF, MLP	-
Active Std	RF	DT	-	DT, MLP	-	KNN, RF, MLP	-	DT, MLP	MLP	DT, MLP
Active Max	MLP	RF, DT, MLP	-	DT	MLP	DT	-	KNN, RF, DT	RF, DT, MLP	DT, MLP
Active Min	KNN, DT	MLP	RF, MLP	RF	DT	DT	MLP	DT, MLP	RF, DT, MLP	RF, MLP
Idle Std	RF, DT	KNN, MLP	MLP	KNN	-	KNN, RF, DT	-	KNN	DT, MLP	RF
Idle Min	MLP	KNN, MLP	MLP	KNN	DT	RF, DT	KNN	RF	DT, MLP	KNN, RF
Label	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP	KNN, RF, DT, MLP

Algoritmaların en iyi sonuçları Çizelge 4.9'da belirtilmiştir. Bu sonuçlara ait özellik seçimleri ise Çizelge 4.12' de detaylı olarak sunulmuştur.

Çizelge 4.12. Seçilen en iyi özellikler

Özellik	MLP	*k*-NN	RF	DT
Fwd Pkt Len Std	Evet	Hayır	Evet	Evet
Bwd Pkt Len Min	Evet	Evet	Hayır	Evet
Fwd URG Flags	Evet	Evet	Hayır	Hayır
Pkt Len Min	Evet	Evet	Hayır	Hayır
RST Flag Cnt	Evet	Evet	Hayır	Hayır
URG Flag Cnt	Evet	Hayır	Hayır	Evet
Pkt Size Avg	Evet	Hayır	Evet	Evet
Bwd Seg Size Avg	Evet	Hayır	Hayır	Evet
Init Bwd Win Byts	Evet	Hayır	Hayır	Hayır
Fwd Act Data Pkts	Evet	Evet	Hayır	Evet
Fwd Seg Size Min	Evet	Evet	Hayır	Hayır
Label	Evet	Evet	Evet	Evet
Dst Port	Hayır	Evet	Hayır	Evet
Fwd PSH Flags	Hayır	Evet	Hayır	Hayır
ACK Flag Cnt	Hayır	Evet	Hayır	Hayır
Down/Up Ratio	Hayır	Evet	Hayır	Hayır
Init Fwd Win Byts	Hayır	Evet	Evet	Evet
Active Max	Hayır	Evet	Evet	Hayır
Flow IAT Std	Hayır	Hayır	Evet	Hayır
Fwd IAT Std	Hayır	Hayır	Evet	Hayır
Bwd Header Len	Hayır	Hayır	Evet	Hayır
FIN Flag Cnt	Hayır	Hayır	Evet	Evet
Active Std	Hayır	Hayır	Evet	Hayır
TotLen Bwd Pkts	Hayır	Hayır	Hayır	Evet
Flow Byts/s	Hayır	Hayır	Hayır	Evet
Flow IAT Min	Hayır	Hayır	Hayır	Evet
Bwd IAT Min	Hayır	Hayır	Hayır	Evet

Çizelge 4.12 incelendiğinde en iyi başarı oranı (Accuarcy) MLP algoritmasına aittir. Bu sebeple siber güvenlik atak tespitinde en iyi özellikler MLP'nin seçtiği 12 özelliktir.

Bu çalışmada yapılan analizler sonucunda, algoritmaların farklı metrikler açısından birbirinden farklı performanslar sergilediği gözlemlenmiştir. Random Forest algoritması, BinZOA ile optimize edilen özellik seçimi sonrasında elde ettiği yüksek doğruluk (0,9999) ve dengeli işlem süresi ile öne çıkmıştır. İki farklı yapılandırmada (2 ve 20 ağaç) da aynı doğruluk ve F1 skoru elde edilmiştir; bu da algoritmanın genel sınıflandırma başarısında tutarlı olduğunu göstermektedir. Özellikle, işlem süresi açısından 20 ağaçlı yapı (1,1715) daha hızlı sonuçlar sunmuştur. Bunun yanında, *k*-Nearest Neighbors (*k*-NN) algoritması, basit yapısı sayesinde en

hızlı işlem süresine (0,2034) sahip olmuştur. Ancak, hız avantajına rağmen doğruluk oranı (0,972) açısından Random Forest ve MLP'nin gerisinde kalmıştır. Bu durum, *k*-NN'nin hızın kritik önem taşıdığı durumlarda daha uygun bir seçenek olduğunu ortaya koymaktadır. MLP algoritması, en yüksek doğruluk (1,0) ve F1 skoru (1,0) ile genel performans açısından en başarılı algoritma olarak öne çıkmıştır. Ancak, işlem süresi (19,9484) diğer algoritmalara kıyasla oldukça yüksek olduğundan, zaman maliyetinin göz önünde bulundurulması gereken senaryolarda dezavantajlı olabilir. Yine de yüksek kesinlik oranı (1,0), yanlış pozitif tahminlerin minimuma indirilmesi gereken durumlar için MLP'yi güçlü bir alternatif haline getirmiştir. Çalışmada elde edilen sonuçlar, sınıflandırıcıların BinZOA ile optimize edilmiş özellik seçimi sonrasında sınıflandırma başarılarının belirgin şekilde arttığını göstermiştir. Özellikle, Random Forest algoritması, doğruluk oranı ve işlem süresi dengesi ile öne çıkarken, MLP algoritması pozitif sınıfları doğru tahmin etme kapasitesi ile dikkat çekmiştir. KNN ise en hızlı sonuçları sunarak zamana duyarlı uygulamalar için uygun bir seçenek olmuştur.

Sonuç olarak, algoritmaların farklı metriklerde güçlü yönlere sahip olduğu ve BinZOA ile optimize edilen özellik seçiminin bu performansı daha ileri taşıdığı gözlemlenmiştir. Bu bulgular, BinZOA'nın veri setindeki anlamlı özellikleri seçmede etkili olduğunu ve algoritmaların genel başarısını önemli ölçüde desteklediğini ortaya koymaktadır.

4.6 Derin Öğrenme Algoritmalarının Sonuçlarının Karşılaştırılması

KNN, RF, MLP, DT algoritmaları sonucunda oluşan dört algoritmanın en az ikisinin başarılı bulduğu özellikler çıkartılarak derin öğrenme algoritmalarına verilmiştir. Seçilen özellikler, zamansal ve yerel örüntüleri yakalama yetenekleri nedeniyle RNN, LSTM ve CNN modelleri ile değerlendirilmiştir. Modellerde gradient clipping, dropout ve uygun öğrenme oranları kullanılarak eğitim stabilitesi ve genelleme başarımı artırılmıştır. Aşağıdaki çizelgede seçilen en iyi özellikler verilmiştir.

Çizelge 4.13. Seçilen en iyi özellikler

Özellik
Fwd Pkt Len Std
Bwd Pkt Len Min
Fwd URG Flags
Pkt Len Min
RST Flag Cnt
URG Flag Cnt
Pkt Size Avg

Bwd Seg Size Avg
Init Bwd Win Byts
Fwd Act Data Pkts
Fwd Seg Size Min
Dst Port
Fwd PSH Flags
ACK Flag Cnt
Down/Up Ratio
Init Fwd Win Byts
Active Max
Flow IAT Std
Fwd IAT Std
Bwd Header Len
FIN Flag Cnt
Active Std
TotLen Bwd Pkts
Flow Byts/s
Flow IAT Min
Bwd IAT Min

Çizelge 4.13'te sunulan özellikler, ZOA tabanlı özellik seçimi süreci sonucunda sınıflandırma başarımına en yüksek katkıyı sağlayan ve ayırt ediciliği en yüksek olan özniteliklerdir. Bu özellikler, ağ trafiği verilerinin hem paket düzeyindeki istatistiksel yapısını hem de akış (flow) tabanlı zamansal davranışlarını yansıtan bileşenlerden oluşmaktadır. Seçilen özellikler incelendiğinde, özellikle paket uzunluğu (Fwd Pkt Len Std, Bwd Pkt Len Min, Pkt Len Min, Pkt Size Avg) ve segment boyutları (Fwd Seg Size Min, Bwd Seg Size Avg) ile ilişkili özniteliklerin ön plana çıktığı görülmektedir. Bu tür özellikler, normal ağ trafiği ile saldırı trafiği arasındaki yapısal farklılıkların belirlenmesinde kritik rol oynamaktadır. Saldırı senaryolarında, paket boyutları ve dağılımlarında meydana gelen düzensizlikler, bu özellikler aracılığıyla etkili biçimde yakalanabilmektedir. Bunun yanı sıra, TCP bayrakları ile ilişkili özelliklerin (URG Flag Cnt, RST Flag Cnt, ACK Flag Cnt, FIN Flag Cnt, Fwd URG Flags, Fwd PSH Flags) seçilmiş olması, saldırı tespitinde protokol davranışlarının ne denli önemli olduğunu göstermektedir. Özellikle servis reddi (DoS/DDoS), tarama (scan) ve bağlantı istismarı türündeki saldırılarda, TCP bayraklarının olağan dışı kullanım biçimleri ayırt edici bir gösterge sunmaktadır. Zamansal ve akış istatistikleri ile ilgili özellikler (Flow IAT Std, Flow IAT Min, Fwd IAT Std, Bwd IAT Min, Active Max, Active Std, Flow Byts/s) ise ağ trafiğinin zaman içindeki davranışını temsil etmektedir. Bu tür özellikler, saldırıların ani, yoğun veya düzensiz trafik üretme eğilimlerini ortaya koyarak özellikle zamana duyarlı derin öğrenme modelleri (LSTM, RNN) için önemli girdiler sağlamaktadır. Ayrıca, pencere boyutları ve veri

aktarım yoğunluğu ile ilişkili özelliklerin (Init Fwd Win Byts, Init Bwd Win Byts, Fwd Act Data Pkts, TotLen Bwd Pkts, Down/Up Ratio) seçilmesi, saldırıların veri aktarım karakteristiklerini yansıtan özniteliklerin de sınıflandırma sürecinde belirleyici olduğunu ortaya koymaktadır. Bu özellikler, özellikle botnet ve brute-force saldırıları gibi senaryolarda normal trafik ile saldırı trafiğinin ayrıştırılmasında etkili olmaktadır. Sonuç olarak, Çizelge 4.13'te yer alan en iyi özellikler; paket yapısı, protokol davranışı ve zamansal akış özelliklerini birlikte temsil ederek siber saldırıların çok boyutlu doğasını yansıtmaktadır. ZOA tabanlı özellik seçimi sayesinde, gereksiz ve ayırt edici olmayan özellikler elenmiş; sınıflandırma modellerinin daha hızlı, daha kararlı ve daha yüksek doğrulukla çalışmasına olanak tanıyan kompakt bir özellik kümesi elde edilmiştir. Bu durum, önerilen yaklaşımın siber saldırı tespit sistemleri için etkili ve uygulanabilir bir çözüm sunduğunu açıkça göstermektedir.

4.6.1 CNN Algoritmasının Perfonmans Değerlendirmesi

Bu bölümde, CNN algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. CNN, özellikle zaman serisi veya akış verilerindeki lokal örüntüleri tespit etme konusunda son derece başarılıdır. Konvolüsyon filtreleri sayesinde veri içindeki kritik özellikleri otomatik olarak öğrenebilir ve yüksek doğruluk oranlarına ulaşabilir.

CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, CNN modeli siber güvenlik saldırılarının sınıflandırılmasında kullanılmıştır. Model, veri akışı içerisindeki lokal ilişkileri değerlendirerek saldırı türlerini başarılı bir şekilde belirlemiştir. Performans analizinde Accuracy, Precision, Recall ve F1-Score metrikleri incelenmiş ve CNN modelinin güçlü bir sınıflandırma performansı sunduğu görülmüştür.

