T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

AĞ TRAFİĞİ KULLANARAK SALDIRI TESPİT EDEBİLEN DERİN ÖĞRENME MODELLERİNİN PERFORMANS DEĞERLENDİRMESİ

YÜKSEK LİSANS

RACHID CHEICK MOHAMED

Temmuz 2026

GÜMÜŞHANE

T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

AĞ TRAFİĞİ KULLANARAK SALDIRI TESPİT EDEBİLEN DERİN ÖĞRENME MODELLERİNİN PERFORMANS DEĞERLENDİRMESİ

PERFORMANCE EVALUATION OF DEEP LEARNING MODELS FOR INTRUSION DETECTION USING NETWORK TRAFFIC

YÜKSEK LİSANS

RACHID CHEICK MOHAMED

Temmuz 2026

GÜMÜŞHANE

GÜMÜŞHANE ÜNİVERSİTESİ
Lisansüstü Eğitim Enstitüsü

T.C.

GÜMÜŞHANE ÜNİVERSİTESİ

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YAPAY ZEKÂ VE AKILLI SİSTEMLER ANA BİLİM DALI

AĞ TRAFİĞİ KULLANARAK SALDIRI TESPİT EDEBİLEN DERİN ÖĞRENME MODELLERİNİN PERFORMANS DEĞERLENDİRMESİ

PERFORMANCE EVALUATION OF DEEP LEARNING MODELS FOR INTRUSION DETECTION USING NETWORK TRAFFIC

YÜKSEK LİSANS

RACHID CHEICK MOHAMED

DANIŞMAN: DR. ÖĞR. ÜYESİ SAMET TONYALI

Temmuz 2026

GÜMÜŞHANE

KABUL VE ONAY

Dr. Öğr. Üyesi Samet TONYALI danışmanlığında, **Rachid CHEICK MOHAMED** tarafından hazırlanan "**Ağ Trafiği Kullanarak Saldırı Tespit Edebilen Derin Öğrenme Modellerinin Performans Değerlendirmesi**" isimli bu çalışma, 03/07/2026 tarihinde yapılan lisansüstü tez savunma sınavı sonucunda **Oy Birliği** ile başarılı bulunarak jürimiz tarafından **Yüksek Lisans Tezi** olarak kabul edilmiştir.

…….
Dr. Öğr. Üyesi Ramazan Özgür DOĞAN (Başkan)

…….
Dr. Öğr. Üyesi Samet TONYALI (Danışman)

…….
Dr. Öğr. Üyesi Gökhan ÇETİN (Üye)

Lisansüstü tez savunma sınavında başarılı bulunarak kabul edilen bu tezin ciltlenmiş hali, ... /..... /…… tarihli ve …… / …… sayılı Enstitü Yönetim Kurulu toplantısında görüşülmüş ve tez yazım kılavuzuna uygun bulunarak onaylanmıştır.

Prof. Dr. Duygu ÖZDEŞ

Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Yüksek Lisans Tezi olarak hazırlamış olduğum "**Ağ Trafiği Kullanarak Saldırı Tespit Edebilen Derin Öğrenme Modellerinin Performans Değerlendirmesi**" isimli bu tezimin tamamen kendi çalışmam olduğunu, danışmanımın sorumluluğunda hazırladığımı, her alıntıya kaynak gösterdiğimi, alıntı yaptığım tüm çalışmaları kaynakçada belirttiğimi ve Gümüşhane Üniversitesi'nin lisanslı kullanıcısı olduğu intihal yazılım programı ile Lisansüstü Eğitim Enstitüsü'nün belirlediği kıstaslara uygun olarak raporladığımı taahhüt ederim. Tezimin kâğıt ve elektronik kopyalarının Gümüşhane Üniversitesi Lisansüstü Eğitim Enstitüsü arşivinde saklanmasına izin verdiğimi onaylarım.

03/07/2026

…….

Rachid CHEICK MOHAMED

TEŞEKKÜR

Her şeyden önce, bu tezin hazırlanması sürecindeki rehberliği, desteği ve değerli geri bildirimleri için danışmanım Dr. Öğr. Üyesi Samet TONYALI'ya içten teşekkürlerimi sunarım. Uzmanlığı ve yapıcı tavsiyeleri, bu çalışmanın yön ve niteliğinin şekillenmesinde belirleyici bir rol oynamıştır.

Bu araştırmayı yürütmem için gerekli akademik ortamı ve kaynakları sağlayan Gümüşhane Üniversitesine ve Lisansüstü Eğitim Enstitüsüne en derin şükranlarımı sunarım. Yapay Zekâ ve Akıllı Sistemler Anabilim Dalı'nda yüksek lisans eğitimi alma fırsatı, benim için gerçekten zenginleştirici bir deneyim olmuştur.

Ayrıca, öğrenimim boyunca destek ve teşviklerini esirgemeyen bölüm öğretim üyeleri ve çalışanlarına da teşekkür ederim.

Koşulsuz sevgileri, sabırları ve fedakârlıklarıyla en büyük güç kaynağım olan aileme yürekten teşekkür ederim. Kamerun'dan Türkiye'ye gelip lisansüstü eğitim almak, onların bana olan sarsılmaz inancı olmadan mümkün olmayacaktı; bu çalışmayı onlara ithaf ediyorum.

CIC-IDS2017 ve UNSW-NB15 veri kümelerini kamuya açık hâle getiren araştırmacılara ve kurumlara da ayrıca teşekkür ederim; bu temel kaynaklar olmaksızın çalışma gerçekleştirilemezdi.

Son olarak, bu yolculuğu tamamlamak için bana azim, sağlık ve zihinsel berraklık bahşeden Allah'a şükranlarımı sunarım.

Rachid CHEICK MOHAMED

GÜMÜŞHANE – 2026

ÖZET

Dijital altyapıların hızla büyümesi ve siber saldırıların giderek karmaşıklaşması, ağ tabanlı saldırı tespit sistemlerini modern siber güvenliğin temel bileşenlerinden biri hâline getirmiştir. İmza tabanlı yöntemler yeni tehditlerin tespitinde yetersiz kalabildiğinden, makine öğrenmesi ve derin öğrenme yöntemleri önemli alternatifler sunmaktadır. Bu çalışmada, tek boyutlu evrişimli sinir ağı (CNN) ile hibrit CNN-BiLSTM mimarisinin ikili ve çok sınıflı ağ saldırısı sınıflandırma performansları incelenmiştir. Modeller, CIC-IDS2017 ve UNSW-NB15 veri kümeleri üzerinde tabakalı beş katlı çapraz doğrulama, odak kaybı, QuantileTransformer ile ölçeklendirme ve SMOTE ile aşırı örnekleme içeren ortak bir deneysel düzen kullanılarak değerlendirilmiştir. Ayrıca rastgele orman ve XGBoost, UNSW-NB15 üzerinde klasik karşılaştırma modelleri olarak test edilmiştir. Bulgular, CNN modelinin CIC-IDS2017 ikili sınıflandırma görevinde %99,75 doğruluk ve 0,9999 ROC-AUC değerine ulaştığını göstermiştir. CNN-BiLSTM, UNSW-NB15 ikili sınıflandırma görevinde yanlış negatifleri yaklaşık %36 azaltmış; ancak çıkarım gecikmesi CNN'e göre 4-6 kat artmıştır. Çok sınıflı sınıflandırmada CNN, CIC-IDS2017 üzerinde 0,701 makro F1 skoruyla CNN-BiLSTM'den (0,671) daha yüksek performans göstermiştir. XGBoost ise derin öğrenme modelleriyle benzer F1 skorlarına daha kısa sürede ulaşmıştır. Azınlık saldırı sınıflarının tespiti tüm yapılandırmalarda temel bir güçlük olmayı sürdürmektedir. Sonuçlar, gerçek uygulama kısıtları altında derin öğrenme ve klasik makine öğrenmesi modelleri arasında seçim yapacak araştırmacı ve uygulayıcılar için karşılaştırmalı çıkarımlar sunmaktadır.

Anahtar Kelimeler: Ağ tabanlı saldırı tespit sistemi, CNN-BiLSTM, Çok sınıflı sınıflandırma, Evrişimli sinir ağı, Odak kaybı,

ABSTRACT

Digital infrastructures are growing rapidly and cyberattacks are becoming increasingly sophisticated. As a result, Network-based Intrusion Detection Systems have become a critical component of modern cybersecurity. Signature-based methods sometimes fail to detect new threats, where machine learning and deep learning can provide a valuable alternative. This study examines a one-dimensional Convolutional Neural Network and a hybrid model called CNN-BiLSTM, investigating how both architectures can detect network attacks in binary and multiclass classification settings. Both architectures were evaluated on the CIC-IDS2017 and UNSW-NB15 benchmark datasets through a rigorous experimental protocol incorporating stratified 5-fold cross-validation, Focal Loss, QuantileTransformer scaling, and SMOTE oversampling. Random Forest and XGBoost were additionally tested on UNSW-NB15 as classical baselines. The results show that CNN achieves near-perfect binary classification performance on CIC-IDS2017, with 99.75% accuracy and a ROC-AUC of 0.9999. CNN-BiLSTM reduces false negatives by approximately 36% in the binary UNSW-NB15 setting, albeit at the cost of 4–6 times higher inference latency. In multiclass classification, CNN outperforms CNN-BiLSTM on CIC-IDS2017 with a higher macro F1-Score (0.701 vs. 0.671). XGBoost achieves comparable F1 performance to deep learning models in significantly less training time, once again demonstrating the enduring value of ensemble methods. Minority attack class detection remains a fundamental unresolved challenge across all configurations, laying the groundwork for future research on class-wise threshold calibration and Transformer-based architectures. The findings provide actionable insights for practitioners choosing between deep learning and classical approaches under real-world NIDS deployment constraints.

Keywords: Network intrusion detection system, CNN-BiLSTM, Multiclass classification, Convolutional neural network, Focal loss

İÇİNDEKİLER

TABLOLAR DİZİNİ

ŞEKİLLER DİZİNİ

SİMGELER VE KISALTMALAR DİZİNİ

∑	: Toplam sembolü
%	: Yüzde
AGA	: Aşırı Gradyan Artırma
AGL-VBK	: Ağ Güvenliği Laboratuvarı – Veritabanlarında Bilgi Keşfi
AİK	: Alıcı İşletim Karakteristiği
AİK-EAA	: Alıcı İşletim Karakteristiği Eğrisi Altındaki Alan (AUC-ROC)
AMSTS	: Ana Makine Tabanlı Saldırı Tespit Sistemi (Host-Based IDS)
AS	: Araştırma Sorusu
ATSTS	: Ağ Tabanlı Saldırı Tespit Sistemi (Network-Based IDS)
BiLSTM	: Çift Yönlü Uzun Kısa Süreli Bellek (Bidirectional LSTM)
CIC-IDS2017	: Kanada Siber Güvenlik Enstitüsü Saldırı Tespit Sistemi 2017 Veri
CNN	: Evrişimli Sinir Ağı (Convolutional Neural Network)
CNN1D	: Tek Boyutlu Evrişimli Sinir Ağı (1D CNN)
ÇD	: Çapraz Doğrulama (Cross-Validation)
ÇKA	: Çok Katmanlı Algılayıcı (MLP- Multilayer Perceptron)
DDoS	: Dağıtılmış Hizmet Engelleme (Distributed Denial of Service)
DDrB	: Doğrultulmuş Doğrusal Birim (ReLU- Rectified Linear Unit)
DoS	: Hizmet Engelleme (Denial of Service)
DÖ	: Derin Öğrenme (Deep Learning)
DSA	: Derin Sinir Ağı (Deep Neural Network)
DVM	: Destek Vektör Makinesi (SVM- Support Vector Machine)
EAA	: Eğri Altındaki Alan (AUC- Area Under the Curve)
ENİ	: Endüstriyel Nesnelerin İnterneti (IIoT- Industrial IoT)
FTP	: Dosya Aktarım Protokolü (File Transfer Protocol)
GİB	: Grafik İşlem Birimi (GPU- Graphics Processing Unit)
GOM	: Güvenlik Operasyon Merkezi (SOC- Security Operations Center)