Çizelge 4.14. CNN Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9982
Kesinlik (Precision)	0,9982
Duyarlılık (Recall)	0,9982
F1 skoru (F1-Score)	0,9982
Süre (Time)	195,8696

- *İterasyon Sayısı (Iteration):* BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngüyü ifade eder.
- *Fitness Değeri (Fitness):* BinZOA'nın ilgili iterasyonda elde ettiği en iyi fitness değerini gösterir ve seçilen özelliklerin ne kadar etkili olduğunu temsil eder.

- *Seçilen Özellik Sayısı:* CNN modelinin eğitiminde kullanılan, BinZOA tarafından seçilmiş özelliklerin sayısını belirtir.
- *Süre (Time):* Her bir iterasyonda CNN modelinin eğitim ve test süresini ifade eder.
- *Accuracy:* İlgili iterasyonda CNN tarafından sınıflandırılan verinin doğruluk oranıdır.
- *Precision, Recall ve F1-Score:* CNN modelinin sınıflandırma performansını değerlendirmek için kullanılan tamamlayıcı metriklerdir.

CNN Mimarisi:

- *Input Layer:* Sequence verisi için uygun şekil
- *Conv1D Layer:* 32 filtre, kernel size 3, ReLU aktivasyon fonksiyonu
- *MaxPooling1D Layer:* Pool size 2
- *Dropout Layer 1:* 0.2 dropout oranı
- *Flatten Layer:* Conv katmanı çıktısını düzleştirme
- *Dense Layer 1:* 16 nöron, ReLU aktivasyon fonksiyonu
- *Dropout Layer 2:* 0.2 dropout oranı
- *Output Layer:* 1 nöron, sigmoid aktivasyon fonksiyonu

Eğitim Parametreleri:

- *Learning Rate:* 0.001
- *Optimizer:* Adam optimizer with gradient clipping (clipnorm=1.0)
- *Loss Function:* Binary cross-entropy
- *Batch Size:* 8 (veya veri boyutuna göre minimum)
- *Epochs:* 10 (tam veri seti), 5 (test veri seti)
- *Sequence Length:* 5 (veya veri boyutuna göre dinamik)

Seçim Gerekçesi:

- *32 filtre:* CNN'ler için yaygın bir filtre sayısıdır. Yeterli özellik haritası sağlarken hesaplama maliyetini kontrol altında tutmaktadır.
- *Kernel Size 3:* 1D convolution için optimal kernel boyutudur. Yerel desenleri yakalayabilirken hesaplama maliyeti düşüktür.
- *Pool Size 2:* Max pooling, özellik haritası boyutunu yarıya indirerek hesaplama maliyetini azaltmaktadır.

- *Sequence Length 5:* CNN'ler, convolution işlemi sayesinde daha uzun sequence'leri işleyebilmektedir. 5, yerel desenleri yakalamak için yeterli bir uzunluktur.
- *Batch Size 8:* CNN'ler, RNN/LSTM'lerden daha hızlı eğitilebildiğinden daha büyük batch size kullanılabilmektedir.

BinZOA ile bağımsız 20 farklı çalıştırma ve 4 farklı makine öğrenmesi sınıflandırıcısından elde edilen en iyi sonuçların değerlendirilmesi ile elde edilen özellikler; CNN modeli ile tekrar sınıflandırılmıştır. Özellikle doğruluk oranları ve süre arasındaki ilişki dikkatle incelenmiş ve elde edilen sonuçlar Çizelge 4.15'de özetlenmiştir.

Çizelge 4.15. CNN Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9995
Kesinlik (Precision)	0,9995
Duyarlılık (Recall)	0,9995
F1 skoru (F1-Score)	0,9995
Süre (Time)	156,8696

Elde edilen sonuçlar, CNN algoritmasının BinZOA ile optimize edilmiş ve makine öğrenmesi sınıflandırıcı algoritmaları ile çalıştırılmış ve seçilmiş en iyi özellikler ile çalıştırdıktan sonra yüksek performans gösteriği ortaya koymaktadır. Özellikle doğru özellik seçimi, CNN modelinin zaman bağımlı akış verilerini daha etkili analiz etmesini sağlamış ve sınıflandırma başarısını arttırmıştır.

4.6.2 RNN Algoritmasının Perfonmans Değerlendirmesi

Bu bölümde, RNN algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. RNN, özellikle zaman bağımlı veya ardışık verilerin analizinde güçlü sonuçlar sunabilen, yapay sinir ağları arasında önemli bir konuma sahiptir. Gelen ağlardan farklı olarak geçmiş bilgiyi belleğinde tutarak gelecekteki tahminlerde bu bilgiyi kullanabilmesi, RNN'nin tercih edilme nedenleri arasındadır. Günümüzde birçok sınıflandırma, metin analizi, zaman serisi tahmini ve siber güvenlik problemleri RNN modelleri ile etkili bir şekilde çözülebilmektedir.

Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, RNN modeli siber güvenlik saldırılarını tahmin etmek ve sınıflandırmak amacıyla uygulanmıştır. Algoritma, akış verilerindeki zaman bağımlı özellikleri değerlendirerek saldırı türlerini başarılı bir şekilde sınıflandırmıştır. Performans değerlendirmesi sırasında doğruluk oranı (Accuracy),

kesinlik (Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel metrikler incelenmiştir. Bu metrikler, RNN modelinin CSE-CIC-IDS2018 veri setinde genel olarak yüksek doğruluk oranı sunduğunu göstermektedir. Özellikle Accuracy değeri modelin genel sınıflandırma başarısını ifade ederken; Precision, Recall ve F1-Score, pozitif sınıfları doğru bir şekilde tahmin etme kapasitesini yansıtmaktadır.

Çizelge 4.16. RNN Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9925
Kesinlik (Precision)	0,9925
Duyarlılık (Recall)	0,9925
F1 skoru (F1-Score)	0,9925
Süre (Time)	219,8190

- *İterasyon Sayısı (Iteration):* BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngüyü ifade eder.
- *Fitness Değeri (Fitness):* BinZOA'nın ilgili iterasyonda elde ettiği en iyi fitness değerini gösterir ve seçilen özelliklerin ne kadar etkili olduğunu temsil eder.
- *Seçilen Özellik Sayısı:* RNN modelinin eğitiminde kullanılan, BinZOA tarafından seçilmiş özelliklerin sayısını belirtir.
- *Süre (Time):* Her bir iterasyonda RNN modelinin eğitim ve test süresini ifade eder.
- *Accuracy:* İlgili iterasyonda RNN tarafından sınıflandırılan verinin doğruluk oranıdır.
- *Precision, Recall ve F1-Score:* RNN modelinin sınıflandırma performansını değerlendirmek için kullanılan tamamlayıcı metriklerdir.

Mimari:

- *Input Layer:* Sequence verisi için uygun şekil
- *SimpleRNN Layer:* 8 nöron, tanh aktivasyon fonksiyonu, return_sequences=False
- *Dense Layer 1:* 4 nöron, ReLU aktivasyon fonksiyonu
- *Output Layer:* 1 nöron, sigmoid aktivasyon fonksiyonu (binary sınıflandırma)

Eğitim Parametreleri:

- *Learning Rate:* 0.01
- *Optimizer:* Adam optimizer with gradient clipping (clipnorm=1.0)
- *Loss Function:* Binary cross-entropy
- *Batch Size:* 4 (veya veri boyutuna göre minimum)

- *Epochs:* 10 (tam veri seti), 5 (test veri seti)
- *Sequence Length:* 3 (veya veri boyutuna göre dinamik)

Seçim Gerekçesi:

- *8 nöron:* RNN modelleri gradient explosion/vanishing problemine eğilimlidir. Küçük nöron sayısı bu riski azaltmaktadır.
- *Learning Rate 0.01:* RNN'ler için nispeten yüksek learning rate, daha hızlı yakınsama sağlamaktadır.
- *Gradient Clipping:* Gradient explosion'ı önlemek için kritik bir tekniktir.
- *Sequence Length 3:* Kısa sequence uzunluğu, hesaplama maliyetini azaltırken temporal bilgiyi korumaktadır.
- *Batch Size 4:* Küçük batch size, gradient hesaplamalarını daha stabil hale getirmektedir.

BinZOA ile bağımsız 20 farklı çalıştırma ve 4 farklı makine öğrenmesi sınıflandırıcısından elde edilen en iyi sonuçların değerlendirilmesi ile elde edilen özellikler; RNN modelinin performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk oranları ve süre arasındaki ilişki dikkatle incelenmiş ve elde edilen sonuçlar Çizelge 4.17'de özetlenmiştir.

Çizelge 4.17. RNN Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	1,0
Kesinlik (Precision)	1,0
Duyarlılık (Recall)	1,0
F1 skoru (F1-Score)	1,0
Süre (Time)	173,1298

Elde edilen sonuçlar, RNN algoritmasının BinZOA ile optimize edilmiş ve makine öğrenmesi sınıflandırıcı algoritmaları ile çalıştırılmış ve seçilmiş en iyi özellikler ile çalıştırdıktan sonra yüksek performans gösteriği ortaya koymaktadır. Özellikle doğru özellik seçimi, RNN modelinin zaman bağımlı akış verilerini daha etkili analiz etmesini sağlamış ve sınıflandırma başarısını artırmıştır.

4.6.3 LSTM Algoritmasının Perfonmans Değerlendirmesi

Bu bölümde, LSTM algoritmasının performans değerlendirmesi detaylı bir şekilde ele alınmıştır. LSTM, klasik RNN yapısının geliştirilmiş bir versiyonu olup, uzun süreli bağımlılık ilişkilerini öğrenme kapasitesi sayesinde ardışık veri analizinde oldukça başarılı bir modeldir. Zaman serisi, doğal dil işleme ve akış tabanlı siber güvenlik verilerinin analizinde yaygın olarak kullanılmaktadır. LSTM'nin geçmiş verileri unutma ve hatırlama mekanizmaları ile kontrol edebilmesi, onu karmaşık örüntüleri öğrenmede oldukça etkili bir hale getirmektedir.

Bu çalışmada, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilen deneylerde, LSTM modeli siber güvenlik saldırılarını tahmin etmek ve sınıflandırmak amacıyla uygulanmıştır. Model, veri setindeki zaman bağımlı özellikleri analiz ederek saldırı türlerini doğru bir şekilde sınıflandırmıştır. Performans değerlendirmesi sırasında doğruluk oranı (Accuracy), kesinlik (Precision), duyarlılık (Recall) ve F1 skoru (F1-Score) gibi temel metrikler incelenmiştir. Bu metrikler, LSTM modelinin genel olarak yüksek doğruluk oranı sunduğunu göstermektedir.

Çizelge 4.18 LSTM Ham Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9985
Kesinlik (Precision)	0,9985
Duyarlılık (Recall)	0,9985
F1 skoru (F1-Score)	0,9985
Süre (Time)	195,2316

- *İterasyon Sayısı (Iteration):* BinZOA tarafından optimize edilen özellik seçimi sürecindeki her bir döngüyü ifade eder.
- *Fitness Değeri (Fitness):* BinZOA'nın ilgili iterasyonda elde ettiği en iyi fitness değerini gösterir ve seçilen özelliklerin ne kadar etkili olduğunu temsil eder.
- *Seçilen Özellik Sayısı:* LSTM modelinin eğitiminde kullanılan, BinZOA tarafından seçilmiş özelliklerin sayısını belirtir.
- *Süre (Time):* Her bir iterasyonda LSTM modelinin eğitim ve test süresini ifade eder.
- *Accuracy:* İlgili iterasyonda LSTM tarafından sınıflandırılan verinin doğruluk oranıdır.
- *Precision, Recall ve F1-Score:* LSTM modelinin sınıflandırma performansını değerlendirmek için kullanılan tamamlayıcı metriklerdir.

Mimari:

- *Input Layer:* Sequence verisi için uygun şekil

- *LSTM Layer:* 32 nöron, tanh aktivasyon fonksiyonu, return_sequences=False
- *Dropout Layer 1:* 0.2 dropout oranı
- *Dense Layer 1:* 16 nöron, ReLU aktivasyon fonksiyonu
- *Dropout Layer 2:* 0.2 dropout oranı
- *Output Layer:* 1 nöron, sigmoid aktivasyon fonksiyonu

Eğitim Parametreleri:

- *Learning Rate:* 0.001
- *Optimizer:* Adam optimizer with gradient clipping (clipnorm=1.0)
- *Loss Function:* Binary cross-entropy
- *Batch Size:* 4 (veya veri boyutuna göre minimum)
- *Epochs:* 10 (tam veri seti), 5 (test veri seti)
- *Sequence Length:* 3 (veya veri boyutuna göre dinamik)

Seçim Gerekçesi:

- *32 nöron:* LSTM, RNN'den daha güçlü olduğundan daha fazla nöron kullanılabilmektedir. 32 nöron, kapasite ve hesaplama maliyeti arasında iyi bir dengedir.
- *Learning Rate 0.001:* LSTM'ler için standart learning rate değeridir. Daha düşük learning rate, daha stabil eğitim sağlamaktadır.
- *Dropout 0.2:* Overfitting'i önlemek için yaygın olarak kullanılan bir değerdir. %20 dropout, modelin genelleme yeteneğini artırmaktadır.
- *Sequence Length 3:* RNN ile aynı gerekçe, ancak LSTM daha uzun bağımlılıkları yakalayabilmektedir.

BinZOA ile bağımsız 20 farklı çalıştırma ve 4 farklı makine öğrenmesi sınıflandırıcısından elde edilen en iyi sonuçların değerlendirilmesi ile elde edilen özellikler; LSTM modelinin performansını detaylı bir şekilde değerlendirmek için kullanılmıştır. Özellikle doğruluk oranları ve süre arasındaki ilişki dikkatle incelenmiş ve elde edilen sonuçlar Çizelge 4.19'de özetlenmiştir.

Çizelge 4.19. LSTM Sonuçları

Parametre	Değerler
Doğruluk (Accuracy)	0,9994
Kesinlik (Precision)	0,9994
Duyarlılık (Recall)	0,9994
F1 skoru (F1-Score)	0,9994
Süre (Time)	145,4624

Elde edilen sonuçlar, LSTM algoritmasının BinZOA ile optimize edilmiş ve makine öğrenmesi sınıflandırıcı algoritmaları ile çalıştırılmış ve seçilmiş en iyi özellikler ile çalıştırdıktan sonra yüksek performans gösteriği ortaya koymaktadır. Özellikle doğru özellik seçimi, LSTM modelinin zaman bağımlı akış verilerini daha etkili analiz etmesini sağlamış ve sınıflandırma başarısını artırmıştır.

4.6.4 Derin Öğrenme Algoritmalarının Sonuçlarının Karşılaştırılması

LSTM, RNN ve CNN derin öğrenme modellerinin hem ham özellikler hem de ZOA tabanlı özellik seçimi sonrasında elde edilen performans sonuçları karşılaştırmalı olarak incelenmiştir. Çizelge 4.20'de sunulan sonuçlar; doğruluk, kesinlik, duyarlılık, F1-skoru ve işlem süresi metrikleri dikkate alınarak değerlendirilmiştir. Ham veri kullanılarak gerçekleştirilen deneylerde, LSTM ve CNN modellerinin yüksek doğruluk oranlarına ulaştığı, RNN modelinin ise diğer modellere kıyasla daha düşük bir performans sergilediği görülmektedir. Özellikle RNN (Ham) modelinin doğruluk ve F1-skoru değerlerinin, LSTM ve CNN modellerinin gerisinde kaldığı dikkat çekmektedir. Bununla birlikte, ham veri kullanımı tüm modeller için daha uzun işlem sürelerine neden olmuştur. ZOA ile optimize edilmiş özellikler kullanıldığında, tüm derin öğrenme modellerinde belirgin bir performans artışı gözlemlenmiştir. LSTM modeli, doğruluk değerini 0,9985'ten 0,9994'e yükseltirken, işlem süresinde de anlamlı bir azalma sağlamıştır. Benzer şekilde, CNN modeli hem doğruluk hem de F1-skoru açısından gelişim göstermiş ve ham veriye kıyasla daha kısa sürede sonuç üretmiştir. Bu durum, ZOA tabanlı özellik seçiminin, gereksiz ve ayırt edici olmayan özellikleri eleyerek derin öğrenme modellerinin daha verimli çalışmasına katkı sağladığını göstermektedir. En dikkat çekici sonuçlar RNN modelinde elde edilmiştir. Ham veri ile yapılan deneylerde görece düşük performans sergileyen RNN modeli, ZOA ile seçilmiş özellikler kullanıldığında %100 doğruluk, kesinlik, duyarlılık ve F1-skoru değerlerine ulaşmıştır. Ayrıca işlem süresinde de kayda değer bir düşüş gözlemlenmiştir. Bu sonuç, özellik seçiminin özellikle zamansal bağımlılıkları modelleyen RNN tabanlı yapılar üzerinde kritik bir etkiye sahip olduğunu ortaya koymaktadır. Genel olarak değerlendirildiğinde, ZOA tabanlı özellik seçimi yaklaşımı, tüm derin öğrenme modellerinde hem sınıflandırma başarımını artırmış hem de hesaplama maliyetlerini azaltmıştır. Elde edilen bulgular, derin öğrenme modellerinin siber saldırı tespiti gibi yüksek boyutlu ve karmaşık problemlerde, uygun bir özellik seçimi yöntemi ile birlikte kullanıldığında daha etkin ve uygulanabilir çözümler sunduğunu açıkça göstermektedir.

Çizelge 4.20. Çizelge Sonuçları Derin Öğrenme Algoritma Sonuçları

Algoritma	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Süre
LSTM(Ham)	0,9985	0,9985	0,9985	0,9985	195,2316
LSTM	0,9994	0,9994	0,9994	0,9994	145,4624
RNN(Ham)	0,9925	0,9925	0,9925	0,9925	219,8190
RNN	1.0	1.0	1.0	1.0	173.1298
CNN(Ham)	0,9982	0,9982	0,9982	0,9982	195,8696
CNN	0,9995	0,9995	0,9995	0,9995	156,8696

5. SONUÇLAR VE ÖNERİLER

Bu çalışmada, ZOA, veri madenciliğinde sınıflandırma problemlerinin çözümünde sıklıkla karşılaşılan bir ön işleme tekniği olan özellik seçimi problemi için kullanılmıştır. ZOA, literatürde genellikle sürekli optimizasyon problemlerinde başarıyla uygulanmış bir algoritma olmasına rağmen, transfer fonksiyon tabanlı yaklaşımlar kullanılarak ikili problemlerdeki performansı yeterince ele alınmamıştır. Bu çalışmada, ZOA'nın ikili arama uzayında transfer fonksiyonları kullanılarak özellik seçimi problemi üzerindeki başarısını göstermektedir. Bu çalışma kapsamında yalnızca bir çeşit S-şekilli (S-shaped) transfer fonksiyonu kullanılmıştır ve bu fonksiyon, sürekli arama uzayını ikili arama uzayına dönüştürmek için uygulanmıştır. Özellik seçimi problemi, büyük özellik setlerine sahip veri kümelerinde sınıflandırma sırasında ortaya çıkan uzun işlem sürelerini ve düşük doğruluk oranlarını iyileştirmek amacıyla en uygun alt özellik setinin oluşturulmasını hedefler. Bu çalışmada, CSE-CIC-IDS2018 veri seti, siber güvenlik alanında yaygın olarak kullanılan bir veri seti olarak seçilmiş ve ZOA'nın performansını değerlendirmek için kullanılmıştır. Bu veri seti, yüksek boyutlu özellikler ve farklı saldırı türlerini içermesi nedeniyle özellik seçimi problemleri için karmaşık ve gerçekçi bir test ortamı sunmaktadır.

Siber güvenlik bağlamında, ağ tabanlı saldırı tespit sistemlerinde kullanılan veri setleri genellikle yüksek boyutlu, gürültülü ve dengesiz sınıf dağılımlarına sahiptir. Bu durum, sınıflandırma algoritmalarının hem hesaplama maliyetini artırmakta hem de genelleme performansını olumsuz etkilemektedir. Özellik seçimi, bu tür sistemlerde gereksiz ve ayırt edici olmayan özelliklerin elenmesini sağlayarak saldırı tespit modellerinin daha hızlı ve daha doğru çalışmasına katkı sunmaktadır. Bu nedenle, ZOA tabanlı özellik seçimi yaklaşımı, siber güvenlik uygulamalarında saldırı tespit performansını artırmaya yönelik etkili bir çözüm olarak değerlendirilmektedir.

Bu çalışmada doğruluğa etki eden en uygun özellikler ikili ZOA algoritması (BinZOA) ile belirlenmiştir ve bir fitness fonksiyonu yardımıyla performans ölçümleri gerçekleştirilmiştir. Kullanılan fitness fonksiyonunda yer alan doğruluk parametresi dört farklı makine öğrenimi algoritmaları ile tespit edilmiştir. Bu makine öğrenimi algoritmaları KNN,DT,RF ve MLP'dir. Bu sınıflandırıcılar yardımıyla BinZOA algoritması 26 özellik sayısına sahip bir alt veri kümesi belirlemiştir. Bu 26 özellik sayısına sahip alt veri kümesi üzerinde RNN, LSTM, CNN algoritmaları kullanılarak karşılaştırmalı analizler yapılmaktadır. Test sonuçlarına göre, S transfer fonksiyonu, ZOA'nın hem fitness değerlerini iyileştirmede hem de sınıflandırma doğruluğunu artırmada etkili olduğunu göstermiştir. En iyi fitness sonuçları, ZOA'nın veri

setindeki en anlamlı özellikleri başarılı bir şekilde seçtiğini ortaya koymaktadır. Ayrıca, S-şekilli transfer fonksiyonu kullanılarak yapılan optimizasyonlarda elde edilen sınıflandırma doğruluğu, büyük veri kümelerinde bile işlem sürelerini kısaltarak algoritmanın etkinliğini artırmıştır. Bu tez çalışması, ZOA'nın sınıflandırma görevlerinde bir ön işlem tekniği uygulayıcısı olarak özellik seçimi için güçlü bir optimizasyon aracı olduğunu göstermiştir.

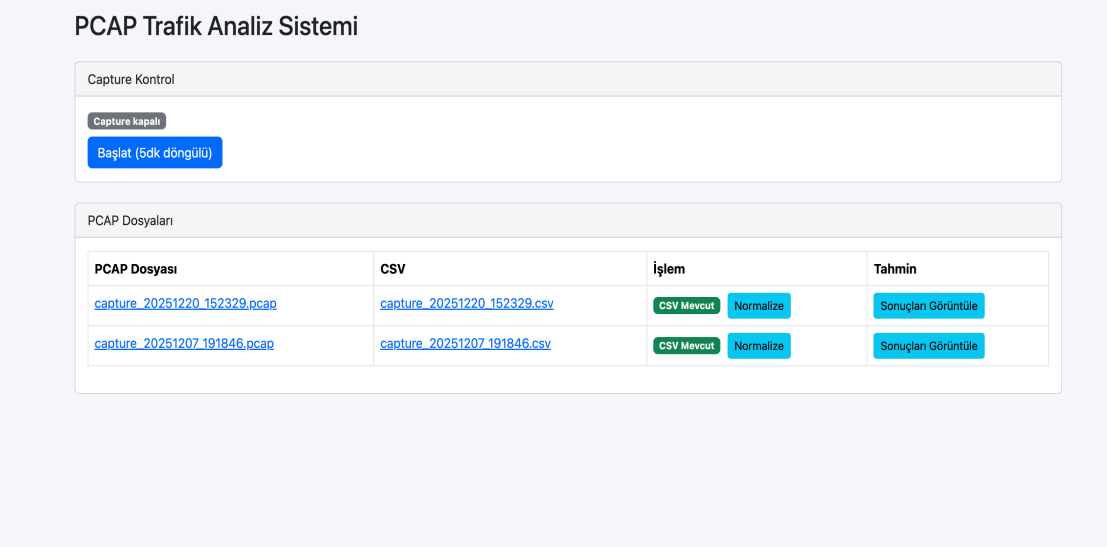
5.1 Tez Çalışmasının Uygulanabilirliği

Bu tez çalışması kapsamında önerilen yöntem, yalnızca teorik ve deneysel analizlerle sınırlı kalmayıp, gerçek zamanlı olarak çalışabilen bütünleşik bir siber saldırı tespit sistemi hâline getirilmiştir. Çalışmada kullanılan tüm makine öğrenmesi ve derin öğrenme modelleri; KNN, MLP, RF, DT, RNN, LSTM, CNN anlık ağ trafiği üzerinde çalışabilecek şekilde bir web tabanlı uygulama içerisinde entegre edilmiştir. Geliştirilen web sitesi, ağ trafiğini (örneğin 80 portu) belirlenen zaman aralıklarına bağlı olarak (örneğin 5 dakikalık döngüler hâlinde) otomatik biçimde yakalayabilmektedir. Bu süreçte, Wireshark aracı kullanılarak ilgili zaman dilimine ait ağ trafiği PCAP formatında toplanmakta; ardından elde edilen ham trafik verileri, CSE-CIC-IDS2018 veri seti formatı ile uyumlu olacak şekilde ön işleme adımlarından geçirilerek özellik çıkarımı yapılmaktadır. Bu sayede, gerçek ağ trafiği ile eğitim sürecinde kullanılan veri yapısı arasında tutarlılık sağlanmaktadır.