GPO	: Gerçek Pozitif Oranı (TPR- True Positive Rate)
İBSA	: İleri Beslemeli Sinir Ağı (Feedforward Neural Network)
KB	: Karşılıklı Bilgi (Mutual Information)
KEK	: K-En Yakın Komşu (KNN- K-Nearest Neighbors)
KG-EAA	: Kesinlik-Geri Çağırma Eğrisi Altındaki Alan (PR-AUC)
KOH	: Küresel Ortalama Havuzlama (Global Average Pooling)
KTB	: Kişisel Tanımlayıcı Bilgi (PII- Personally Identifiable Information)
KYB	: Kapılı Yinelemeli Birim (GRU- Gated Recurrent Unit)
MÖ	: Makine Öğrenmesi (Machine Learning)
OKF	: Odak Kayıp Fonksiyonu (Focal Loss Function)
PÖ	: Pekiştirmeli Öğrenme (Reinforcement Learning)
R2L	: Uzaktan Yerel'e (Remote to Local) – saldırı türü
RO	: Rastgele Orman (Random Forest)
SAÖT	: Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE)
STS	: Saldırı Tespit Sistemi (IDS- Intrusion Detection System)
TB	: Tam Bağlantılı (Fully Connected)
TBA	: Temel Bileşen Analizi (PCA- Principal Component Analysis)
TO	: Tespit Oranı (Detection Rate)
U2R	: Kullanıcıdan Köke (User to Root) – saldırı türü
UKSB	: Uzun Kısa Süreli Bellek (LSTM- Long Short-Term Memory)
UNSW-NB15	: Yeni Güney Galler Üniversitesi Ağ Tabanlı 2015 Veri Seti
VBK	: Veritabanlarında Bilgi Keşfi (KDD- Knowledge Discovery in Databases)
XSS	: Siteler Arası Betik Çalıştırma (Cross-Site Scripting)
YPO	: Yanlış Pozitif Oranı (FPR- False Positive Rate)
YSD	: Yapılandırılmış Sorgu Dili (SQL- Structured Query Language)
YSA	: Yinelemeli Sinir Ağı (RNN- Recurrent Neural Network)
YZ	: Yapay Zekâ (AI- Artificial Intelligence)

1. GİRİŞ

Dijital altyapıların hızla genişlemesi ve birbirine bağlı sistemlerin üstel büyümesi, siber güvenlik tehdit ortamını kökten dönüştürmüştür. Modern ağlar, finans, sağlık, enerji ve kamu yönetimi gibi kritik sektörleri kapsayarak günlük milyarlarca veri paketini işlemektedir. Bu devasa ölçek, kötü niyetli aktörlerin iletişim protokollerindeki, uygulama katmanlarındaki ve insan davranışındaki güvenlik açıklarını sürekli olarak istismar eden karmaşık saldırı stratejileri geliştirip rafine iyileştirdiği bir ortam yaratmaktadır. Bu arka plan karşısında, saldırıları doğru, hızlı ve güvenilir biçimde tespit etme yeteneği, uygulamalı bilgisayar biliminin temel zorluklarından biri hâline gelmiştir.

Bu tez, iki sinir ağı modelini tek boyutlu Evrişimli Sinir Ağı (CNN) ve hibrit CNN-BiLSTM mimarisini yaygın olarak kullanılan iki kıyaslama veri kümesi olan CIC-IDS2017 ve UNSW-NB15 üzerinde değerlendirip karşılaştırarak bu zorluğu ele almaktadır. Değerlendirme hem ikili hem de çok sınıflı sınıflandırma görevlerini kapsamakta; sonuçları daha geniş NIDS literatürü içinde konumlandırmak amacıyla klasik makine öğrenmesi taban çizgileriyle (Rastgele Orman ve XGBoost) tamamlanmaktadır.

1.1. Problem Tanımı ve Bağlamı

Saldırı tespiti, bir güvenlik ihlalini, politika ihlalini ya da sistem bütünlüğünü tehlikeye atacak kötü niyetli girişimi işaret edebilecek olayları belirlemek amacıyla ağ trafiğini ve sistem etkinliğini izleme sürecini ifade etmektedir. Geleneksel saldırı tespit yaklaşımları, gözlemlenen trafik örüntülerini bir veritabanında saklanan bilinen saldırı imzalarıyla eşleştiren kural tabanlı veya imza tabanlı sistemlere büyük ölçüde dayanmaktaydı. Bilinen tehditlere karşı etkili olmakla birlikte, bu yöntemler özünde reaktiftir: önceden tanımlanmış kural kümelerinin dışında kalan yeni saldırı türlerini, sıfır gün açıklarını ve davranışsal anormallikleri tespit edemezler.

İmza tabanlı yöntemlerin sınırlılıkları, makine öğrenmesi (MÖ) ve derin öğrenme (DÖ) yaklaşımlarına yönelik önemli araştırma ilgisini yönlendirmiştir. Bu yaklaşımlar, elle tasarlanmış kurallara dayanmaksızın trafik verilerinden doğrudan ayırt edici öznitelikleri öğrenebilmektedir. Sinir ağlarının yüksek boyutlu ve heterojen ağ akışı verilerinden hiyerarşik temsilleri otomatik olarak çıkarma yeteneği, onları bu uygulama için özellikle umut verici adaylar konumuna taşımaktadır.

1.2. Siber Güvenlik ve IDS İçinde NIDS'in Önemi

(Magazine, 2018)'e göre, siber güvenlik dijital çağda kuruluşlar için stratejik bir öncelik hâline gelmiştir. Küresel siber suçun 2023 yılına kadar 8 trilyon ABD dolarının üzerinde maliyete yol açması ve bu rakamın 2025 itibarıyla 10 trilyon ABD dolarını aşması beklenmektedir. Siber saldırılar yalnızca finansal bilgileri değil, kritik altyapıları da etkileyebilmekte; milyonlarca kişinin kişisel bilgisinin ele geçirilmesine ya da çalınmasına yol açabilmekte ve dijital hizmetlere duyulan güveni zedeleyebilmektedir. Öne çıkan siber saldırılar arasında (Coco vd., 2022) tarafından bildirilen SolarWinds tedarik zinciri saldırısı, (*Colonial Pipeline*, t.y.) yayınında yer alan Colonial Pipeline fidye yazılımı saldırısı ve sağlık ile kamu sektörlerini hedef alan çok sayıda siber güvenlik ihlali sayılabilir. Bu bağlamda Ağ Tabanlı Saldırı Tespit Sistemleri, katmanlı bir siber güvenlik savunma stratejisinin merkezinde yer almaktadır. Stratejik ağ izleme noktalarına konuşlandırılan NIDS'ler, trafiği pasif biçimde izleyerek güvenlik operasyon ekiplerini şüpheli etkinlikler konusunda uyarmaktadır. Ağ sınırlarında erişim politikalarını uygulayan güvenlik duvarı gibi önleyici kontrollerden farklı olarak NIDS'ler, trafik davranışına sürekli görünürlük sağlayarak çevre savunmalarını aşmış tehditlerin tespitini mümkün kılmaktadır. Etkinlikleri, doğrudan olay müdahalesinin hızını ve kalitesini belirlemektedir.

Gelenekselden yapay zekâ destekli NIDS'e geçiş artık bir araştırma ufku değil, operasyonel bir zorunluluktur. Modern ağ trafiğinin hacmi ve hızı, manuel analizi olanaksız kılmakta; gelişmiş kalıcı tehditlerin (APT) karmaşıklığı ise statik kural tabanlı sistemleri yetersiz bırakmaktadır. Uyum sağlayabilen, genelleştirebilen ve ölçeklenebilen makine öğrenmesi tabanlı NIDS'ler, bir sonraki nesil ağ güvenliği izleme araçlarını temsil etmekte ve bunların titiz biçimde değerlendirilmesi hem araştırma topluluğuna hem de siber güvenlik sektörüne anlamlı katkılar sunmaktadır.

1.3. Motivasyonlar

Bu araştırmanın motivasyonları, birbiriyle örtüşen iki gözlemden kaynaklanmaktadır: biri akademik, diğeri teknik.

Akademik açıdan, NIDS literatürü zengin olmakla birlikte, kontrollü ve tekrarlanabilir deneysel koşullar altında sistemli mimari karşılaştırmalardan yoksun kalmaktadır. Tutarsız ön işleme ardışık düzenleri, değişen eğitim/test bölümleri ve farklı değerlendirme metrikleri nedeniyle yayımlanan sonuçların büyük çoğunluğu karşılaştırılamamaktadır. Bu tez, gelecekteki çalışmalar için güvenilir bir başvuru

noktası işlevi görebilecek titiz ve metodolojik açıdan tutarlı bir karşılaştırmalı değerlendirmeye katkıda bulunmayı amaçlamaktadır.

Teknik açıdan ise sıralı girdiyi hem ileri hem geri yönde işleyen BiLSTM mimarisi, ağ trafiği analizi için teorik açıdan cazip bir kapasite sunmaktadır: akış düzeyinde zamansal bağımlılıkları çift yönlü olarak yakalayabilme yeteneği. Bununla birlikte, tablosal ağ akışı verileri üzerinde daha basit evrişimsel mimarilere kıyasla pratik üstünlüğü ampirik olarak yeterince araştırılmamıştır. Mevcut çalışma, bu üstünlüğü görevler ve veri kümeleri genelinde ölçen doğrudan ve kontrollü bir karşılaştırma sunmaktadır.

1.4. Problem İfadesi

Bu tezde ele alınan temel problem şu şekilde ifade edilebilir: literatürde derin öğrenme tabanlı NIDS önerilerinin yaygınlaşmasına karşın, hangi mimari ailelerin farklı kıyaslama koşullarında tespit performansı, yanlış pozitif oranı, saldırı türü ayrımı ve çıkarım hızı açısından en iyi dengeyi sunduğu hâlâ belirsizliğini korumaktadır. Özellikle üç çözüme kavuşturulamamış gerilim bu araştırmayı motive etmektedir.

Birinci gerilim, model karmaşıklığı ile genelleme yeteneği arasındaki ilişkiyi kapsamaktadır. CNN-BiLSTM gibi temsil kapasitesi daha yüksek mimariler sahiptir; ancak bu avantaj, örnek düzeyinde zamansal bağımlılıkların sınırlı olduğu tablosal ağ akışı verilerinde üstün performansa dönüşmeyebilir. İlave karmaşıklığın gerçek kazanımlar sağladığı durumları anlamak, pratik model seçimi açısından zorunludur.

İkinci gerilim, küresel doğruluk ile sınıf düzeyinde ayrım kapasitesi arasındaki ilişkiyi ele almaktadır. Sınıf dengesizliği varlığında yüksek genel doğruluk, operasyonel açıdan tam da en kritik olabilecek azınlık saldırı sınıflarında zayıf tespiti gizleyebilmektedir. Bu nedenle yalnızca doğruluk için optimizasyon yapmak yetersizdir ve titiz bir değerlendirme, tüm trafik kategorilerindeki performans dağılımını hesaba katmalıdır.

Üçüncü gerilim, klasik taban çizgilerinin rolünü kapsamaktadır. Derin öğrenme topluluğu zaman zaman klasik MÖ modellerini sinir ağı yaklaşımlarının gerisinde bırakılmış olarak değerlendirmiş olsa da özenle tasarlanmış öznitelikler sahip orta ölçekli veri kümelerinde Rastgele Orman ve XGBoost son derece rekabetçi olmayı sürdürmektedir. Derin öğrenmenin NIDS alanında bu taban çizgileri üzerinde açık ve ölçülebilir bir avantaj sağladığı koşulların açıkça araştırılması gerekmektedir.

1.5. Araştırma Soruları

Yukarıdaki problem formülasyonu ışığında bu tez, aşağıdaki araştırma sorularıyla yönlendirilmektedir:

AS1: CNN ve CNN-BiLSTM mimarileri hem ikili hem de çok sınıflı sınıflandırma ayarlarında CIC-IDS2017 ve UNSW-NB15 kıyaslama veri kümelerinde ne ölçüde yüksek doğruluk, düşük yanlış pozitif oranı ve güçlü saldırı türü ayrımı elde etmektedir?

AS2: Hibrit CNN-BiLSTM mimarisi, daha basit CNN mimarisine kıyasla istatistiksel olarak anlamlı ve operasyonel açıdan ilgili bir performans avantajı sağlamakta mıdır ve bu avantaj hangi veri kümesi ya da görev koşullarında en belirgin biçimde ortaya çıkmaktadır?

AS3: Derin öğrenme tabanlı NIDS modelleri, aynı kıyaslama veri kümesi üzerinde tespit performansı ve hesaplama verimliliği açısından iyi optimize edilmiş klasik makine öğrenmesi taban çizgileriyle (Rastgele Orman, XGBoost) nasıl karşılaştırılmaktadır?

Bu tezin geri kalanı aşağıdaki şekilde düzenlenmiştir. Tablo 1, her bölümün içeriğine genel bir bakış sunarak mevcut çalışmanın genel yapısını özetlemektedir.

1.6. Organizasyon

Tablo 1. Çalışmanın genel yapısını özetlemektedir.