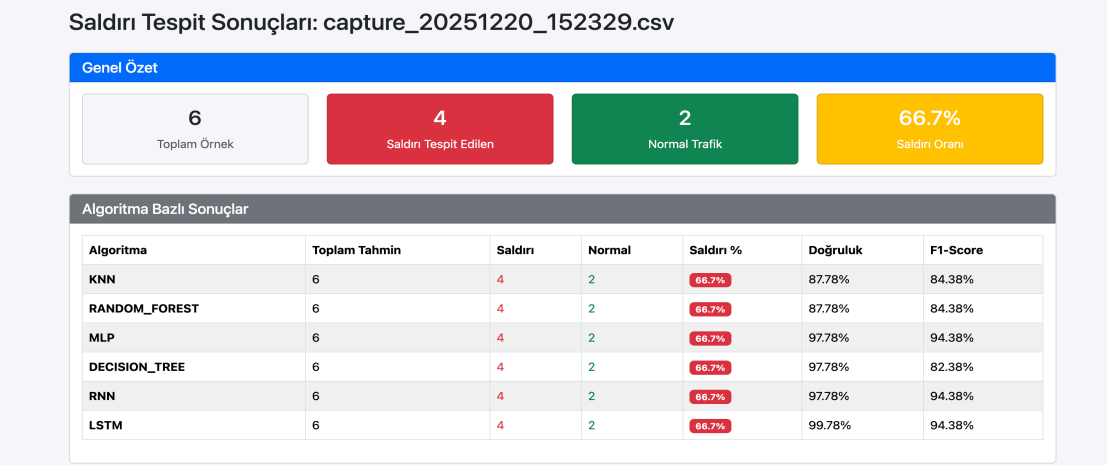
Ön işleme ve özellik çıkarımı tamamlanan veriler, tez kapsamında eğitilmiş ve optimize edilmiş yapay zekâ modellerine girdi olarak verilmekte; her bir model, gelen trafiği "normal" veya "saldırı" olarak sınıflandırmaktadır. Web tabanlı arayüz üzerinden, her bir algoritmanın ürettiği tahminler ayrı ayrı görüntülenebilmekte; aynı zamanda toplam örnek sayısı, tespit edilen saldırı sayısı ve saldırı oranı gibi özet bilgiler kullanıcıya sunulmaktadır. Bu yapı, farklı modellerin gerçek zamanlı performanslarının karşılaştırılmasına da olanak tanımaktadır.

Geliştirilen sistemin önemli avantajlarından biri, ZOA tabanlı özellik seçimi ile belirlenen en iyi özelliklerin kullanılması sayesinde, hem işlem süresinin azaltılması hem de sınıflandırma doğruluğunun yüksek tutulmasıdır. Bu durum, özellikle gerçek zamanlı çalışması gereken siber güvenlik uygulamaları için kritik bir gereksinimi karşılamaktadır. Ayrıca sistem, statik bir veri kümesiyle sınırlı kalmayıp, sürekli akan ağ trafiği üzerinde çalışabildiği için dinamik saldırı senaryolarına uyum sağlayabilen bir yapı sunmaktadır. Sonuç olarak, bu tez çalışmasında geliştirilen web tabanlı saldırı tespit sistemi; kullanılan algoritmaların gerçek hayatta uygulanabilirliğini ortaya koymakta ve önerilen ZOA tabanlı yaklaşımın yalnızca akademik düzeyde değil, pratik siber güvenlik uygulamalarında da etkin bir şekilde

kullanılabileceğini göstermektedir. Bu yönüyle çalışma, gerçek zamanlı saldırı tespit sistemlerinin geliştirilmesine yönelik somut ve uygulanabilir bir çözüm sunmaktadır.



Şekil 5.1. PCAP Trafik Analiz Sistemi Ana Arayüzü



Şekil 5.2. Gerçek Zamanlı Saldırı Tespit Sonuçları Ekranı

Şekil 5.1 ve Şekil 5.2'de, tez kapsamında geliştirilen web tabanlı gerçek zamanlı siber saldırı tespit sisteminin arayüzü ve elde edilen sonuçlar sunulmaktadır. İlk olarak, PCAP Trafik Analiz Sistemi arayüzü üzerinden ağ trafiği belirlenen zaman aralıklarında (örneğin 5 dakikalık döngüler hâlinde) otomatik olarak yakalanmakta ve elde edilen PCAP dosyaları sistem tarafından listelenmektedir. Bu dosyalar, CSE-CIC-IDS2018 veri seti formatı ile uyumlu olacak şekilde ön işleme adımlarından geçirilerek CSV formatına dönüştürülmekte ve normalize edilmektedir. Ardından, işlenen veriler k-NN, RF, MLP, DT, RNN ve LSTM modellerine girdi olarak verilerek saldırı tespiti gerçekleştirilmektedir.

Elde edilen sonuçlarda Şekil 5.2'de görüldüğü üzere, toplam 6 ağ trafiği örneğinin 4'ünün saldırı, 2'sinin ise normal trafik olarak sınıflandırıldığı ve genel saldırı oranının %66,7 olduğu görülmektedir. Algoritma bazlı değerlendirmelerde, özellikle LSTM modelinin %99,78 doğruluk ve %94,38 F1-skoru ile en yüksek performansı sunduğu dikkat çekmektedir. Benzer şekilde MLP ve RNN modelleri de yüksek doğruluk ve F1-skoru değerleriyle saldırı tespitinde başarılı sonuçlar üretmiştir. Bu bulgular, ZOA tabanlı özellik seçimi ile optimize edilen modellerin yalnızca çevrimdışı deneylerde değil, gerçek zamanlı ağ trafiği üzerinde de etkin ve uygulanabilir bir saldırı tespit çözümü sunduğunu açıkça ortaya koymaktadır.

5.2 Tez Üzerine Öneriler

Bu tez çalışmasında, ZOA ve ikili versiyonu olan BinZOA kullanılarak siber saldırı tespitinde özellik seçimi problemi ele alınmış ve farklı makine öğrenmesi ile derin öğrenme modelleri üzerinden kapsamlı bir değerlendirme gerçekleştirilmiştir. Elde edilen sonuçlar, önerilen yaklaşımın etkinliğini ortaya koymakla birlikte, çalışmanın bazı sınırlılıkları ve gelecekte geliştirilebilecek yönleri de bulunmaktadır. Öncelikle, bu çalışmada ZOA'nın ikili arama uzayına uyarlanmasında yalnızca S-şekilli (S-shaped) tek bir transfer fonksiyonu kullanılmıştır. Gelecek çalışmalarda, V-şekilli (V-shaped) veya farklı olasılıksal transfer fonksiyonlarının kullanılması, BinZOA'nın özellik seçimi performansının daha kapsamlı biçimde analiz edilmesine olanak sağlayabilir. Ayrıca, farklı transfer fonksiyonlarının sınıflandırma başarımı ve hesaplama maliyeti üzerindeki etkilerinin karşılaştırmalı olarak incelenmesi, algoritmanın davranışını daha iyi anlamaya katkı sunacaktır. Bu tez kapsamında deneyler, CSE-CIC-IDS2018 veri seti üzerinde gerçekleştirilmiştir. Her ne kadar bu veri seti siber güvenlik alanında yaygın olarak kullanılsa da, farklı veri setleri (örneğin UNSW-NB15, NSL-KDD veya gerçek ağ trafiğinden elde edilen veriler) üzerinde yapılacak çalışmalar, önerilen yöntemin genellenebilirliğini ve farklı saldırı senaryolarındaki başarımını daha net biçimde ortaya koyacaktır. Ayrıca, veri setlerindeki sınıf dengesizliğinin giderilmesine yönelik ek stratejilerin değerlendirilmesi de gelecek çalışmalar için önemli bir araştırma alanı oluşturmaktadır. Çalışmada, BinZOA ile seçilen özellikler; KNN, MLP, RF ve gibi klasik algoritmaların yanı sıra LSTM, RNN ve CNN gibi derin öğrenme modelleri kullanılarak değerlendirilmiştir. Ancak, daha güncel ve karmaşık derin öğrenme mimarilerinin (örneğin hibrit CNN-LSTM yapıları veya dikkat mekanizmaları içeren modeller) kullanılması, saldırı tespit performansının daha da artırılmasına katkı sağlayabilir. Bununla birlikte, geliştirilen sistem gerçek zamanlı saldırı tespit senaryolarına uygun bir yapı sunsa da, gerçek ağ ortamlarında uzun süreli ve yüksek hacimli trafik üzerinde yapılacak ek testler, sistemin

ölçeklenebilirliği ve kararlılığı hakkında daha kapsamlı bilgiler sağlayacaktır. Özellikle farklı ağ topolojileri ve değişen saldırı profilleri altında sistemin davranışının incelenmesi, uygulama açısından önemli bir katkı sunacaktır. Son olarak, bu çalışmada ele alınan optimizasyon yaklaşımı ağırlıklı olarak özellik seçimi problemi üzerine odaklanmıştır.

Gelecek çalışmalarda, ZOA ve benzeri meta-sezgisel algoritmaların hiperparametre optimizasyonu, model mimarisi seçimi veya çok amaçlı optimizasyon gibi farklı siber güvenlik problemlerine uygulanması, algoritmanın kullanım alanlarını genişletebilir. Özetle, bu tez çalışması siber saldırı tespitinde ZOA tabanlı özellik seçimi yaklaşımının etkinliğini ortaya koymakla birlikte, önerilen yöntem farklı veri setleri, transfer fonksiyonları, derin öğrenme mimarileri ve gerçek zamanlı senaryolar ile desteklenerek daha kapsamlı hâle getirilebilir.

KAYNAKLAR

Abd Elaziz, M., Al-qaness, M. A. A., Dahou, A., Ibrahim, R. A., & El-Latif, A. A. A. (2023). Intrusion detection approach for cloud and IoT environments using deep learning and Capuchin Search Algorithm. *Advances in Engineering Software*, *176*. https://doi.org/10.1016/j.advengsoft.2022.103402

Abdel-Basset, M., Abdel-Fatah, L., … A. S. data on the cloud with, & 2018, undefined. (2018). Metaheuristic algorithms: A comprehensive review. *Elsevier*, 185-231. https://doi.org/10.1016/B978-0-12-813314-9.00010-4

Abdelmalek, F., Afghoul, H., Krim, F., Zabia, D. E., Trabelsi, H., Bajaj, M., Zaitsev, I., & Blazek, V. (2024a). Experimental validation of effective zebra optimization algorithm-based MPPT under partial shading conditions in photovoltaic systems. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-77488-2

Abdelmalek, F., Afghoul, H., Krim, F., Zabia, D. E., Trabelsi, H., Bajaj, M., Zaitsev, I., & Blazek, V. (2024b). Experimental validation of effective zebra optimization algorithm-based MPPT under partial shading conditions in photovoltaic systems. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-77488-2

Abdi, N., Albaseer, A., & Abdallah, M. (2024). *The Role of Deep Learning in Advancing Proactive Cybersecurity Measures for Smart Grid Networks: A Survey*. http://arxiv.org/abs/2401.05896

Abdullah-Al-Kafi, M., Tasnova, I. J., Wadud Islam, M., & Banshal, S. K. (2022). Performances of different approaches for fake news classification: an analytical study. *SpringerM Abdullah-Al-Kafi, IJ Tasnova, M Wadud Islam, SK BanshalInternational Conference on Advanced Network Technologies and Intelligent, 2021•Springer*, *1534 CCIS*, 700-714. https://doi.org/10.1007/978-3-030-96040-7_53

Abu Bakar, R., Paolucci, F., Cugini, F., Jose, J., Olmos, V., & De Marinis, L. (2025). *IDS-NGNN: A SmartNIC-based Intrusion Detection System Based on Reduce and Merge Nested Graph Neural Networks*.

Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection. *IEEE Access*, *6*, 33789-33795. https://doi.org/10.1109/ACCESS.2018.2841987

Al Razib, M., Javeed, D., Khan, M. T., Alkanhel, R., & Muthanna, M. S. A. (2022). Cyber Threats Detection in Smart Environments Using SDN-Enabled DNN-LSTM Hybrid Framework. *IEEE Access*, *10*, 53015-53026. https://doi.org/10.1109/ACCESS.2022.3172304

Ali, A. (2026). A Hybrid Grey Wolf Optimizer-Zebra Optimization Algorithm for Solving Optimization Problems. *Jambura J. Math*, *8*(1), 7-25. https://doi.org/10.37905/jjom.v8i1.34499

Alsharif, M., & Rawat, D. B. (2023). Machine Learning Enabled Intrusion Detection for Edge Devices in the Internet of Things. *2023 IEEE 13th Annual Computing and Communication Workshop and Conference, CCWC 2023*, 361-367. https://doi.org/10.1109/CCWC57344.2023.10099276

Amin, R., El-Taweel, G., Ali, A. F., & Tahoun, M. (2024). Hybrid Chaotic Zebra Optimization Algorithm and Long Short-Term Memory for Cyber Threats Detection. *IEEE Access*, *12*, 93235-93260. https://doi.org/10.1109/ACCESS.2024.3397303

Amin, R., El-Taweel, G., Fouad Ali, A., & Tahoun, M. (2023). A Hybrid Extreme Gradient Boosting and Long Short-Term Memory Algorithm for Cyber Threats Detection.

MENDEL-Soft Computing Journal, *29*, 2571-3701. https://doi.org/10.13164/mendel.2023..307

Appiahene, P., Berchie, S. O., Botchway, E., Ayitey, M. J., Dawson, J. K., Mettle, H. N. A., & Afrifa, S. (2026a). Network intrusion detection using a hybrid graph-based convolutional network and transformer architecture. *PLOS ONE*, *21*(1 January). https://doi.org/10.1371/journal.pone.0340997

Appiahene, P., Berchie, S. O., Botchway, E., Ayitey, M. J., Dawson, J. K., Mettle, H. N. A., & Afrifa, S. (2026b). Network intrusion detection using a hybrid graph-based convolutional network and transformer architecture. *journals.plos.orgP Appiahene, SO Berchie, E Botchway, MJ Ayitey, JK Dawson, HNA Mettle, S AfrifaPloS one, 2026•journals.plos.org*, *21*(1 January). https://doi.org/10.1371/journal.pone.0340997

Arkwazee, A., & Abttan, R. (2025). Enhanced zebra optimization algorithm for sustainable combined economic and emission dispatch in power systems. *Array*, *27*. https://doi.org/10.1016/j.array.2025.100484

Atefinia, R., & Ahmadi, M. (2021). Network intrusion detection using multi-architectural modular deep neural network. *Journal of Supercomputing*, *77*(4), 3571-3593. https://doi.org/10.1007/s11227-020-03410-y

Ayantayo, A., Kaur, A., Kour, A., Schmoor, X., Shah, F., Vickers, I., Kearney, P., & Abdelsamea, M. M. (2023). Network intrusion detection using feature fusion with deep learning. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00834-0

Ayyadevara, V. K. (2018). Gradient boosting machine. *SpringerVK AyyadevaraPro machine learning algorithms: A hands-on approach to implementing, 2018•Springer*, 117-134. https://doi.org/10.1007/978-1-4842-3564-5_6

Bacevicius, M., Sutiene, K., Malakauskas, L., & Paulauskaite-Taraseviciene, A. (2026a). CoLIME with 2D Copulas for Reliable Local Explanations on Imbalanced Network Data. *Applied Sciences (Switzerland)*, *16*(1). https://doi.org/10.3390/app16010119

Bacevicius, M., Sutiene, K., Malakauskas, L., & Paulauskaite-Taraseviciene, A. (2026b). CoLIME with 2D Copulas for Reliable Local Explanations on Imbalanced Network Data. *Applied Sciences (Switzerland)*, *16*(1). https://doi.org/10.3390/app16010119

Bağımsız, D., Yansımalarının, D., Güvenlik, S., Değerlendirilmesi, Y., Yel, T., & Atasoy, A. (t.y.). Dijitalleşmenin bağımsız denetime yansımalarının siber güvenlik yönünden değerlendirilmesi. *dergipark.org.trT Yel, A AtasoyMuhasebe ve Finansman Dergisi, 2021•dergipark.org.tr*. Geliş tarihi 20 Şubat 2026, gönderen https://dergipark.org.tr/en/download/article-file/1924180

Baig, M. M., Awais, M. M., & El-Alfy, E. S. M. (2017). AdaBoost-based artificial neural network learning. *Neurocomputing*, *248*, 120-126. https://doi.org/10.1016/j.neucom.2017.02.077

Balci, Y., & Duman, S. (2025). Modified Zebra Optimization Algorithm via Design Operators for Modern Power System Planning Problem Solution. *IEEE Access*, *13*, 107343-107376. https://doi.org/10.1109/ACCESS.2025.3581676

Basnet, R., Shash, R., … C. J.-J. I. Serv. I., & 2019, undefined. (2019). Towards Detecting and Classifying Network Intrusion Traffic Using Deep Learning Frameworks. *researchgate.netRB Basnet, R Shash, C Johnson, L Walgren, T DoleckJ. Internet Serv. Inf. Secur., 2019•researchgate.net*. https://www.researchgate.net/profile/Ram-Basnet/publication/337936440_Towards_Detecting_and_Classifying_Network_Intrusion_Traffic_Using_Deep_Learning_Frameworks/links/5df5c7b992851c83647e7dc1/Towards-Detecting-and-Classifying-Network-Intrusion-Traffic-Using-Deep-Learning-Frameworks.pdf

Bilgic, B., Chatnuntawech, I., Fan, A. P., Setsompop, K., Cauley, S. F., Wald, L. L., & Adalsteinsson, E. (2014). Fast image reconstruction with L2-regularization. *Journal of Magnetic Resonance Imaging*, *40*(1), 181-191. https://doi.org/10.1002/jmri.24365

Cao, L., & Wei, Q. (2025). SZOA: An Improved Synergistic Zebra Optimization Algorithm for Microgrid Scheduling and Management. *Biomimetics*, *10*(10). https://doi.org/10.3390/biomimetics10100664

Catillo, M., Rak, M., & Villano, U. (2020). 2L-ZED-IDS: A two-level anomaly detector for multiple attack classes. *SpringerM Catillo, M Rak, U VillanoWorkshops of the international conference on advanced information networking, 2020•Springer*, *1150 AISC*, 687-696. https://doi.org/10.1007/978-3-030-44038-1_63

Chadza, T., … K. K.-2019 17th I., & 2019, undefined. (2019). Contemporary sequential network attacks prediction using hidden Markov model. *ieeexplore.ieee.orgT Chadza, KG Kyriakopoulos, S Lambotharan2019 17th International Conference on Privacy, Security and Trust, 2019•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8949035/

Chan, L. H., Law, N. F., & Siu, W. C. (2013). A confidence map and pixel-based weighted correlation for PRNU-based camera identification. *Digital Investigation*, *10*(3), 215-225. https://doi.org/10.1016/j.diin.2013.04.001

Chao, K. H., Huang, K. H., & Guo, Y. H. (2025). Design of a Brushless DC Motor Drive System Controller Integrating the Zebra Optimization Algorithm and Sliding Mode Theory. *Electronics (Switzerland)*, *14*(17). https://doi.org/10.3390/electronics14173353

Chaudhary, A., Kolhe, S., agriculture, R. K.-I. processing in, & 2016, undefined. (2016). An improved random forest classifier for multi-class classification. *ElsevierA Chaudhary, S Kolhe, R KamalInformation processing in agriculture, 2016•Elsevier*. https://www.sciencedirect.com/science/article/pii/S2214317316300099

Chen, J., Xie, B., Zhang, H., & Zhai, J. (2019). Deep autoencoders in pattern recognition: a survey. *World ScientificJ Chen, B Xie, H Zhang, J ZhaiBio-inspired computing models and algorithms, 2019•World Scientific*, 229-255. https://doi.org/10.1142/9789813143180_0009

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-August-2016*, 785-794. https://doi.org/10.1145/2939672.2939785

Chen, W., Zhang, S., Li, R., & Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling. *Science of the Total Environment*, *644*, 1006-1018. https://doi.org/10.1016/j.scitotenv.2018.06.389

Chi, W. (2025a). Multidimensional covert traffic attack detection via coupled spatio-temporal transformer and causal convolutional networks. *International Journal of Reasoning-based Intelligent Systems*, *17*(10). https://doi.org/10.1504/ijris.2025.10074693

Chi, W. (2025b). Multidimensional covert traffic attack detection via coupled spatio-temporal transformer and causal convolutional networks. *International Journal of Reasoning-based Intelligent Systems*, *17*(10). https://doi.org/10.1504/ijris.2025.10074693

Chorowski, J., & Bahdanau, D. (2015). *Attention-Based Models for Speech Recognition*.