Bölüm	İçerik
Bölüm 1	Giriş. Araştırma bağlamını, problem ifadesini, motivasyonları, araştırma sorularını ve tezin genel yapısını sunmaktadır.
Bölüm 2	Kuramsal Arka Plan. Makine öğrenmesinin temellerini, sinir ağı mimarilerini ve saldırı tespit sistemlerinin altında yatan temel kavramları kapsamlı biçimde incelemektedir. Denetimli öğrenme, öznitelik mühendisliği, sınıf dengesizliği işleme ve değerlendirme metriklerini ele almaktadır. Ayrıca, yazın özeti de bu bölümde sunulmuştur.
Bölüm 3	NIDS için Derin Öğrenme Mimarileri. Bu tezde incelenen CNN ve CNN-BiLSTM mimarilerini; tasarım ilkeleri, evrişimsel ve tekrarlayan bileşenler ve ağ trafiği verilerine uygulama gerekçesi dahil olmak üzere ayrıntılı biçimde tanımlamaktadır. Bir de veri kümeleri, ön işleme, mimari tanımları (CNN, CNN-BiLSTM, RF, XGBoost) ve eğitim yapılandırması.

Tablo 1. (Devamı)

Bölüm	İçerik
Bölüm 4	Deneysel Sonuçlar ve Analiz. Deneysel kurulumu, veri kümelerini, ön işleme ardışık düzenini ve tüm model yapılandırmalarının kapsamlı değerlendirmesini sunmaktadır. Sonuçlar, klasik MÖ taban çizgilerine ayrılmış bir alt bölümle birlikte her mimari, veri kümesi ve sınıflandırma görevi için raporlanmakta ve yorumlanmaktadır.
Bölüm 5	Sonuç ve Gelecek Çalışmalar. Temel bulguları sentezlemekte, araştırma sorularını yanıtlamakta, çalışmanın sınırlılıklarını kabul etmekte ve eşik optimizasyonu, mimari uzantılar ve gerçek dünya dağıtım değerlendirmeleri dahil olmak üzere gelecekteki araştırma yönlerini ortaya koymaktadır.

1.7. Etik ve Sürdürülebilirlik

Bu tezde yürütülen araştırma, tüm deneylerin yalnızca kamuya açık, önceden kaydedilmiş ağ trafiği veri kümeleri (CIC-IDS2017 ve UNSW-NB15) üzerinde gerçekleştirildiğinden, veri toplama veya insan deneklere ilişkin doğrudan bir etik endişe doğurmamaktadır. Bu veri kümeleri kişisel tanımlayıcı bilgi (KTB) içermemekte olup ilgili kurumlar tarafından açıkça siber güvenlik alanında akademik araştırma amacıyla yayımlanmıştır.

Bununla birlikte, daha kapsamlı bir etik ve sürdürülebilirlik değerlendirmesi açıkça kabul edilmeyi hak etmektedir. Algoritmik adalet açısından bu tezde benimsenen değerlendirme metrikleri özellikle makro ortalamalı F1 Skoru ve sınıf bazlı duyarlılık trafik kategorileri genelinde performans eşitsizliklerini açığa çıkarmak amacıyla tasarlanmıştır. Bu tercih, şeffaflığa yönelik bilinçli bir bağlılığı yansıtmaktadır: nadir ama operasyonel açıdan kritik saldırı sınıflarında sistematik olarak başarısız olan, ortalama doğruluğu yüksek bir model, küresel performansı ne olursa olsun kabul edilemez kabul edilmektedir.

2. ARKA PLAN

2.1. Giriş

Bağlı sistemlerin hızla çoğalması ve ağlar üzerinden akan verilerin üstel büyümesiyle birlikte, ağ altyapısını hedef alan siber saldırılar günümüzde kuruluşların karşı karşıya olduğu en acil tehditlerden biri hâline gelmiştir. Bu gerçekliğe yanıt olarak araştırma topluluğu, imza tabanlı yaklaşımlardan makine öğrenmesi (MÖ) ve daha yakın zamanda derin öğrenme (DÖ) temelli daha uyarlamalı yöntemlere doğru aşamalı bir evrim geçiren çok çeşitli saldırı tespit sistemleri (IDS) geliştirmiştir. Bu bölüm, mevcut yaklaşımların, literatürde en yaygın kullanılan veri kümelerinin, önceki çalışmalarda tespit edilen sınırlılıkların ve mevcut araştırmayı motive eden daha geniş bağlamın yapılandırılmış bir incelemesini sunmaktadır.

2.2. Ağ Trafiği Analizi ve Saldırı Tespit Sistemleri (IDS)

2.2.1. IDS'nin Tanımı ve Rolü

Saldırı Tespiti, sistem kaynaklarına yetkisiz erişim girişimlerini bulmak ve bunlara ilişkin gerçek zamanlı ya da gerçek zamana yakın uyarılar sağlamak amacıyla sistem olaylarını izleyen ve analiz eden bir güvenlik hizmetidir (Farhan vd., 2025a). Başka bir deyişle Saldırı Tespit Sistemi (IDS), bir bilgisayar sistemi veya ağ içindeki şüpheli ya da kötü niyetli etkinlikleri izlemek, analiz etmek ve tanımlamaktan sorumlu bir güvenlik mekanizması olarak genel biçimde tanımlanmaktadır. Birincil hedefi, kaynakların gizliliğini, bütünlüğünü veya erişilebilirliğini tehlikeye atabilecek anormal davranışları, güvenlik politikası ihlallerini ya da saldırı girişimlerini tespit etmektir.

Bir IDS, ağ trafiğini, sistem günlüklerini veya ana bilgisayar davranışını gerçek zamanlı ya da geriye dönük olarak inceleyerek bu bilgiyi bilinen saldırı imzalarıyla veya anormallikleri tespit eden istatistiksel/algoritmik modellerle karşılaştırır. Modern IDS'ler, doğruluğu artırmak, yanlış pozitifleri azaltmak ve yeni ya da karmaşık saldırıları belirlemek amacıyla giderek daha fazla makine öğrenmesi ve derin öğrenme gibi ileri teknikler kullanmaktadır.

2.2.2. IDS Türleri

2.2.2.1. Ağ Tabanlı Saldırı Tespit Sistemi

Ağ Tabanlı Saldırı Tespit Sistemi (NIDS), ağ trafiğini gerçek zamanlı olarak analiz etmek için tasarlanmış bir sistemdir. NIDS, trafiği gerçek zamanlı ya da gerçek

zamana yakın biçimde paket paket inceleyerek saldırı örüntülerini tespit etmeye çalışır. NIDS, ağ, taşıma ve/veya uygulama katmanı protokol etkinliğini inceleyebilir (Brown, 2008). NIDS'in temel özelliği, birden fazla ana bilgisayarı eş zamanlı olarak izleyebilmesi ve bunu diğer sistemlere kıyasla daha etkili kılmasıdır.

2.2.2.2. Ana Bilgisayar Tabanlı Saldırı Tespit Sistemi

Ana Bilgisayar Tabanlı Saldırı Tespit Sistemi (HIDS), doğrudan bir ana bilgisayara (sunucu, bilgisayar, IoT cihazı) kurulur. Günlük dosyaları, süreçler, dosya erişimi, oturum açma girişimleri ve sistem değişiklikleri gibi dahili sistem etkinliklerini analiz eder. Özünde, bir cihaza doğrudan gömülü bir ajan yazılımı olarak işlev görür. Bu sistem, günlük dosyaları, çalışan süreçler ve sistem dosyalarının bütünlüğü dahil olmak üzere çeşitli unsurlara yönelik derinlemesine analiz gerçekleştirir.

2.2.2.3. Hibrit IDS

Hibrit IDS, NIDS ve HIDS özelliklerini birleştirerek güvenliğe kapsamlı bir bakış açısı sunar. Hem ağ trafiğinden hem de dahili ana bilgisayar olaylarından elde edilen verileri ilişkilendirerek tespit doğruluğunu artırır ve yanlış pozitif sayısını azaltır. Birden fazla kaynaktan gelen verileri eş zamanlı olarak birleştirebilir; böylece NIDS ve HIDS'in bireysel sınırlılıklarını giderir.

2.2.3. Tespit Yöntemleri

2.2.3.1. Anomali Tabanlı

Anomali tespiti, bir veri kümesindeki örneklerin büyük çoğunluğundan belirgin biçimde sapan veri noktalarını belirleme sürecidir. Bu tür veri noktaları anomali olarak adlandırılır ve bunların tanımlanıp izole edilmesi, uygulamaların kararlılığını ve sağlamlığını kapsamlı biçimde artırabilir (Ioannidou, 2021). Gerçekte IDS, bu yönteme dayalı iki ilke doğrultusunda çalışır. Ağın normal davranışını öğrenir; bu profilden sapan her davranış şüpheli kabul edilir. Ne var ki sıfır gün saldırılarını tespit edebildiği hâlde yanlış pozitiflerin tespitinde zaman zaman yetersiz kalmaktadır.

2.2.3.2. İmza Tabanlı

İmza tabanlı tespit, ağ trafiğini veya sistem olaylarını bilinen saldırı örüntülerinden oluşan bir veritabanıyla karşılaştırmaya dayanır. Antivirüs yazılımına benzer şekilde bu yaklaşım, daha önce belgelenmiş kötü niyetli etkinliklerle birebir

eşleşme arar ve her imzayı belirli bir tehdide özgü kılar. Bir imzanın somut biçimi, tespit sisteminin konuşlandırıldığı ortama büyük ölçüde bağlıdır. Ağ tabanlı IDS'te imzalar genellikle paket düzeyinde gözlemlenebilir özellikleri tanımlar. Ana bilgisayar tabanlı IDS'te ise imzalar sistem düzeyinde çalışır; bir rootkit tarafından kurcalanmış kritik bir dosyanın kriptografik karması ya da bir denetim günlüğünde yakalanan anormal sistem çağrılarından oluşan tanınabilir bir sekans biçimini alabilir.

2.3. Saldırı Tespit Sistemi İçin Makine Öğrenmesi Yaklaşımı

Makine öğrenmesini açıklamanın en basit yolu şudur: bir bilgisayara izlemesi için açık kurallar yazmak yerine, ona veri verirsiniz ve kuralları kendi kendine çözmesine izin verirsiniz. Kavramsal olarak yalın görünen bu fikrin derin çıkarımları vardır. Bir makine öğrenmesi modeli ise örneklerden öğrenir. Örüntüler bulur, dahili temsiller oluşturur ve bu temsilleri daha önce görmediği yeni verilere ilişkin tahminler yapmak için kullanır.

Bu ayrım, özellikle siber güvenlik gibi alanlarda son derece önem kazanmaktadır. Her olası ağ saldırısı türünü tespit edecek kural tabanlı bir sistem oluşturmak pratik açıdan olanaksızdır; saldırılar evrim geçirir, her gün yeni varyantlar ortaya çıkar ve saldırganlar tekniklerini bilinçli olarak bilinen imzaları aşacak biçimde tasarlar. Makine öğrenmesi bir alternatif sunar: modeli normal ve kötü niyetli trafik örnekleri üzerinde eğiterek ikisi arasında ayrım yapmayı öğrenmesini sağlar.

Biçimsel olarak bir makine öğrenmesi algoritması, bir giriş uzayı X'i (paket boyutu, protokol türü, süre gibi öznitelikler) bir çıkış uzayı Y'ye (Normal veya Saldırı gibi etiket) eşleyen bir $f: X \rightarrow Y$ fonksiyonu olarak tanımlanabilir. Öğrenme süreci, bir eğitim veri kümesi üzerinde kayıp fonksiyonunu en aza indirerek f'nin en uygun parametrelerini bulmaktan oluşur.

Makine öğrenmesi, eğitim sırasında mevcut geri bildirim türüne bağlı olarak genel olarak üç temel paradigmaya ayrılır: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme. Her paradigma farklı türde problemlere uygundur ve aralarındaki farkları anlamak, doğru mimari seçimler yapmak açısından zorunludur.

- *Denetimli öğrenme:* Muhtemelen en yaygın kullanılan paradigma olan ve bu tezdeki çalışmaların büyük bölümünün temelini oluşturan denetimli öğrenmede temel fikir basittir: eğitim verisi etiket içerir. Her örnek, x'in giriş (öznitelik vektörü) ve y'nin beklenen çıkış (doğru sınıf veya değer) olduğu bir (x, y) çiftidir. Model, tahminleri ile gerçek etiketler arasındaki tutarsızlığı en aza indirecek şekilde eğitilir ve zamanla hiç görmediği örneklere genelleme

yapmayı öğrenir. NIDS bağlamında denetimli öğrenme, girdi olarak bir ağ akışını tanımlayan öznitelik vektörü alıp bu akışın normal mi yoksa belirli bir saldırı türüne mi karşılık geldiğini gösteren bir etiket üreten bir sınıflandırıcıyı eğitmek için kullanılır. (Moustafa & Slay, 2015) ; (Sharafaldin vd., 2018) yayınlarında UNSW-NB15 ve CIC-IDS2017 gibi veri kümeleri bu amaç için özel olarak tasarlanmış; hem iyi huylu hem de saldırı örneklerini birden fazla saldırı kategorisi boyunca etiketli ağ trafiği biçiminde sunmuşlardır.

- *Denetimsiz öğrenme:* Denetimsiz öğrenme tamamen farklı bir yaklaşım benimser: etiket yoktur. Algoritma yalnızca giriş verisi X'i alır ve yapıyı kendi başına keşfetmek zorundadır. Girişlerden bilinen çıktılara yönelik bir eşleme öğrenmek yerine model, verilerin altta yatan dağılımını anlamaya çalışır; kümeler bulur, boyutları azaltır veya sıkıştırılmış temsiller öğrenir. Siber güvenlikte denetimsiz yöntemler, özellikle anomali tespiti için değerlidir. Bilinen saldırı imzalarına karşı trafiği sınıflandırmak yerine, bir anomali dedektörü normal trafiğin nasıl göründüğünü öğrenir ve bu modelden belirgin biçimde sapan her şeyi işaretler. Bu yaklaşımın önemli bir avantajı, hiçbir etiketli eğitim kümesinde yer alamayacak sıfır gün saldırılarını tespit edebilmesidir. NIDS araştırmalarında kullanılan yaygın denetimsiz teknikler.
- *Pekiştirmeli öğrenme:* Pekiştirmeli öğrenme, temelden farklı bir ilkeyle çalışan üçüncü paradigmadır. Bir ajan bir ortamla etkileşime girer, eylemler gerçekleştirir ve bu eylemlerin sonucuna bağlı olarak ödüller veya cezalar alır. Amaç, zaman içindeki kümülatif ödülü en üst düzeye çıkaracak bir politika durumlardan eylemlere yönelik bir eşleme öğrenmektir.

Bu bölümde, önerilen NIDS'in taban çizgisini oluşturan iki makine öğrenmesi algoritması sunulmaktadır: Rastgele Orman ve XGBoost.

Tablo 2. Bu tablo, üç temel makine öğrenmesi paradigmasının ve NIDS'teki uygulamalarının kısa bir karşılaştırmasını sunmaktadır

Paradigma	Etiketler	Amaç	NIDS Uygulaması
Denetimli	Evet	Giriş-çıkış eşlemesini öğrenme	Saldırı sınıflandırma (ikili / çok sınıflı)
Denetimsiz	Hayır	Gizli yapıyı keşfetme	Anomali tespiti, sıfır gün tehditleri
Pekiştirmeli	Ödül sinyali	En uygun politikayı öğrenme	Uyarlamalı IDS, otomatik yanıt

Ağ tabanlı saldırı tespiti bağlamında çeşitli klasik makine öğrenmesi algoritmaları yaygın biçimde değerlendirilmiştir. Karar Ağaçları hızlı ve yorumlanabilir olmaları nedeniyle denetimli öğrenmede güçlü bir başlangıç noktası sunar. Ancak aşırı öğrenmeye eğilimlidirler; bu da KDD veri kümesinde makul bir taban çizgisi sağlamalarına karşın iyi genelleme yapamadıkları anlamına gelir. Rastgele Orman, birden fazla ağacı bir topluluk hâlinde birleştirerek bu sorunu çözer ve gürültülü verilere karşı daha sağlam hâle gelir. Ödünleşim ise yorumlanabilirliğin azalmasıdır. CIC-IDS2017 üzerinde %97'nin üzerinde doğruluk elde ederek en güçlü klasik yöntemlerden biri olmaya devam etmektedir. XGBoost da topluluk artırma kullanır ve aşırı öğrenmeyi kontrol altında tutmak için düzenlileştirme ekleyerek özellikle çok sınıflı NIDS sınıflandırma görevlerinde son derece etkili olmaktadır.

2.4. IDS İçin Çeşitli Derin Öğrenme Modellerinin Değerlendirilmesi

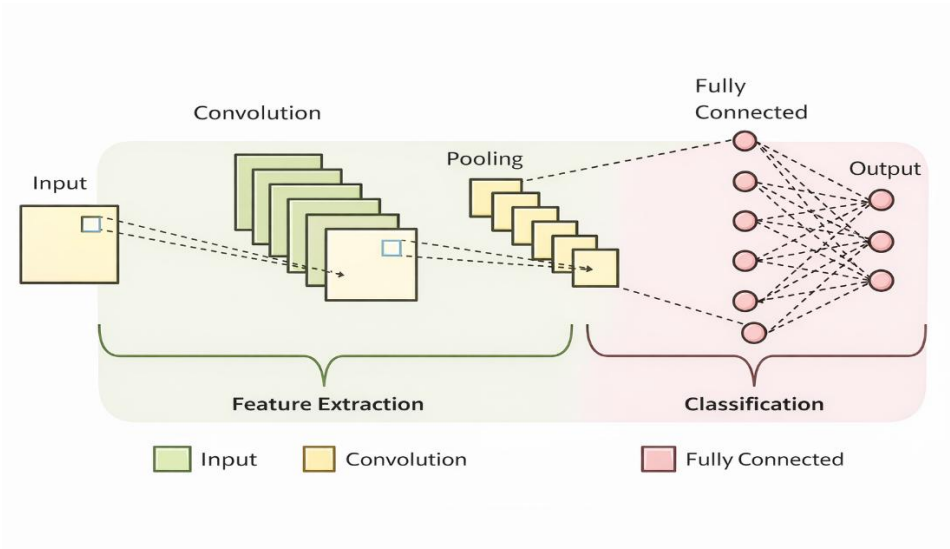
Derin öğrenme, birden fazla işlem katmanından oluşan yapay sinir ağlarına dayanan bir makine öğrenmesi alt alanıdır. (Krizhevsky vd., 2012), 2012 ImageNet yarışmasında modern derin öğrenme çağını başlattığı yaygın biçimde kabul edilen AlexNet'i tanıtmış; tüm rekabetçi klasik yöntemler üzerinde dikkat çekici bir performans farkı ortaya koymuştur. O tarihten itibaren derin öğrenme, karmaşık ve yüksek boyutlu verileri içeren hemen her görev için başvurulan yaklaşım hâline gelmiştir. Derin öğrenmeyi klasik makine öğrenmesinden ayırt eden temel özellik, hiyerarşik temsilleri doğrudan ham verilerden otomatik olarak öğrenme yeteneğidir. Bu bölümde, önerilen NIDS'in temelini oluşturan iki derin öğrenme mimarisi sunulmaktadır: Evrişimli Sinir Ağı (CNN) ve Çift Yönlü Uzun-Kısa Süreli Bellek ağı (BiLSTM). Bölüm, CNN ve BiLSTM'nin tek bir hibrit modelde nasıl birleştirildiğinin açıklanmasıyla son bulmaktadır.

2.4.1. Evrişimli Sinir Ağları (CNN)

Evrişimli Sinir Ağları başlangıçta görüntü işleme görevleri için tasarlanmıştır. Ağ tabanlı saldırı tespitinde ise giriş bir görüntü değil, bir ağ akışının öznitelik vektör temsilidir. Bu bağlamda uygulama için 2D evrişimler kolayca 1D ile değiştirilir; bir evrişimsel filtre öznitelik vektörü üzerinde kayar ve komşu öznitelikler arasındaki yerel korelasyonları bulur. Örneğin, yalnızca bir port tarama saldırısında gözlemlenebilecek yüksek port numarası, küçük akış süresi ve büyük paket sayısı kombinasyonu başlı başına bir filtre hâline gelebilir.

2.4.2. Mimari Bileşenler

NIDS için temel CNN, üst üste yığılmış bileşenlerden oluşur. Evrişimsel katman bazı filtreler alır (bunlar öğrenilebilir) ve bunları girişe uygulayarak belirli örüntüleri sergileyen öznitelik haritaları üretir; bu yamalar, farklı örüntü türlerine göre birden fazla filtre tarafından paralel olarak tespit edilir. Toplu Normalleştirme, mini toplu iş genelinde aktivasyonları normalleştirerek her evrişimsel katmanı ayırır; bu da eğitimi kararlı kılar ve yakınsamayı hızlandırır. Her öznitelik haritası için, boyutu azaltırken maksimum değeri kayan bir pencerede alan ve hesaplama maliyetini düşürürken öteleme değişmezliği özelliği sunan Maksimum Havuzlama uygulanır.



Şekil 1. NIDS bağlamında Evrişimli Sinir Ağı

CNN'in NIDS bağlamındaki temel gücü, elle öznitelik mühendisliği gerektirmeksizin ham trafik istatistiklerinden ilgili öznitelik kombinasyonlarını otomatik olarak çıkarma yeteneğidir. Temel sınırlılığı ise ardışık ağ akışları arasındaki zamansal bağımlılıkları doğal olarak modelleyememesidir; bu sınırlılığı gidermek amacıyla özel olarak tasarlanan LSTM mimarisi bir sonraki bölümde sunulmaktadır.

Öte yandan, bir IDS içinde CNN kullanan saldırı tespit mekanizması üç ana aşamada işler: otomatik öznitelik çıkarma, derin temsil öğrenme ve trafik sınıflandırma.

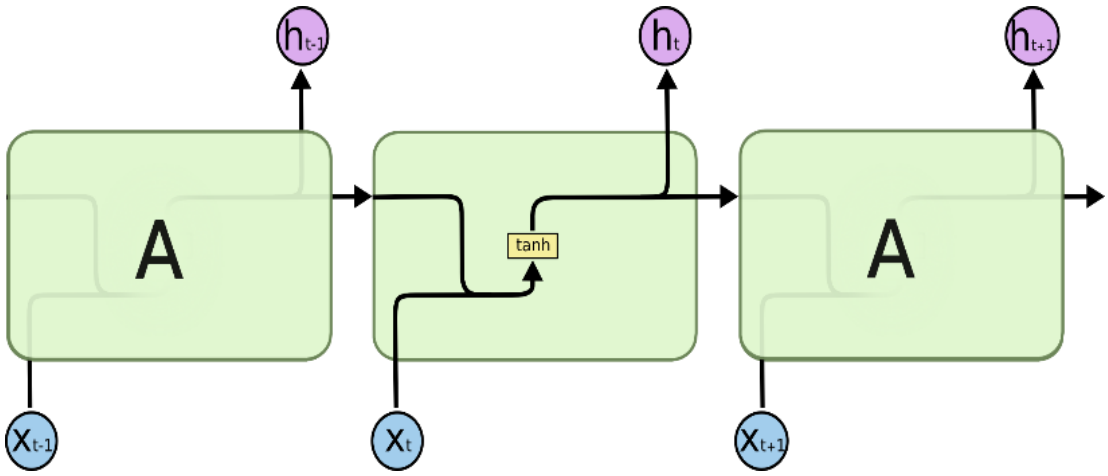
2.4.3. Tekrarlayan Sinir Ağları (RNN)

2.4.3.1. Motivasyon: Standart RNN'lerin Sınırları

Ağ trafiği özünde ardışıklıdır. Paketler belirli bir sırayla gelir ve bir bağlantının zaman içindeki davranışı nasıl başladığı, nasıl geliştiği, nasıl sonlandığı kötü niyetli mi

yoksa iyi niyetli mi olduğuna ilişkin kritik bilgiler taşır. Bir DoS saldırısı, örneğin, tek bir şüpheli paket ile değil; zaman içinde gelişen sürekli yüksek hacimli trafik örüntüsü ile karakterize edilir. Bu tür zamansal bağımlılıkları yakalamak için zaman adımları genelinde belleği koruyabilen bir modele ihtiyaç duyulmaktadır.

Tekrarlayan Sinir Ağları (RNN) bu amaç için tasarlanmıştır. Her zaman adımında bir RNN, mevcut gözlemi ve önceki adımdan gelen gizli durumu girdi olarak alarak geçmiş tüm bilgileri özetleyen yeni bir gizli durum üretir. Prensip itibarıyla bu, ağın sıranın bellekten koşan bir özetini tutmasına olanak tanır. Ancak pratikte standart RNN'ler kaybolan gradyan sorunuyla karşı karşıyadır: zaman içinden geriye yayılım sırasında gradyanlar, birçok zaman adımından geçerken üstel olarak küçülme eğilimindedir; bu da uzun vadeli bağımlılıkları öğrenmeyi son derece güçleştirir. 50 paket önce başlamış bir ağ akışını işleyen bir RNN, o erken paketlere ilişkin bilgiyi neredeyse tamamen yitirmiş olacaktır.



Şekil 2. Bu şekil, basit bir Tekrarlayan Sinir Ağının gösterimini sunmaktadır.

1997 yılında Hochreiter ve Schmidhuber, Tekrarlayan Sinir Ağına yeni bir mimari tanıtmış ve Uzun-Kısa Süreli Bellek ağını (LSTM) önererek standart RNN'lerdeki kaybolan gradyan sorununu ele almıştır. Bir LSTM ağında her hücre yalnızca RNN çıktısına eşdeğer bir gizli durumu (h_t) değil, aynı zamanda zaman adımları boyunca standart RNN'lerdeki gibi çarpımsal değil, toplamalı etkileşimler aracılığıyla güncellenen bir hücre durumunu (C_t) da barındırır.