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *26-June-2016*, 2201-2206. https://doi.org/10.1145/2882903.2912574

CNBC. (2019, October 13). Cyberattacks now cost companies $200,000 on average, putting many out of business. . (t.y.). Geliş tarihi 20 Şubat 2026, gönderen https://www.cnbc.com/2019/10/13/cyberattacks-cost-small-companies-200k-putting-many-out-of-business.html

Coşar, H., Arısoy, Ç., and, H. U.-A. I. T., & 2024, undefined. (2024). Intrusion detection on CSE-CIC-IDS2018 dataset using machine learning methods. *dergipark.org.trHİ Coşar, Ç Arısoy, H UlutaşArtificial Intelligence Theory and Applications, 2024•dergipark.org.tr*. https://dergipark.org.tr/en/pub/aita/issue/87553/1553769

Demirci, H. (2023). *Elektriğin dirençli ortamda hareketini temel alan yeni bir meta sezgisel algoritma tasarımı= Design of a new metaheuristic algorithm based on the movement of*. https://acikerisim.sakarya.edu.tr/handle/20.500.12619/101461

Devnath, M. K. (2023). *GCNIDS: Graph Convolutional Network-Based Intrusion Detection System for CAN Bus*. http://arxiv.org/abs/2309.10173

Devprabhakar, S. M., Parameswarappa, M., & Naik, A. K. J. (2025). Energy-aware multi-objective triple strategy-based zebra optimization algorithm for clustering and routing in wireless sensor network. Içinde *Iran Journal of Computer Science* (C. 8, Sayı 4, ss. 2965-2986). Springer International Publishing. https://doi.org/10.1007/s42044-025-00314-z

D'hooge, L., Wauters, T., Volckaert, B., & De Turck, F. (2019). In-depth comparative evaluation of supervised machine learning approaches for detection of cybersecurity threats. *IoTBDS 2019 - Proceedings of the 4th International Conference on Internet of Things, Big Data and Security*, 125-136. https://doi.org/10.5220/0007724801250136

El-Hasnony, I. M., Barakat, S. I., Elhoseny, M., & Mostafa, R. R. (2020). Improved Feature Selection Model for Big Data Analytics. *IEEE Access*, *8*, 66989-67004. https://doi.org/10.1109/ACCESS.2020.2986232

Elymany, M. M., Enany, M. A., & Elsonbaty, N. A. (2024). Hybrid optimized-ANFIS based MPPT for hybrid microgrid using zebra optimization algorithm and artificial gorilla troops optimizer. *Energy Conversion and Management*, *299*. https://doi.org/10.1016/j.enconman.2023.117809

Ervural, B., & Hakli, B. (2023). A binary reptile search algorithm based on transfer functions with a new stochastic repair method for 0–1 knapsack problems. *ElsevierB Ervural, H HakliComputers & Industrial Engineering, 2023•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0360835223001043

Faheem, M., & Al-Khasawneh, M. A. (2024). Multilayer cyberattacks identification and classification using machine learning in internet of blockchain (IoBC)-based energy networks. *Data in Brief*, *54*. https://doi.org/10.1016/j.dib.2024.110461

Faizan, M., & Mohamad, M. A. (2026). *Zebra Optimization Algorithm for Feature Selection*. 12-17. https://doi.org/10.1109/icsecs65227.2025.11278999

Faria, V. da S., Gonçalves, J. A., da Silva, C. A. M., Vieira, G. de B., & Mascarenhas, D. M. (2020). Sdtow: A slowloris detecting tool for WMNS. *Information (Switzerland)*, *11*(12), 1-18. https://doi.org/10.3390/info11120544

Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. Içinde *Journal of Artificial Intelligence Research* (C. 61).

Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*. https://doi.org/10.1016/j.jisa.2019.102419

Fitni, Q. R. S. , & Ramli, K. (2020). Implementation of ensemble learning and feature selection for performance improvements in anomaly-based intrusion detection systems. *ieeexplore.ieee.orgQRS Fitni, K Ramli2020 IEEE International Conference on Industry 4.0, Artificial, 2020•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/9172014/

Gamage, S., & Samarabandu, J. (2020). Deep learning methods in network intrusion detection: A survey and an objective comparison. *ElsevierS Gamage, J*

SamarabanduJournal of Network and Computer Applications, 2020•Elsevier. https://www.sciencedirect.com/science/article/pii/S1084804520302411

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. https://link.springer.com/content/pdf/10.1007/978-3-319-10247-4.pdf

Geem, Z. W., Kim, J. H., & Loganathan, G. V. (2001). A new heuristic optimization algorithm: harmony search. *journals.sagepub.comZW Geem, JH Kim, GV Loganathansimulation, 2001•journals.sagepub.com*, *76*(2), 60-68. https://doi.org/10.1177/003754970107600201

George, G., & Thampi, S. M. (2018). A Graph-Based Security Framework for Securing Industrial IoT Networks from Vulnerability Exploitations. *IEEE Access*, *6*, 43586-43601. https://doi.org/10.1109/ACCESS.2018.2863244

Ghadi, Y., Neamah, N., & Hossam-Eldin, N. (2023). State-of-the-Art Frequency Control Strategy Based on an Optimal Fuzzy PI-FOPDFλ for SMES and UPFC Integrated Smart Grids Using Zebra Optimization Algorithm. *ieeexplore.ieee.orgYY Ghadi, NM Neamah, AA Hossam-Eldin, M Alqarni, KM AboRasIEEE Access, 2023•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/10302284/

Göcs, L., & Johanyák, Z. C. (2023). Feature Selection with Weighted Ensemble Ranking for Improved Classification Performance on the CSE-CIC-IDS2018 Dataset. *Computers*, *12*(8). https://doi.org/10.3390/computers12080147

Groff Z, & Schwartz S. (2019). Data preprocessing and feature selection for an intrusion detection system dataset. *engr.ship.eduZ Groff, S Schwartz34th Annual Conference of The Pennsylvania Association of Computer and, 2019•engr.ship.edu*. https://www.engr.ship.edu/downloads/pacise/pacise-proceedings-2019.pdf#page=103

Gupta, & Bhavsar. (2017). Random forest-based feature importance for hep-2 cell image classification. *SpringerV Gupta, A BhavsarAnnual Conference on Medical Image Understanding and Analysis, 2017•Springer*, *723*, 922-934. https://doi.org/10.1007/978-3-319-60964-5_80

Gupta, S., Kumar, S., Singh, K., Chastikova, V. A., & Sotnikov, V. V. (2019). Method of analyzing computer traffic based on recurrent neural networks. *iopscience.iop.orgVA Chastikova, VV SotnikovJournal of Physics: Conference Series, 2019•iopscience.iop.org*, *1353*, 12133. https://doi.org/10.1088/1742-6596/1353/1/012133

Habibipour, M., Kamel Tabbakh, S. R., Naeen, H. M., & Moattar, M. H. (2026a). ATS-IDS: an adaptive teacher-student framework for lightweight IDS in IoT using incremental learning. *Cluster Computing*, *29*(2). https://doi.org/10.1007/s10586-025-05878-w

Habibipour, M., Kamel Tabbakh, S. R., Naeen, H. M., & Moattar, M. H. (2026b). ATS-IDS: an adaptive teacher-student framework for lightweight IDS in IoT using incremental learning. *SpringerM Habibipour, SR Kamel Tabbakh, HM Naeen, MH MoattarCluster Computing, 2026•Springer*, *29*(2). https://doi.org/10.1007/s10586-025-05878-w

Hamilton, R., Gray, W., Sibanda, C., Kandasamy, S., Kirner, R., & Tsokanos, A. (2020, Kasım 24). Deep packet inspection in firewall clusters. *2020 28th Telecommunications Forum, TELFOR 2020 - Proceedings*. https://doi.org/10.1109/TELFOR51502.2020.9306651

Happel, B., networks, J. M.-N., & 1994, undefined. (1994). Design and evolution of modular neural network architectures. *ElsevierBLM Happel, JMJ MurreNeural networks, 1994•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0893608005801558

Hart, W. (1994). *Adaptive global optimization with local search*. https://search.proquest.com/openview/d9f98abd9e41ae4bc492e5a5f9e73694/1?pq-origsite=gscholar&cbl=18750&diss=y

Hashemi, A., Dowlatshahi, M. B., & Nezamabadi-pour, H. (2022). Ensemble of feature selection algorithms: a multi-criteria decision-making approach. *SpringerA Hashemi, MB*

Dowlatshahi, H Nezamabadi-pourInternational Journal of Machine Learning and Cybernetics, 2022•Springer, *13*(1), 49-69. https://doi.org/10.1007/s13042-021-01347-z

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, *97*, 849-872. https://doi.org/10.1016/j.future.2019.02.028

Holland, J. H. (1992). *Genetic Algorithms*. *267*(1), 66-73. https://doi.org/10.2307/24939139

Hoque, N., Singh, M., & Bhattacharyya, D. K. (2018). EFS-MI: an ensemble feature selection method for classification. *Complex & Intelligent Systems*, *4*(2), 105-118. https://doi.org/10.1007/s40747-017-0060-x

Hua, Y. (2020). An efficient traffic classification scheme using embedded feature selection and lightgbm. *ieeexplore.ieee.orgY Hua2020 Information Communication Technologies Conference (ICTC), 2020•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/9123302/

Huang, H., Du, B., Zhou, H., Wang, M., Ming, Y., & Hu, G. (2025). A path planning method in three-dimensional complex space based on Bézier curves and a hybrid zebra optimization algorithm. *SpringerH Huang, B Du, H Zhou, M Wang, Y Ming, G HuCluster Computing, 2025•Springer*, *28*(2). https://doi.org/10.1007/s10586-024-04683-1

Huang, L., Chen, C., Yun, J., Sun, Y., Tian, J., Hao, Z., Yu, H., & Ma, H. (2022). Multi-Scale Feature Fusion Convolutional Neural Network for Indoor Small Target Detection. *Frontiers in Neurorobotics*, *16*. https://doi.org/10.3389/fnbot.2022.881021

Jagathbala, R., Bharath Kumar, G., Geethanjali, D., & Nanthini, N. (2025a). ForenXplorer:A Smart Digital Forensic Triage Tool for Fast and Automated Incident Response. *2025 10th International Conference on Smart Structures and Systems (ICSSS)*, 1-7. https://doi.org/10.1109/ICSSS66939.2025.11346333

Jagathbala, R., Bharath Kumar, G., Geethanjali, D., & Nanthini, N. (2025b). ForenXplorer:A Smart Digital Forensic Triage Tool for Fast and Automated Incident Response. *2025 10th International Conference on Smart Structures and Systems (ICSSS)*, 1-7. https://doi.org/10.1109/ICSSS66939.2025.11346333

Jazi, H. H., Gonzalez, H., Stakhanova, N., & Ghorbani, A. A. (2017). Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling. *Computer Networks*, *121*, 25-36. https://doi.org/10.1016/j.comnet.2017.03.018

Jia, H., Peng, X., & Lang, C. (2021). Remora optimization algorithm. *Expert Systems with Applications*, *185*. https://doi.org/10.1016/j.eswa.2021.115665

Joshi, J. C., Kumar, T., Srivastava, S., & Sachdeva, D. (2017). Optimisation of hidden Markov Model using Baum–Welch algorithm for prediction of maximum and minimum temperature over Indian Himalaya. *Journal of Earth System Science*, *126*(1). https://doi.org/10.1007/s12040-016-0780-0

Kanimozhi, & Jacob, T. P. (2019). Calibration of various optimized machine learning classifiers in network intrusion detection system on the realistic cyber dataset CSE-CIC-IDS2018 using cloud. *researchgate.netV Kanimozhi, TP JacobInternational Journal of Engineering Applied Sciences and Technology, 2019•researchgate.net*. https://www.researchgate.net/profile/Prem-Jacob/publication/368085892_CALIBRATION_OF_VARIOUS_OPTIMIZED_MACHINE_LEARNING_CLASSIFIERS_IN_NETWORK_INTRUSION_DETECTION_SYSTEM_ON_THE_REALISTIC_CYBER_DATASET_CSE-CIC-IDS2018_USING_CLOUD_COMPUTING/links/64f8c281e098013a83da13b4/CALIBR

ATION-OF-VARIOUS-OPTIMIZED-MACHINE-LEARNING-CLASSIFIERS-IN-NETWORK-INTRUSION-DETECTION-SYSTEM-ON-THE-REALISTIC-CYBER-DATASET-CSE-CIC-IDS2018-USING-CLOUD-COMPUTING.pdf

Karatas, G., Demir, O., & Sahingoz, O. K. (2020). Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset. *IEEE Access*, *8*, 32150-32162. https://doi.org/10.1109/ACCESS.2020.2973219

Keserwani, P. K., Govil, M. C., & Pilli, E. S. (2023). An effective NIDS framework based on a comprehensive survey of feature optimization and classification techniques. *Neural Computing and Applications*, *35*(7), 4993-5013. https://doi.org/10.1007/s00521-021-06093-5

Khraisat, A., Alazab, A., Singh, S., Jan, T., & Gomez, A. (2024). Survey on Federated Learning for Intrusion Detection System: Concept, Architectures, Aggregation Strategies, Challenges, and Future Directions. *ACM Computing Surveys*, *57*(1). https://doi.org/10.1145/3687124