LSTM ağları, üç kapı (*z, i, f*) aracılığıyla dahili durumların zaman sabitlerini kontrol eder; bu kapılar sigmoid aktivasyonlu doğrusal katmanlar olarak uygulanır.

Unut Kapısı, önceki hücre durumunun ne kadarının atılacağına karar verir. Çıktısı 0'a yakın olduğunda bellek silinir; 1'e yakın olduğunda bellek tamamen korunur. Ağ

trafiği açısından unut kapısı, bir bağlantının sona erip yeni bir bağlantının başladığını tespit ettiğinde belleğini sıfırlamayı öğrenebilir.

Giriş Kapısı, mevcut girdiden ne kadar yeni bilginin hücre durumuna dahil edileceğini kontrol eder. Bu iki aşamada hesaplanır: hangi değerlerin güncelleneceğine karar veren bir sigmoid kapısı ve potansiyel olarak eklenecek aday değerler üreten bir dönüşüm.

Çıkış Kapısı, hücre durumunun ne kadarının mevcut zaman adımında gizli durum olarak ortaya çıkacağını belirler. Bu gizli durum, ağın bir sonraki katmana veya çıktıya ilettiği değerdir.

İleri yönlü bir LSTM hücresinin altı temel denklemi şöyle özetlenebilir:

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right) \quad \text{(Unut kapısı)} \quad (1)$$

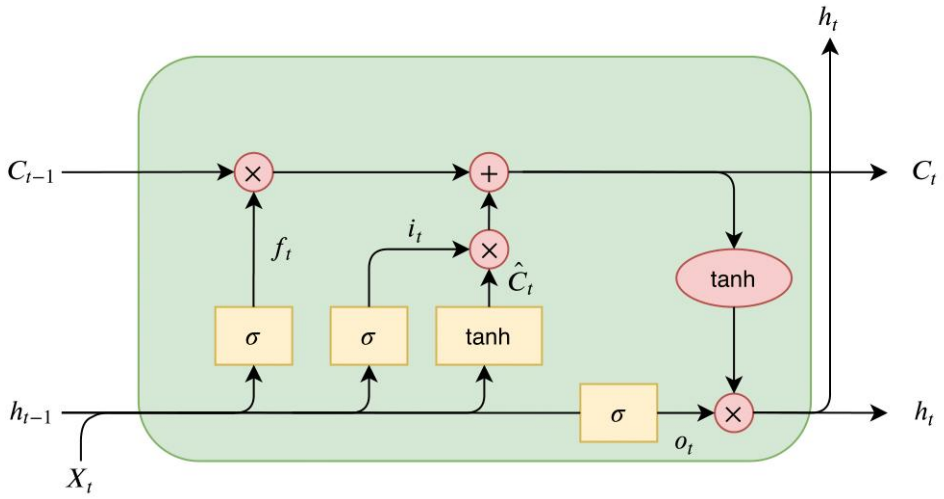
$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad \text{(Giriş kapısı)} \quad (2)$$

$$\widetilde{C_t} = tanh(W_C . [h_{t-1}, x_t] + b_c) \quad \text{(Aday hücre durumu)} \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \quad \text{(Güncellenmiş hücre durumu)} \quad (4)$$

$$O_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \quad \text{(Çıkış kapısı)} \quad (5)$$

$$h_t = ot \odot \tanh(C_t) \quad \text{(Gizli durum çıkısı)} \quad (6)$$



Şekil 3. LSTM'nin (Uzun-Kısa Süreli Bellek) dahili mimarisi.

Bu mimari, LSTM'ye uzun menzilli bağımlılıkları öğrenmede olağanüstü bir yetenek kazandırır. BiLSTM'yi KDDCUP-99 ve UNSW-NB15 üzerinde özel olarak test eden bir çalışma, LSTM tabanlı modellerin her iki veri kümesinde de %99 doğruluk

elde ederek özellikle azınlık saldırı sınıflarında sığ makine öğrenmesi yaklaşımlarını geride bıraktığını göstermiştir (Vinayakumar vd., 2017).

2.4.3.2. Çift Yönlü LSTM (BiLSTM)

Standart LSTM ağları, dizileri ilk zaman adımından son zaman adımına kadar işler. Çift Yönlü LSTM ağları ise aynı dizileri hem geriye hem ileriye doğru işler ve her zaman adımında her iki ağ da çıktıya katkıda bulunur. İleri LSTM'ye (ileri beslemeli ağ) ek olarak bir geri LSTM (geri bildirimli ağ) de bulunur. Herhangi bir zaman adımının çıktısı, iki ağın birleşimidir.

Tekrarlayan sinir ağlarına ilişkin mevcut literatürde, bir dizideki tüm zaman adımları eşit derecede önemli kabul edilmekte ve dizideki her eleman yalnızca öncesi ve kendinden önceki tüm elemanlar tarafından tanımlanmaktadır. Bu çalışmada RNN mimarisinde yeni bir yön araştırılmaktadır: ağın bir dizideki bir elemandan sonra gelenlere dikkat etme olasılıkları incelenmektedir. Bu tür bir mimari için pek çok uygulama mevcuttur örneğin doğal dil işleme, kelimeleri hem önceki hem de sonraki komşularıyla modelleyerek yapılabilir. Ağ trafiği de bu şekilde modellenebilir: bir paket patlaması (trafik artışı) iyi niyetli mi (örneğin bir bağlantı sonlandırmanın parçası) yoksa kötü niyetli mi (örneğin yanal hareketin göstergesi), Geriye dönük bir LSTM hücresinin altı temel denklemi şöyle özetlenebilir:

$$f_t = \sigma\left(W_f.[h_{t+1}, x_t] + b_f\right) \quad \text{(Unut kapısı)} \quad (7)$$

$$i_t = \sigma(W_i.[h_{t+1}, x_t] + b_i) \quad \text{(Giriş kapısı)} \quad (8)$$

$$\widetilde{C_t} = tanh(W_C.[h_{t+1}, x_t] + b_c) \quad \text{(Aday hücre durumu)} \quad (9)$$

$$C_t = f_t \odot C_{t+1} + i_t \odot \widetilde{C_t} \quad \text{(Güncellenmiş hücre durumu)} \quad (10)$$

$$O_t = \sigma(W_o.[h_{t+1}, x_t] + b_o) \quad \text{(Çıkış kapısı)} \quad (11)$$

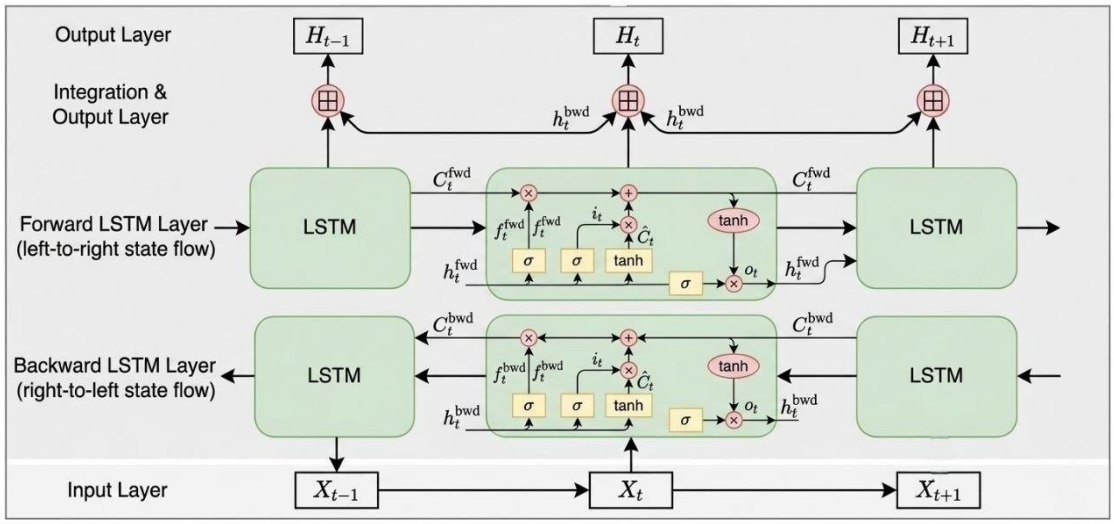
$$h_t = ot \odot \tanh(C_t) \quad \text{(Gizli durum çıktısı)} \quad (12)$$

Unut kapısı $f_t = \sigma\left(W_f.[h_{t+1}, x_t] + b_f\right)$, önceki hücre durumundan hangi bilgilerin atılacağına karar verir. Sigmoid fonksiyonu σ, 0 ile 1 arasında bir değer üretir. Formül şu şekilde verilmektedir:

$$\sigma = \frac{1}{1+e^{-x}}. \quad (13)$$

Aday Hücre Durumu $\widetilde{C_t} = tanh(W_C.[h_{t+1}, x_t] + b_c)$, hücre durumuna eklenebilecek yeni aday değerleri hesaplar. $tanh$ Fonksiyonu, yeni bilginin yönünü ve büyüklüğünü kodlayarak değerleri -1 ile +1 arasında sıkıştırır. Formül şu şekilde verilmektedir:

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}. \tag{14}$$



Şekil 4. Çift yönlü uzun-kısa süreli bellek (BiLSTM).

2.4.3.3. NIDS Literatüründe BiLSTM

Standart LSTM üzerinde BiLSTM'nin NIDS için üstünlüğü birden fazla çalışmayla doğrulanmıştır.(Udurume vd., 2024), NSL-KDD ve UNSW-NB15 üzerinde kapsamlı karşılaştırmalı bir çalışma sunmuş; CNN-BiLSTM modelinin sırasıyla %99,89 ve %98,95 doğruluk elde ederek geleneksel MÖ yöntemlerini %1,5 ila %3,5 arasında değişen bir farkla geride bıraktığını bildirmiştir. (Peng vd., 2023) ise sınıf dengesizliğini ele almak amacıyla CNN, BiLSTM ve odak kaybını birleştiren CBF-IDS modelini önermiş; NSL-KDD, UNSW-NB15 ve CIC-IDS2017 olmak üzere üç kıyaslama veri kümesinde eş zamanlı olarak rekabetçi sonuçlar elde etmiştir.

Mevcut çalışma açısından özellikle ilgili bir çalışma, UNSW-NB15 veri kümesi üzerinde öznitelik seçiminde ki-kare kategorik öznitelik puanlaması ile minimal CNN'i BiLSTM ile birleştirmiş ve bu kombinasyonun yüksek tespit doğruluğunu korurken çıkarım süresini azalttığını göstermiştir. Hibrit modeller günümüzde hâlâ kullanılmakla birlikte, derin öğrenme tabanlı saldırı tespitini anlamak için temel bir zemin oluşturmaktadır.

2.4.4. Hibrit Modeller: CNN-LSTM

2.4.4.1. Hibrit Yaklaşımın Gerekçesi

CNN ve LSTM'yi ayrı ayrı inceledikten sonra, bunları birleştirmenin her birinden tek başına daha iyi sonuçlar verip vermeyeceğini sormak doğaldır. Son literatürdeki kayda değer miktarda çalışmaya göre yanıt evettir; bunun nedeni ise iki mimarinin gerçekten tamamlayıcı nitelikte olmasıdır.

CNN, tek bir trafik akışındaki öznitelikler arasındaki yerel örüntüleri ve uzamsal korelasyonları tespit etmede üstündür. Boyutsallığı azaltır, gürültüyü filtreler ve sıkıştırılmış, bilgi yüklü bir öznitelik temsili üretir. Ancak her akışı izole bir gözlem olarak ele alır ve birden fazla akış üzerinde ağ davranışının zamansal evrimini modellemez.

LSTM ise zaman adımları arasındaki ardışık bağımlılıkları modellemede üstündür. Ancak doğrudan ham ve yüksek boyutlu öznitelikleri girdi olarak aldığında, en ilgili örüntüleri tanımlamakta güçlük çekebilir ve hesaplama açısından maliyetli olabilir.

Bir CNN bloğunun çıktısını bir LSTM bloğuna besleyerek her ikisinin de avantajlarını elde ederiz: CNN her akıştan sıkıştırılmış ve anlamlı öznitelik temsilleri çıkarır; LSTM ise bu temsiller arasındaki zamansal ilişkileri modeller. Bu hiyerarşik öznitelik çıkarma ardışık düzeni, yakın dönem NIDS araştırmalarında baskın paradigmalardan biri hâline gelmiştir.

2.4.4.2. CNN-BiLSTM Modelimizin Mimarisi

Önerilen CNN-BiLSTM modelimiz dört aşamadan oluşmaktadır. Birinci aşamada, ön işlenmiş ağ akışı öznitelikleri CNN bloğuna girdi olarak verilir. CNN omurgası, artık yapılara sahip iki ayrı 1D evrişimsel bloktan oluşmaktadır. Her blok, artan kanal sayılarına ($1 \rightarrow 64 \rightarrow 128$) sahip iki 1D evrişimsel katman ve ardından bir Toplu Normalleştirme katmanı, ReLU aktivasyonu, MaksimumHavuzlama katmanı ve 0,3'lük bir Hattan Düşürme katmanı içermektedir.

İkinci aşamada, CNN bloğunun çıktısı BiLSTM bloğuna beslenir. BiLSTM, her yönde 128 gizli nörona sahip olup bu sıkıştırılmış öznitelik temsilini her iki yönde işleyerek ağ trafiği dizisindeki zamansal bağımlılıkları yakalar. İleri ve geri LSTM çıktıları her zaman adımında birleştirilerek 256 boyutlu birleşik bir temsil oluşturulur. LSTM çıktısına 0,3'lük bir Hattan Düşürme uygulanmaktadır.

Üçüncü aşamada, BiLSTM çıktısı iki tam bağlantılı katmandan (256 birim → 128 birim) oluşan Yoğun bloğa iletilir. Bu katmanlar ReLU aktivasyonu ve 0,3'lük Hattan Düşürme ile düzenlileştirilmiştir.

Son olarak çıkış katmanı, hedef sınıflar üzerinde olasılık dağılımları üretmek amacıyla Softmax aktivasyonu kullanır; bu, ikili sınıflandırma için 2 sınıf (Normal/Saldırı), UNSW-NB15 üzerinde çok sınıflı sınıflandırma için 10 sınıf veya CIC-IDS2017 için 15 sınıf olabilir.

2.5. İlgili Teknolojinin Geldiği Son Nokta

2.5.1. Genel Bakış ve Bağlam

Son on yılda ağ tabanlı saldırı tespiti, kural tabanlı yöntemlerden veri güdümlü yöntemlere doğru evrim geçirmiştir. Öncelikle karar ağaçları, rastgele ormanlar, destek vektör makineleri (DVM) ve AdaBoost, torbalama ve yığma gibi topluluk öğrenme yöntemleri gibi klasik makine öğrenmesi teknikleri geliştirilmiş ve kullanılmıştır. Ardından evrişimli sinir ağları (CNN), tekrarlayan sinir ağları (RNN), tam bağlantılı sinir ağları (TBSN), artık sinir ağları (ResNet), otokodlayıcılar, üretici çekişmeli ağlar (GAN) ve sinir ağı toplulukları dahil olmak üzere derin öğrenme tabanlı çözümlere olan ilgi önemli ölçüde artmıştır. Son dönemde ise makine öğrenmesi ve derin öğrenmeyi birleştiren hibrit modeller, dikkat mekanizmaları ile güçlendirilmiş mimariler ve grafik sinir ağlarına yönelik ilk girişimler literatürde yer almaya başlamıştır. Çalışılan ve ayrıntılı biçimde ele alınan çalışmalar mimariler, veri kümeleri ve yaklaşımlar açısından bu tezle ilişkilidir.

Derin öğrenme yaklaşımlarına geçmeden önce, klasik makine öğrenmesi yöntemlerinin kalıcı önemini kabul etmek yerinde olacaktır. (Türk, 2023), hem UNSW-NB15 hem de NSL-KDD veri kümeleri üzerinde birkaç makine öğrenmesi algoritmasını kapsamlı biçimde değerlendirmiş; UNSW-NB15 üzerinde iki sınıflı ve çok sınıflı sınıflandırma için sırasıyla %98,6 ve %98,3, NSL-KDD üzerinde ise %97,8 ve %93,4 doğruluk elde etmiştir. Bu çalışma, mevcut tezde kullanılan aynı veri kümeleri üzerinde doğrudan bir taban çizgisi sunması nedeniyle özellikle ilgilidir.

2.5.2. Tek Modelli ve Görece Basit Derin Öğrenme Yaklaşımları

Derin öğrenmeyi kullanmanın en doğrudan motivasyonlarından biri, ham trafik verilerinden hiyerarşik öznitelikleri otomatik olarak çıkarma yeteneğidir.

Özellikle UNSW-NB15 üzerinde (Farhan vd., 2025b), Scientific Reports'taki bir çalışma, Ekstra Ağaç Sınıflandırıcı tabanlı öznitelik seçimi ile birleştirilmiş ReLU aktivasyonu kullanan ardışık Derin Sinir Ağı (DSA) mimarisi önermiş; önceki DSA tabanlı yaklaşımların sigmoid veya tanh aktivasyon fonksiyonları nedeniyle kaybolan gradyan sorununa yol açtığını ve ReLU'nun hedefli öznitelik seçimiyle birleştirildiğinde bu veri kümesinde yorumlanabilirliği ve hesaplama verimliliğini önemli ölçüde artırdığını ileri sürmüştür. Bu mimari tercih, alanın genel eğilimiyle örtüşmektedir ve ön işleme adımı olarak öznitelik seçiminin vurgulanması, kendi UNSW-NB15 deneyimlerimizde kullanılan 40 öznitelikli ardışık düzen ile doğrudan ilişkilidir.

2.5.3. CIC-IDS2017 Üzerinde Öznitelik Seçimi ve Hibrit Derin Öğrenme

Birçok çalışma özellikle CIC-IDS2017 üzerine odaklanmış ve derin öğrenmeyi açık öznitelik optimizasyonu ile birleştirmiştir. 2023 tarihli bir çalışma, CIC-IDS2017 üzerinde öznitelik seçimiyle birlikte CNN-GRU hibrit mimarisini önermiş; üç CNN katmanı ve iki GRU katmanını gereksiz özniteliklerin kaldırılmasıyla birleştirmiştir. Elde edilen model, 0,075'lik YPO (Yanlış Pozitif Oranı) ile %98,73 doğruluk elde etmiştir. (Henry vd., 2023), bu yaklaşım dikkat çekicidir; çünkü öznitelik seçimini bir ön işleme eki olarak değil, modelin ayrılmaz bir parçası olarak ele almaktadır. Bu, kendi çalışmamızın kantil normalleştirme ve evrişimsel katmanların doğasında bulunan öznitelik seçme kapasitesi aracılığıyla kısmen ele aldığı bir tasarım stratejisidir. Çalışma aynı zamanda GRU'nun, tıpkı BiLSTM gibi, saf CNN mimarilerinin karşı karşıya olduğu uzun vadeli bağımlılık sorununa çözüm sunduğunu vurgulamaktadır. Kendi çalışmamızdaki fark, GRU yerine BiLSTM tercihidir. Her ikisi de geçerli seçeneklerdir; ancak BiLSTM'nin çift yönlü işlemi, biraz daha yüksek hesaplama maliyeti pahasına daha zengin bağlam sağlamaktadır.

2.5.4. Topluluk Yöntemleri ve Klasik Yaklaşımların Yeniden Değerlendirilmesi

Son dönemdeki tüm ilerlemeler yalnızca derin öğrenmeden gelmemektedir. Topluluk yöntemleri, özellikle yorumlanabilirlik ve eğitim verimliliğinin öncelik taşıdığı durumlarda yüksek rekabetçiliğini sürdürmektedir. 2025 tarihli bir çalışma, CIC-IoT-2023 üzerinde IoT saldırı tespiti için hibrit İleri Beslemeli Sinir Ağı (TBSN)-

XGBoost çerçevesi önermiş; boyutsallık azaltma için temel bileşen analizi (TBA) ile iki aşamalı yeniden örnekleme stratejisini bir araya getirmiş ve hem ikili hem de çok sınıflı sınıflandırma için %99 doğruluk elde etmiştir. (Alashjaee & Alqahtani, 2025), buradaki temel içgörü, derin öznitelik çıkarma (TBSN) ile gradyan artırma (XGBoost) arasındaki tamamlayıcılıktır: sinir ağı karmaşık temsilleri öğrenirken XGBoost yüksek hassasiyetli ayrım sağlar. Bu hibrit felsefe, uçtan uca derin öğrenmeden mimari olarak ayrışmakta ancak karşılaştırılabilir sonuçlar elde etmektedir.

2.5.5. Sentez ve Mevcut Çalışmanın Konumlandırılması

Mevcut çalışma, tutarlı bir ön işleme ve değerlendirme protokolü çerçevesinde iki kıyaslama veri kümesi olan CIC-IDS2017 ve UNSW-NB15 üzerinde CNN ile CNN-BiLSTM'yi çok sınıflı sınıflandırma, 5-katlı çapraz doğrulama ve çıkarım işlem hacmi ile YPO dahil olmak üzere operasyonel metriklere odaklanarak titiz ve yan yana karşılaştırarak bu literatüre katkıda bulunmaktadır. İncelenen literatürde bu titizlikle nadiren gerçekleştirilen bu sistematik karşılaştırma, gerçek dünya dağıtım kısıtlamaları altında mimariler arasında seçim yapan uygulayıcılara uygulanabilir içgörüler sunmaktadır.

3. YÖNTEM VE ÖNERİLEN YAKLAŞIM

3.1. Veri Kümelerinin Tanımı

Bu çalışma, önerilen derin öğrenme tabanlı saldırı tespit sistemini değerlendirmek amacıyla yaygın olarak kullanılan iki kıyaslama veri kümesini UNSW-NB15 ve CIC-IDS2017 kullanmaktadır. Bu veri kümeleri, gerçekçilikleri, saldırı senaryolarının çeşitliliği ve saldırı tespit araştırmalarında yaygın kullanımları nedeniyle seçilmiştir.

3.1.1. UNSW-NB15 Veri Kümesi

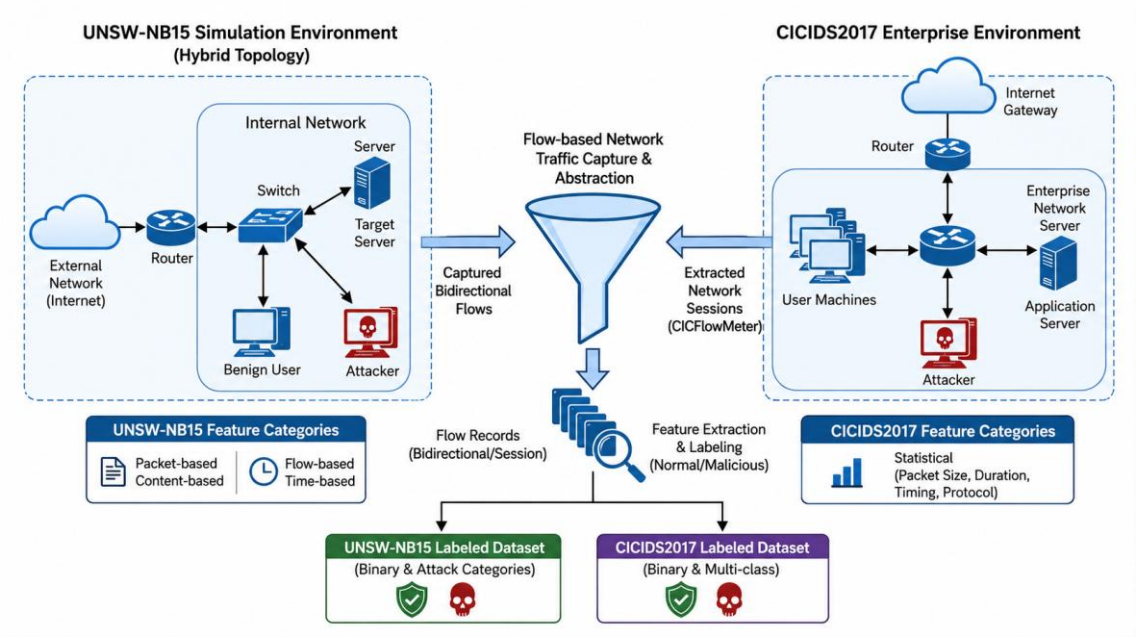
UNSW-NB15 veri kümesi, gerçekçi ağ trafiği üretmek amacıyla IXIA PerfectStorm aracı kullanılarak Avustralya Siber Güvenlik Merkezi tarafından geliştirilmiştir. DoS, Açıklar, Fuzzer, Keşif ve Arka Kapılar gibi birden fazla modern saldırı kategorisini kapsayan hem normal hem de kötü niyetli etkinlikleri içermektedir. Veri kümesi, sayısal ve kategorik özniteliklerle tanımlanan akış tabanlı kayıtlar içermekte olup hem ikili hem de çok sınıflı saldırı tespit görevlerini desteklemektedir.