Kim, J., Kim, J., Kim, H., & Shim, M. (2020). CNN-based network intrusion detection against denial-of-service attacks. *mdpi.comJ Kim, J Kim, H Kim, M Shim, E ChoiElectronics, 2020•mdpi.com*. https://www.mdpi.com/2079-9292/9/6/916

Knab, P., Marton, S., Schlegel, U., & Bartelt, C. (2026). Which LIME Should I Trust? Concepts, Challenges, and Solutions. *Communications in Computer and Information Science*, *2577 CCIS*, 28-52. https://doi.org/10.1007/978-3-032-08324-1_2

Koroniotis, N., Moustafa, N., … E. S.-F. G., & 2019, undefined. (2019). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *ElsevierN Koroniotis, N Moustafa, E Sitnikova, B TurnbullFuture Generation Computer Systems, 2019•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0167739X18327687

Kosugi, A., Teranishi, K., & Kogiso, K. (2024). Experimental Validation of the Attack-Detection Capability of Encrypted Control Systems Using Man-in-the-Middle Attacks. *IEEE Access*, *12*, 10535-10547. https://doi.org/10.1109/ACCESS.2024.3353289

Krishnan, A., Sampath, L. P. M. I., Foo Eddy, Y. S., & Gooi, H. B. (2018). Optimal Scheduling of a Microgrid Including Pump Scheduling and Network Constraints. *Complexity*, *2018*. https://doi.org/10.1155/2018/9842025

Lee, K., Filannino, M., … Ö. U. W. e-N. for A., & 2019, undefined. (2019). An Empirical Test of GRUs and Deep Contextualized Word Representations on De. *books.google.comK Lee, M Filannino, Ö UzunerMEDINFO 2019: Health and Wellbeing e-Networks for All: Proceedings, 2019•books.google.com*. https://books.google.com/books?hl=en&lr=&id=m--9DwAAQBAJ&oi=fnd&pg=PA218&dq=Lee+K,+Filannino+M,+Uzuner+%C3%96.+An+empirical+test+of+grus+and+deep+contextualized+word+representations+on+de-+identification.+Stud+Health+Technol+Inform.+2019%3B264:218%E2%80%9322.&ots=qhuXzJeiwr&sig=lpNYkMmgbBFwbEalfqqYokmx8yY

Leevy, J., … T. K.-I. J. of, & 2022, undefined. (2022). Feature evaluation for IoT botnet traffic classification. *inderscienceonline.comJL Leevy, TM Khoshgoftaar, J HancockInternational Journal of Internet of Things and Cyber-Assurance, 2022•inderscienceonline.com*, *2*(1), 87. https://doi.org/10.1504/ijitca.2022.124374

Lember, J., Probability, J. S.-J. of T., & 2021, undefined. (2021). Regenerativity of Viterbi process for pairwise Markov models. *SpringerJ Lember, J SovaJournal of Theoretical Probability, 2021•Springer*, *34*(1), 1-33. https://doi.org/10.1007/s10959-020-01022-z

Li, C., Wang, X., Zhang, Q., Liang, J., & Zhang, A. (2025a). DDoS Attack Autonomous Detection Model Based on Multi-Strategy Integrate Zebra Optimization Algorithm.

Computers, Materials and Continua, *82*(1), 645-674. https://doi.org/10.32604/cmc.2024.058081

Li, C., Wang, X., Zhang, Q., Liang, J., & Zhang, A. (2025b). DDoS Attack Autonomous Detection Model Based on Multi-Strategy Integrate Zebra Optimization Algorithm. *Computers, Materials and Continua*, *82*(1), 645-674. https://doi.org/10.32604/cmc.2024.058081

Li, J., Xi, B., Li, Y., Du, Q., & Wang, K. (2018). Hyperspectral classification based on texture feature enhancement and deep belief networks. *Remote Sensing*, *10*(3). https://doi.org/10.3390/rs10030396

Li, X., Chen, W., & Zhang, Q. (2020). Building auto-encoder intrusion detection system based on random forest feature selection. *ElsevierXK Li, W Chen, Q Zhang, L WuComputers & Security, 2020•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0167404820301231

Lima Filho, F. S. De, Silveira, F. A. F., De Medeiros Brito Junior, A., Vargas-Solar, G., & Silveira, L. F. (2019). Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning. *Security and Communication Networks*, *2019*. https://doi.org/10.1155/2019/1574749

Lin, P., Ye, K., & Xu, C. Z. (2019). Dynamic network anomaly detection system by using deep learning techniques. *SpringerP Lin, K Ye, CZ XuInternational conference on cloud computing, 2019•Springer*, *11513 LNCS*, 161-176. https://doi.org/10.1007/978-3-030-23502-4_12

Liu, Y., Tong, C., Wu, S., Shang, W., Xing, X., & Yuan, W. (2025). Characteristic wavelength selection based on multi-strategy fusion zebra optimization algorithm for PLSR. *Measurement: Journal of the International Measurement Confederation*, *253*. https://doi.org/10.1016/j.measurement.2025.117486

Lu, N., Li, T., Ren, X., neural, H. M.-I. transactions on, & 2016, undefined. (2017). A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. *ieeexplore.ieee.orgN Lu, T Li, X Ren, H MiaoIEEE transactions on neural systems and rehabilitation engineering, 2016•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/7546909/

Ma, M., Lahmadi, A., & Chrisment, I. (2020). *Detecting a Stealthy Attack in Distributed Control for Microgrids using Machine Learning Algorithms*. https://inria.hal.science/hal-02980115v1

Madan, A., George, A., … A. S.-2018 7th I., & 2018, undefined. (2018). Redaction of protected health information in ehrs using crfs and bi-directional lstms. *ieeexplore.ieee.orgA Madan, AM George, A Singh, MPS Bhatia2018 7th International Conference on Reliability, Infocom, 2018•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8748713/

Mei, Y., Li, B., Wang, H., Wang, X., & Negnevitsky, M. (2022). Multi-objective optimal scheduling of microgrid with electric vehicles. *Energy Reports*, *8*, 4512-4524. https://doi.org/10.1016/j.egyr.2022.03.131

Mendonça, R. V., Silva, J. C., Rosa, R. L., Saadi, M., Rodriguez, D. Z., & Farouk, A. (2022). A lightweight intelligent intrusion detection system for industrial internet of things using deep learning algorithms. *Expert Systems*, *39*(5). https://doi.org/10.1111/exsy.12917

Meyer, H., Reudenbach, C., … T. H.-E. general, & 2018, undefined. (2018). How to detect and avoid overfitting in spatio-temporal machine learning applications. *ui.adsabs.harvard.edu*. https://ui.adsabs.harvard.edu/abs/2018EGUGA..20.8365M/abstract

Mhawi, D. N., Aldallal, A., & Hassan, S. (2022). Advanced Feature-Selection-Based Hybrid Ensemble Learning Algorithms for Network Intrusion Detection Systems. *Symmetry*, *14*(7). https://doi.org/10.3390/sym14071461

Mirsadeghi, S. M. H., Bahsi, H., Vaarandi, R., & Inoubli, W. (2023). Learning From Few Cyber-Attacks: Addressing the Class Imbalance Problem in Machine Learning-Based Intrusion Detection in Software-Defined Networking. *IEEE Access*, *11*, 140428-140442. https://doi.org/10.1109/ACCESS.2023.3341755

Mirsky, Y., Doitshman, T., Elovici, Y., & Shabtai, A. (2018). Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. *25th Annual Network and Distributed System Security Symposium, NDSS 2018*. https://doi.org/10.14722/ndss.2018.23204

Natarajan. (2020). Cyber secure man-in-the-middle attack intrusion detection using machine learning algorithms. *igi-global.comJ NatarajanAI and Big Data's Potential for Disruptive Innovation, 2020•igi-global.com*. https://www.igi-global.com/chapter/cyber-secure-man-in-the-middle-attack-intrusiondetection-using-machine-learning-algorithms/236343

Nebaba, A. N., Savvas, I. K., Butakova, M. A., Chernov, A. V., & Shevchuk, P. S. (2021). Improving Multiclass Classification of Cybersecurity Breaches in Railway Infrastructure using Imbalanced Learning. *ACM International Conference Proceeding Series*, 100-105. https://doi.org/10.1145/3501774.3501789

Negi, S., Kumar, Y., & Mishra, V. M. (2016). *2016 2nd International Conference on Advances in Computing, Communication, and Automation*. IEEE.

Nguyen, H. K., Khodaei, A., & Han, Z. (2016). A big data scale algorithm for optimal scheduling of integrated microgrids. *ieeexplore.ieee.orgHK Nguyen, A Khodaei, Z HanIEEE Transactions on Smart Grid, 2016•ieeexplore.ieee.org*. https://doi.org/10.1109/TSG.2017.2550422

Ozbay, F. A. (2025). An Enhanced Zebra Optimization Algorithm with Multiple Strategies for Global Optimization and Feature Selection Problems: A Hepatocellular Carcinoma Case Study. *IEEE Access*, *13*, 30036-30057. https://doi.org/10.1109/ACCESS.2025.3541975

Panse, T., & Bandhu, K. C. (2025). *Evaluating Next-Generation Firewalls Using Machine Learning-Based Hybrid Feature Selection for Threat Detection and Risk Mitigation*. 149-156. https://doi.org/10.5220/0014275000004928

Pasupa, K., Vatathanavaro, S., & Tungjitnob, S. (2023). Convolutional neural networks based focal loss for class imbalance problem: a case study of canine red blood cells morphology classification. *SpringerK Pasupa, S Vatathanavaro, S TungjitnobJournal of Ambient Intelligence and Humanized Computing, 2023•Springer*, *14*(11), 15259-15275. https://doi.org/10.1007/s12652-020-01773-x

Patil, A., Laturkar, A., Athawale, S. V, Takale, R., & Tathawade, P. (2017). *A Multilevel System to Mitigate DDoS, Brute force and SQL Injection Attack for Cloud Security*.

Pavithra, S., & Venkata Vikas, K. (2024). Detecting Unbalanced Network Traffic Intrusions With Deep Learning. *IEEE Access*, *12*, 74096-74107. https://doi.org/10.1109/ACCESS.2024.3405187

Pawlicki, M., Choraś, M., Kozik, R., & Hołubowicz, W. (2020). On the impact of network data balancing in cybersecurity applications. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12140 LNCS*, 196-210. https://doi.org/10.1007/978-3-030-50423-6_15

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R. B., Peng, S., Poulter, B., Shi, X., Xiao, Z., Zeng, N., Zeng, Z., & Wang, Y. (2015). Detection and

attribution of vegetation greening trend in China over the last 30 years. *Global Change Biology*, *21*(4), 1601-1609. https://doi.org/10.1111/gcb.12795
Ponkumar, D. D. N., Arun, M. R., Bharadwaj, M. H., Nagendrababu, G., Reddy, A. D., & Saravanakumar, R. (2025). Maximizing the Reliability of Machine Learning Based Invasion Detection Systems on a Modern, Unbalanced Dataset. *2025 3rd International Conference on Communication, Security, and Artificial Intelligence, ICCSAI 2025*, 1055-1060. https://doi.org/10.1109/ICCSAI64074.2025.11064570
Punia, P., Raj, A., & Kumar, P. (2025). Enhanced zebra optimization algorithm for reliability redundancy allocation and engineering optimization problems. *Cluster Computing*, *28*(4). https://doi.org/10.1007/s10586-024-04931-4
Ragab. (2024). Hybrid firefly particle swarm optimisation algorithm for feature selection problems. *Wiley Online LibraryM RagabExpert Systems, 2024•Wiley Online Library*, *41*(7). https://doi.org/10.1111/exsy.13363
Rahman, M. A. , Hossain, M. A. , Kabir, M. R. , Sani, M. H. , & Awal, M. A. (2019). Optimization of sleep stage classification using single-channel eeg signals. *ieeexplore.ieee.orgMA Rahman, MA Hossain, MR Kabir, MH Sani, MA Awal2019 4th International Conference on Electrical Information and, 2019•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/9068825/
Ramos, K. H., Sensors, M. S. M.-, & 2020, undefined. (2020). Benchmark-based reference model for evaluating botnet detection tools driven by traffic-flow analytics. *mdpi.comKS Huancayo Ramos, MA Sotelo Monge, J Maestre VidalSensors, 2020•mdpi.com*. https://www.mdpi.com/1424-8220/20/16/4501
Ranjan, Amar Kumar Verma, & Sudha Radhika. (2019). *K-Nearest Neighbors and Grid Search CV Based Real Time Fault Monitoring System for IndustriesCT) : Pune, India, Mar 29-31, 2019*. IEEE.
Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2009). GSA: A Gravitational Search Algorithm. *Information Sciences*, *179*(13), 2232-2248. https://doi.org/10.1016/j.ins.2009.03.004
Reddy, R., Kulkarni, A. J., Krishnasamy, G., Shastri, A. S., & Gandomi, A. H. (2023). LAB: a leader–advocate–believer-based optimization algorithm. *Soft Computing*, *27*(11), 7209-7243. https://doi.org/10.1007/s00500-023-08033-y
Rossmann, M. G., & Van Beek, C. G. (1999). *D55, 1631±1640 Rossmann & van Beek Data processing 1631 research papers Acta Crystallographica Section D*. http://bilbo.bio.purdue.edu/~viruswww/
Roy, A., Cruz, R. M. O., Sabourin, R., & Cavalcanti, G. D. C. (2018). A study on combining dynamic selection and data preprocessing for imbalance learning. *Neurocomputing*, *286*, 179-192. https://doi.org/10.1016/j.neucom.2018.01.060
Rymarczyk, T., Kozłowski, E., Kłosowski, G., & Niderla, K. (2019). Logistic regression for machine learning in process tomography. *Sensors (Switzerland)*, *19*(15). https://doi.org/10.3390/s19153400
Rynkiewicz, J. (2019). Asymptotic statistics for multilayer perceptron with ReLU hidden units. *Neurocomputing*, *342*, 16-23. https://doi.org/10.1016/j.neucom.2018.11.097
Saikia, T., Brox, T., & Schmid, C. (2020). *Optimized Generic Feature Learning for Few-shot Classification across Domains*. http://arxiv.org/abs/2001.07926
Saxena, A., Sinha, S., conference, P. S.-2017 I., & 2017, undefined. (2017). General study of intrusion detection system and survey of agent based intrusion detection system. *ieeexplore.ieee.orgAK Saxena, S Sinha, P Shukla2017 International conference on computing, communication and, 2017•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8229866/

Shah, S., Systems, B. I.-F. G. C., & 2018, undefined. (2017). Performance comparison of intrusion detection systems and application of machine learning to Snort system. *ElsevierSAR Shah, B IssacFuture Generation Computer Systems, 2018•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0167739X17323178

Shang-Guan, Y. P., Sun, Y. F., Wang, J. S., Gao, Y. Z., & Yan, B. (2025). Zebra optimization algorithm with chaos convergence factor and Gaussian mutation for MLP soft-sensor model of debutanizer column. *Cluster Computing*, *28*(11). https://doi.org/10.1007/s10586-025-05319-8

Sharafaldin, I., Lashkari, A., ICISSp, A. G.-, & 2018, undefined. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *scitepress.orgI Sharafaldin, AH Lashkari, AA GhorbaniICISSp, 2018•scitepress.org*. https://doi.org/10.5220/0006639801080116

Sharma, A. (2018). Guided Stochastic Gradient Descent Algorithm for inconsistent datasets. *Applied Soft Computing Journal*, *73*, 1068-1080. https://doi.org/10.1016/j.asoc.2018.09.038

Shiravi, A., Shiravi, H., Tavallaee, M., & Ghorbani, A. A. (2012). Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computers and Security*, *31*(3), 357-374. https://doi.org/10.1016/j.cose.2011.12.012

Simhon, Z., Weiss, M., Marbel, R., Hajaj, C., Dvir, A., & Dubin, R. (2026). Real-Time Network Security: Integrating ANN and Dynamic Graph-Based Clustering. *Computer Networks*, 112016. https://doi.org/10.1016/j.comnet.2026.112016

Singh, A., … M. S.-J. of C. N. and I., & 2014, undefined. (2014). Analysis of host-based and network-based intrusion detection system. *mecs-press.orgAP Singh, MD SinghInternational Journal of Computer Network and Information Security, 2014•mecs-press.org*, *8*, 41-47. https://doi.org/10.5815/ijcnis.2014.08.06

Somani, G., Gaur, M. S., Sanghi, D., Conti, M., & Buyya, R. (2017). DDoS attacks in cloud computing: Issues, taxonomy, and future directions. Içinde *Computer Communications* (C. 107, ss. 30-48). Elsevier B.V. https://doi.org/10.1016/j.comcom.2017.03.010

Stephan, T., Punitha, S., & Mishra, V. (2025). Improved Zebra optimization strategies for efficient feature selection in high-dimensional spaces. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-025-05008-9

Sulaiman, S., Abdul Wahid, R., Hanee Ariffin, A., & Zalina Zulkifli, C. (2020). *Question Classification Based on Cognitive Levels using Linear SVC*.

Sumant, A. S., & Patil, D. (2022). Ensemble Feature Subset Selection: Integration of Symmetric Uncertainty and Chi-Square techniques with RReliefF. *Journal of The Institution of Engineers (India): Series B*, *103*(3), 831-844. https://doi.org/10.1007/s40031-021-00684-5

Swain, P. K., Satpathy, S., & Pokuri, S. R. (2025). MACHINE LEARNING BASED STRATEGY FOR MITM ATTACK MITIGATION UTILIZING HYBRID FEATURE SELECTION. *Proceedings on Engineering Sciences*, *7*(1), 615-624. https://doi.org/10.24874/PES07.01D.017

Taherdoost, H. (2024). Insights into Cybercrime Detection and Response: A Review of Time Factor. *Information (Switzerland)*, *15*(5). https://doi.org/10.3390/info15050273

Taherkhani, A., Cosma, G., & McGinnity, T. M. (2018). Deep-FS: A feature selection algorithm for Deep Boltzmann Machines. *Neurocomputing*, *322*, 22-37. https://doi.org/10.1016/j.neucom.2018.09.040

Talbi, E.-G. (2009). *Metaheuristics: from design to implementation*. https://books.google.com/books?hl=en&lr=&id=SIsa6zi5XV8C&oi=fnd&pg=PR7&dq=Talbi,+E.+G.+(2009).+Metaheuristics:+From+Design+to+Implementation.+*John+Wiley+%26+Sons*.&ots=-cWGwLjrIq&sig=gdj1XgelmR5NNNvZhb5o0iANt6k

Tan, Z., Tian, Y., & Li, J. (2023). *GLIME: General, Stable and Local LIME Explanation*.

Tarek Abdelaziz, M., Radwan, A., Mamdouh, H., Saad, A. S., Abdulrahman, ·, Abuzaid, S., Ahmed, ·, Abdelhakeem, A., Salma Zakzouk, ·, Moussa, K., & Saeed Darweesh, · M. (2025). *Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance*. *33*(2). https://doi.org/10.1007/s10922-024-09874-0

Taş, O. (t.y.). *İSTANBUL SABAHATTİN ZAİM ÜNİVERSİTESİ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ BİLİM DALI NESNELERİN İNTERNETİ İÇİN AKILLI SALDIRI TESPİT SİSTEMLERİ GELİŞTİRİLMESİ DOKTORA TEZİ*.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). *A Detailed Analysis of the KDD CUP 99 Data Set Tavallaee A Detailed Analysis of the KDD CUP 99 Data Set*. http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/jsp/nparc_cp.jsp?lang=frL'accèsàcesiteWebetl'utilisationdesoncontenusontassujettisauxconditionsprésentéesdanslesitehttp://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/jsp/nparc_cp.jsp?lang=en

Topbaş, Z., Hazırlanan, T., Bir, ", Evrişimli, B., & Ağları, S. (t.y.). *TEZ KABUL VE ONAYI*.

Trojovska, E., Dehghani, M., & Trojovsky, P. (2022). Zebra Optimization Algorithm: A New Bio-Inspired Optimization Algorithm for Solving Optimization Algorithm. *IEEE Access*, *10*, 49445-49473. https://doi.org/10.1109/ACCESS.2022.3172789

Vajda, S., image, K. S.-I. conference on recent trends in, & 2016, undefined. (2017). A fast k-nearest neighbor classifier using unsupervised clustering. *SpringerS Vajda, KC SantoshInternational conference on recent trends in image processing and pattern, 2016•Springer*, *709*, 185-193. https://doi.org/10.1007/978-981-10-4859-3_17

Varsamopoulos, S., Criger, B., & Bertels, K. (2017). *Decoding Small Surface Codes with Feedforward Neural Networks*. https://doi.org/10.1088/2058-9565/aa955a

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.

Wang, M., Xu, X., Yue, Q., arXiv:2101.12631, Y. W. preprint, & 2021, undefined. (2021). A comprehensive survey and experimental comparison of graph-based approximate nearest neighbor search. *arxiv.orgM Wang, X Xu, Q Yue, Y WangarXiv preprint arXiv:2101.12631, 2021•arxiv.org*, *14*(1). https://arxiv.org/abs/2101.12631

Wang, R., Zeng, S., Wang, X., & Ni, J. (2019). Machine learning for hierarchical prediction of elastic properties in Fe-Cr-Al system. *Computational Materials Science*, *166*, 119-123. https://doi.org/10.1016/j.commatsci.2019.04.051

Wang, Z., Ye, X., Jiang, G., & Yi, Y. (2025). Improved Zebra Optimization Algorithm with Multi Strategy Fusion and Its Application in Robot Path Planning. *Biomimetics*, *10*(6). https://doi.org/10.3390/biomimetics10060354

Yadav, S., & Shukla, S. (2016). Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. *ieeexplore.ieee.orgS Yadav, S Shukla2016 IEEE 6th International conference on advanced computing (IACC), 2016•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/7544814/

Yuanyuan, S., & Yongming, W. (2017). The comparison of optimizing SVM by GA and grid search. *ieeexplore.ieee.orgS Yuanyuan, W Yongming, G Lili, M Zhongsong, J Shan2017 13th IEEE international conference on electronic measurement, 2017•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8265815/

Zhang. (2018). Improved adam optimizer for deep neural networks. *ieeexplore.ieee.orgZ Zhang2018 IEEE/ACM 26th international symposium on quality of service, 2018•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8624183/

Zhang, H., on, T. C.-2024 14th I. C., & 2024, undefined. (2024). A hybrid approach to network intrusion detection based on graph neural networks and transformer

architectures. *ieeexplore.ieee.orgH Zhang, T Cao2024 14th International Conference on Information Science and, 2024•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/10805457/

Zhao, F., Zhang, H., Peng, J., Zhuang, X., & Na, S. G. (2020). A semi-self-taught network intrusion detection system. *Neural Computing and Applications*, *32*(23), 17169-17179. https://doi.org/10.1007/s00521-020-04914-7

Zhao, Y., Li, H., Wan, S., … A. S.-I. journal of, & 2019, undefined. (2019). Knowledge-aided convolutional neural network for small organ segmentation. *ieeexplore.ieee.orgY Zhao, H Li, S Wan, A Sekuboyina, X Hu, G Tetteh, M Piraud, B MenzeIEEE journal of biomedical and health informatics, 2019•ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/8606255/

Zhu, R., Ji, X., Yu, D., Tan, Z., Zhao, L., Li, J., & Xia, X. (2020). KNN-Based Approximate Outlier Detection Algorithm over IoT Streaming Data. *IEEE Access*, *8*, 42749-42759. https://doi.org/10.1109/ACCESS.2020.2977114