3.1.2. CIC-IDS2017 Veri Kümesi

CIC-IDS2017 veri kümesi, eski saldırı tespit veri kümelerinin sınırlılıklarını gidermek amacıyla Kanada Siber Güvenlik Enstitüsü tarafından oluşturulmuştur. Gerçek dünya ağ trafiğini simüle etmekte ve kaba kuvvet, DoS/DDoS, web saldırıları, botnet ve sızma gibi çeşitli saldırı türlerini kapsamaktadır. Veri kümesi, CICFlowMeter kullanılarak çıkarılan zengin akış düzeyi öznitelikler sunmakta ve derin öğrenme modellerinin değerlendirilmesine uygun hâle gelmektedir.

3.1.3. Veri Kümesi Seçiminin Gerekçesi

UNSW-NB15 ve CIC-IDS2017, ağ trafiği ve saldırı davranışı üzerine tamamlayıcı bakış açıları sunmaktadır. Bu iki veri kümesinin bir arada kullanılması, derin öğrenme tabanlı saldırı tespit modellerinin kapsamlı ve güvenilir bir şekilde değerlendirilmesini mümkün kılmaktadır. Trafik üretimi ve veri toplama için kullanılan deneysel ağ ortamına ilişkin net bir anlayış sağlamak amacıyla, Şekil 5'te ilgili ağ topolojisi gösterilmektedir.

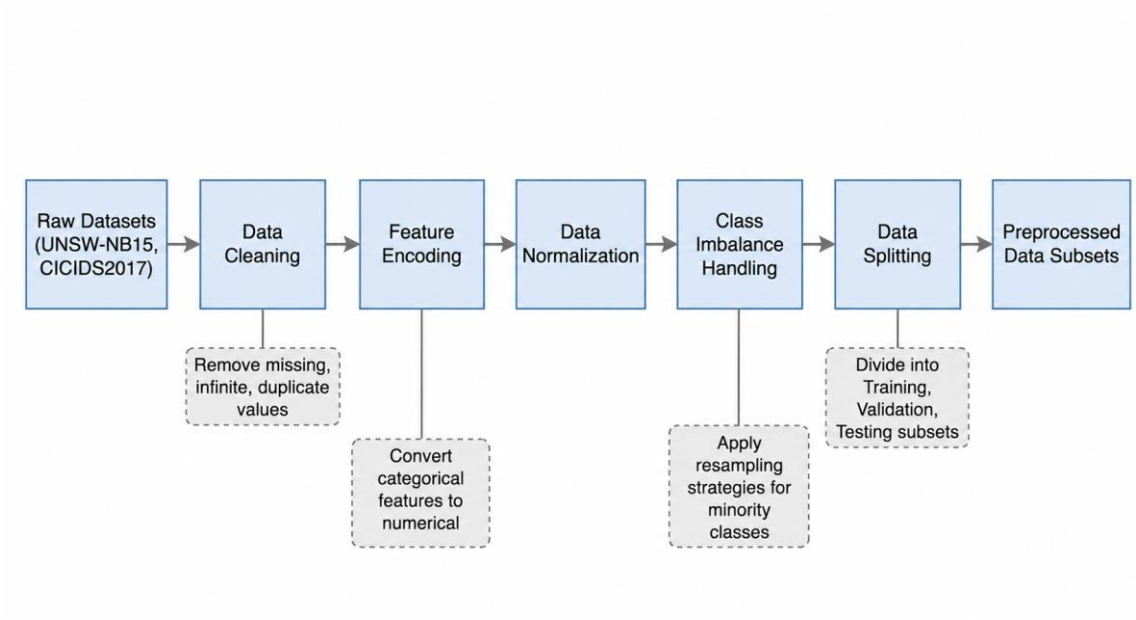


Şekil 5. Hibrit ağ topolojisi ve akış tabanlı trafik analizi (UNSW-NB15 ve CIC-IDS2017 kavramları).

3.2. Veri Ön İşleme

Önerilen saldırı tespit sisteminin güvenilirliğini ve etkinliğini sağlamak amacıyla, hem UNSW-NB15 hem de CIC-IDS2017 veri kümelerine yapılandırılmış bir veri ön işleme ardışık düzeni uygulanmaktadır. Bu ön işleme süreci beş ana aşamadan oluşmaktadır: veri temizleme, öznitelik kodlama, veri normalleştirme, sınıf dengesizliği giderme ve veri bölme.

Bu bölüm, her iki ikili ve çok sınıflı sınıflandırma görevinde modellerimizi eğitmek amacıyla CIC-IDS2017 ve UNSW-NB15 veri kümelerine uygulanan ön işleme ardışık düzenini ayrıntılı biçimde sunmaktadır. Veri ön işleme, saldırı tespit modellerinin kalitesini ve performansını doğrudan etkileyen kritik bir adımdır.



Şekil 6. Saldırı tespit veri kümeleri için veri ön işleme ardışık düzeni; model geliştirme ve değerlendirme için alt kümeler oluşturmak üzere ham veriden temizleme, öznitelik kodlama, normalleştirme, sınıf dengesizliği giderme ve son veri bölme adımlarına kadar

3.2.1. UNSW-NB15 Veri Kümesine Genel Bakış

UNSW-NB15 veri kümesi, 10 trafik kategorisi genelinde 40 sayısal ve üç kategorik öznitelik (proto, service, state) içermektedir. İkili sınıflandırma için normal (65.100) ve saldırı (115.271) örnekleri, ılımlı bir 1,77:1 dengesizlik oranı oluşturmaktadır. Çok sınıflı ayar çok daha zorlu bir tablo sunmakta olup Normal sınıfı, Worms sınıfını 533:1 oranında geride bırakmaktadır. Ek olarak, bayt sayısı öznitelikleri (sbytes, dbytes) 10^9'a kadar ulaşan değerlere sahipken, kararlı model eğitimini güvence altına almak için kantil tabanlı normalleştirme zorunludur.

Tablo 3. UNSW-NB15 ön işleme genel bakışı.

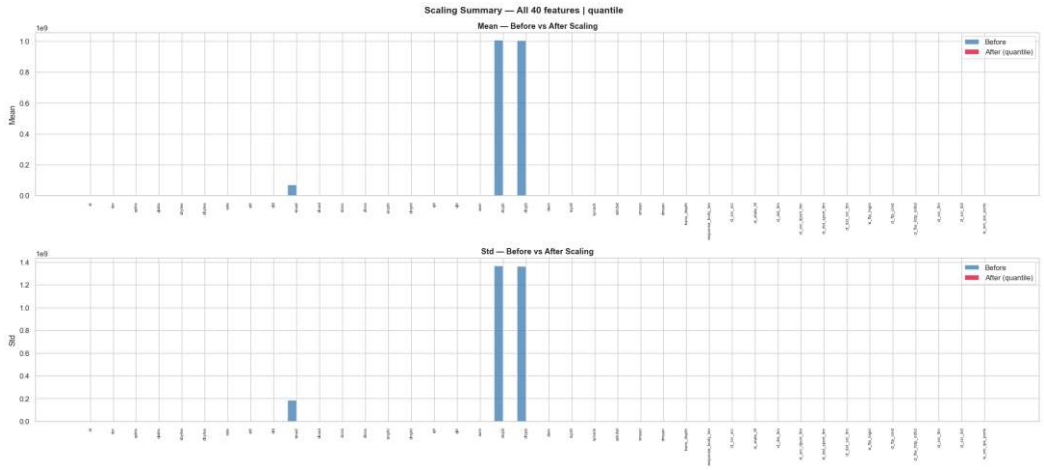
Özellik	Değer
Sayısal öznitelikler	40 (+ kategorik: proto, service, state)
İkili hedef	'label' -> 0 = Normal, 1 = Saldırı
Çok sınıflı hedef	'attack_cat' -> 10 kategori
Saldırı türleri	Normal, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms
Temel zorluk	Aşırı ölçek eşitsizliği: sbytes / dbytes 10^9'a ulaşmaktadır

Tablo 3. (Devamı).

Özellik	Değer
Sınıf dengesizliği (ikili)	NORMAL 65.100'e karşı SALDIRI 115.271 (oran 1,77:1)
Sınıf dengesizliği (çok sınıflı)	Normal 65.100'e karşı Worms 122 (oran 533:1)

3.2.2. UNSW-NB15 Üzerinde Ölçeklendirme Sonuçları

Bu şekil, kantil normalleştirme uygulandıktan sonra 40 özniteliğin tümünün nasıl değiştiğini özetlemektedir. Üst panel ortalama değerleri, alt panel ise normalleştirme öncesi ve sonrasındaki standart sapmayı göstermektedir. Normalleştirme öncesinde sbytes ve dbytes gibi bazı öznitelikler milyarlara (10^9) ulaşan son derece büyük değerlere sahiptir.



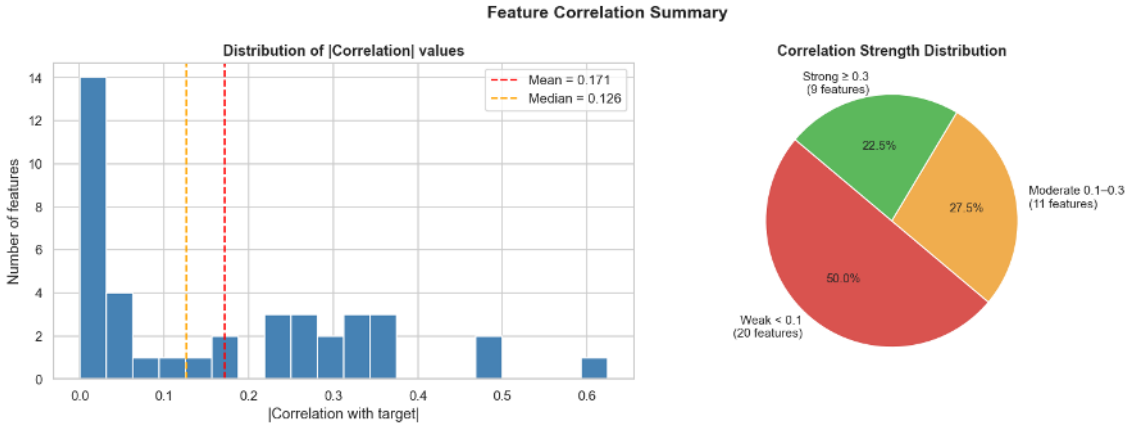
Şekil 7. Kantil normalleştirme kullanılarak tüm 40 öznitelik için ölçeklendirme özeti.

Ölçeklendirmeden sonra (kırmızıyla gösterilmiştir) tüm öznitelikler sıfır etrafında merkezlenmiş çok daha küçük ve tutarlı bir aralığa çekilmekte; bu da verinin daha kararlı ve çalışılabilir hâle gelmesini sağlamaktadır.

Bu çalışmada ölçeklendirme yöntemi olarak IDS için en çok önerilen ve ağ trafiği öznitelikleri için en uygun olan QuantileTransformer kullanılmıştır; aykırı değerleri ele almak ve güçlü öznitelik çok ölçekli çarpıklığının etkisini azaltmak amacıyla tercih edilmiştir.

3.2.3. Öznitelik Korelasyon Analizi

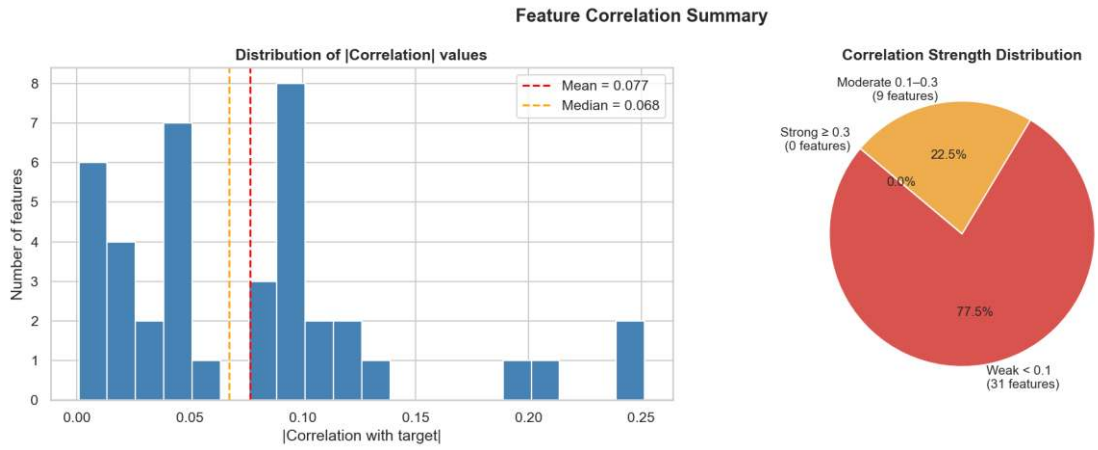
Korelasyon analizi, bilgi içeren özniteliklerin tanımlanması ve gereksiz olanların tespit edilmesi olmak üzere iki amaca hizmet etmektedir.



Şekil 8. İkili sınıflandırma için öznitelik korelasyonu özeti.

UNSW-NB15 için Karşılıklı Bilgi Tabanlı Öznitelik Seçimi: Bu özet, özniteliklerin hedef değişkenle ne derece güçlü ilişki içinde olduğuna göre nasıl seçildiğini göstermektedir. Toplamda 9 öznitelik son derece bilgi içerici (KB ≥ 0,3) olarak öne çıkarken, 11 öznitelik orta düzeyde yararlı bilgi sunmakta ve aynı şekilde tutulmaktadır. Öte yandan çok düşük düzeyde ilgililik (KB <0,1) gösteren 20 öznitelik kaldırılmak üzere değerlendirilmektedir. Ek olarak, 9 öznitelik çiftinin birbirleriyle son derece benzer olduğu (korelasyon 0,95'in üzerinde) belirlenerek her çiftten bir özniteliğin gereksizliği azaltmak amacıyla kaldırılabileceği tespit edilmiştir.

Gereksiz öznitelikler: sbytes, dbytes, sloss, dloss, dwin, ct_src_dport_ltm, ct_dst_src_ltm, ct_ftp_cmd, ct_srv_dst. Özellikle sbytes ve dbytes, hem büyüklük açısından baskın hem de birbirleriyle fazlalıklıdır; bu nedenle dikkatli biçimde ele alınmaları için iki ayrı gerekçe mevcuttur.

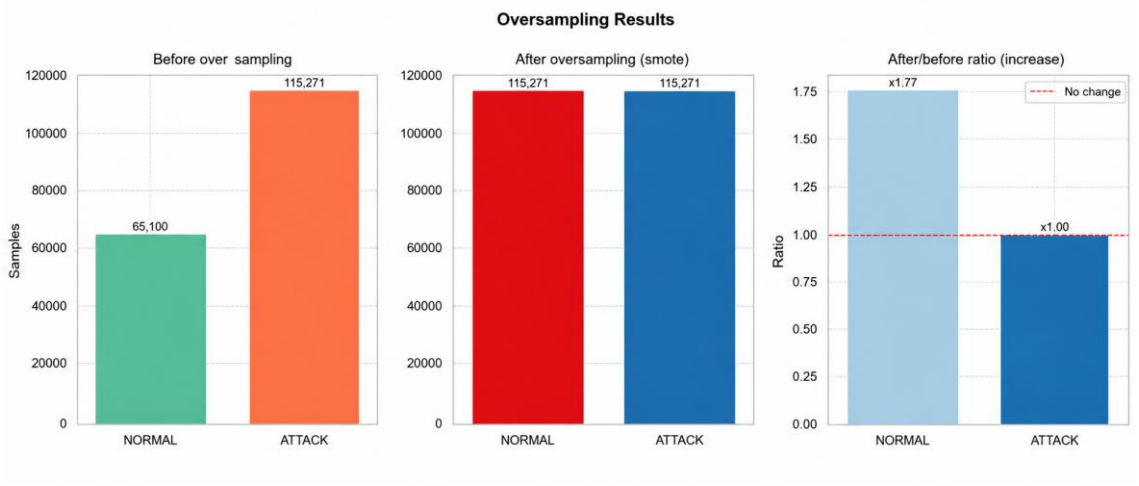


Şekil 9. Çok sınıflı sınıflandırma için öznitelik korelasyonu özeti.

3.2.4. UNSW-NB15 Sınıf Dengesizliği ve Aşırı Örnekleme

3.2.4.1. İkili Görev

İkili sınıflandırma görevinde UNSW-NB15 veri kümesi, normal trafik örneklerinin saldırı örneklerine kıyasla belirgin biçimde fazla olduğu ciddi bir sınıf dengesizliği sergilemektedir. Bu dengesizlik, modelin azınlık sınıfını yani saldırıları öğrenmesini güçleştireceğinden, eğitim aşamasında uygun bir aşırı örnekleme stratejisinin benimsenmesi zorunlu hâle gelmektedir.

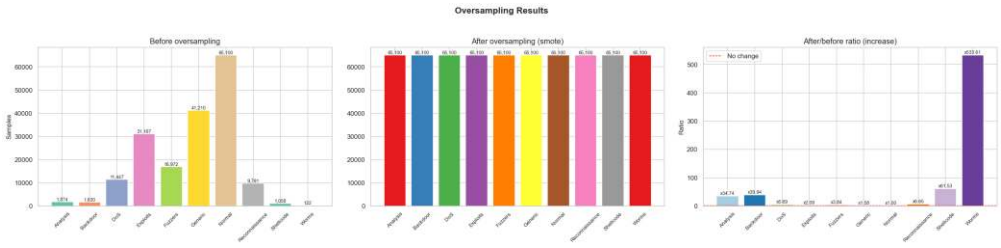


Şekil 10. SMOTE öncesi ve sonrası aşırı örnekleme ikili.

3.2.4.2. Çok Sınıflı Görev Aşırı Dengesizliği

Çok sınıflı görevde sınıf dengesizliği çok daha belirgin bir hal almaktadır. Şekil 11'de görüldüğü üzere, SMOTE uygulanmadan önce "Normal" ve "Generic" gibi baskın sınıflar onlarca bin örneğe ulaşırken "Backdoor" ve "Shellcode" gibi nadir saldırı türleri

yalnızca birkaç yüz örnekle temsil edilmekte olup bu dengesizlik bazı sınıflarda 400 katı aşan artış oranlarıyla giderilmiştir.



Şekil 11. SMOTE öncesi ve sonrası aşırı örnekleme çok sınıflı.

3.2.5. CIC-IDS2017 Veri Kümesine Genel Bakış

CIC-IDS2017 veri kümesi, toplam 78 sayısal öznitelik ve Etiket sütunundan oluşmaktadır. İkili sınıflandırma için Etiket hedefi, iyi huylu ve saldırı türleri arasında ayrım yapmak amacıyla kullanılmaktadır. Ancak çok sınıflı bir sınıflandırma çalışmasında Etiket, iyi huylu artı DoS, DDoS, PortScan, BruteForce, Bot, Infiltration, Web Saldırıları, Heartbleed gibi 14 alt kategoriyi kapsamaktadır. İyi huylu ve saldırı sınıfları arasındaki parametre farkı yaklaşık 4,1:1 oranına karşılık gelecek kadar yüksektir. Buna karşın çok sınıflı sınıflandırmada oran, yaklaşık 1,6 milyon iyi huylu saldırıya karşı 10'dan az örnek içeren en küçük saldırı sınıfıyla (oran> 200.000:1) son derece zorlayıcı bir tablo ortaya çıkarmaktadır.

Tablo 4. CIC-IDS2017 ön işleme genel bakışı.

Özellik	Değer
Özellik türleri	Yalnızca sayısal ; kategorik kodlama gerekmemektedir
İkili hedef	Etiket → BENIGN veya SALDIRI
Çok sınıflı hedef	Etiket → BENIGN + 14 saldırı alt kategorisi
Saldırı türleri	DoS, DDoS, PortScan, BruteForce, Bot, Infiltration, Web Attacks, Heartbleed.
İkili sınıf oranı	BENIGN 1.591.188 — SALDIRI 390.352 (oran ≈ 4,1:1)
Çok sınıflı zorluk	NORMAL ~1,6M — en küçük saldırı sınıfı < 10 örnek (oran > 200.000:1)

3.2.6. Ardışık Düzen Sırası ve Tasarım Tercihleri.

CIC-IDS2017 veri kümesi için işlem zinciri 5 farklı adım üzerine kurulmuştur. İlk olarak eksik değerler ele alınır; eğitim katındaysak ortalama atama gerçekleştirilebilir, ancak bunun doğrulama veya test katına bilgi sızdırmamak için eğitim katının kendisi üzerinde hesaplanması gerekmektedir. Eksik değerler ele alındıktan sonra veri, eğitim için tabakalı %70 ve doğrulama/test için %10/%20 olacak şekilde bölünür. Ölçeklendirme, QuantileTransformer ile gerçekleştirilir. Bu bileşen, ağ trafiği öznitelikleri için tipik olan güçlü öznitelik çok ölçekli çarpıklığının etkisini büyük ölçüde azaltırken büyük aykırı değerlerin etkisini de önemli ölçüde azaltmaktadır. Ardından ikili sınıfı dengelemek amacıyla SMOTE aşırı örnekleme tekniği uygulanır (saldırı sınıfında ×4,08), ancak yalnızca eğitim kümesi üzerinde. Son olarak kategorik öznitelikler kodlanır; düşük kardinaliteli öznitelikler için TekHot kodlama, yüksek kardinaliteli olanlar için ise Hedef Kodlama kullanılır. Burada kritik olan nokta, Hedef Kodlamanın yalnızca eğitim katında uyarlanmasıdır.

3.2.7. Öznitelik Korelasyon Analizi.

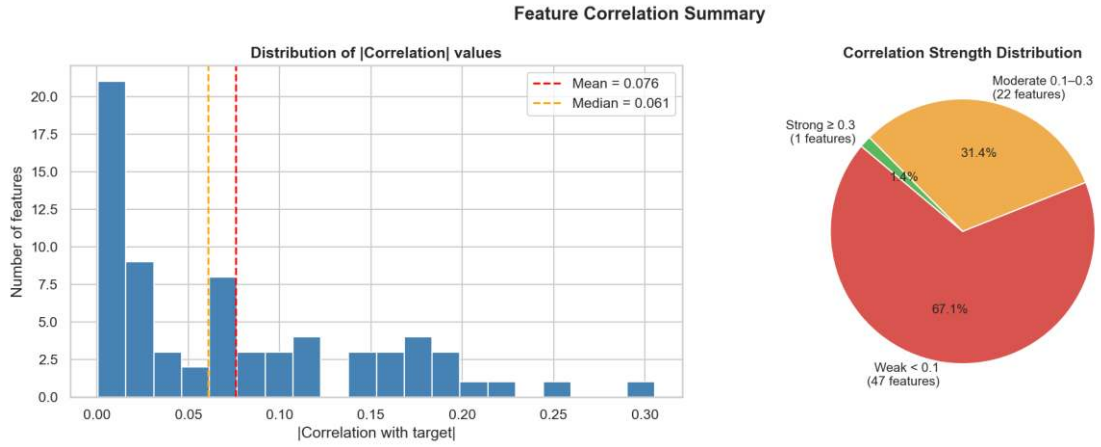
Her öznitelik ile hedef değişken arasındaki korelasyon analizi, Pearson mutlak korelasyon katsayısı kullanılarak gerçekleştirilmiştir. Çok sınıflı ve ikili görevler için ayrı ayrı yürütülen iki analiz, belirgin biçimde farklı dağılımlar ortaya koymuştur.

3.2.7.1. Çok Sınıflı Görev Korelasyonu.

Pearson korelasyonu analizinden, CIC-IDS2017 veri kümesinin 70 özniteliğinin büyük çoğunluğunun çok sınıflı etiketle tek tek ele alındığında dahi düşük korelasyon sergilediği görülmektedir. Tüm 70 öznitelik için ortalama korelasyon 0,076, medyan korelasyon ise 0,061'dir. Tablo 5, tüm öznitelikler için elde edilen korelasyon değerlerinin dağılımına ilişkin genel bir bakış sunmaktadır. Tablo 5'den yalnızca 1 özniteliğin (0,357) 0,3'ün üzerinde korelasyon değeri taşıdığı görülebilmektedir. 22 öznitelik (0,314), 0,1 ile 0,3 arasında yer alan orta düzey korelasyon değerine sahiptir. Öte yandan 70 öznitelikten yalnızca 5'i, yani toplam özniteliklerin %7'si 0,7'nin üzerinde değer almaktadır. Bu durum, söz konusu özniteliklerin problemimiz açısından gürültü niteliği taşıdığını göstermektedir.

Tablo 5. Çok sınıflı görev korelasyonu

Kategori	Eşik	Sayı	Oran	Yorum
Güçlü	≥ 0,3	1	%1,4	Yalnızca 1 özellik çok sınıflı etiketle güçlü doğrusal ilişki sergilemektedir.
Orta	0,1–0,3	22	%31,4	Baskın olmamakla birliktebelirlidüzeyde sinyal taşımaktadır.
Zayıf	< 0,1	47	%67,1	Görünürde gürültü; ancak Pearson sınırlılığına dikkat ediniz.



Şekil 12. Çok sınıflı sınıflandırma için öznitelik korelasyonu özeti.

Sayısal olarak bu öznitelik, yalnızca 0,04'lük bir puan ve negatif Pearson öznitelik önemiyle oldukça zayıf görünmektedir. Bununla birlikte Pearson yalnızca verideki doğrusal ilişkileri ölçmekte olup, bu öznitelik kötü puan almasına karşın özellikle söz konusu güç doğrusal olmayan ilişkilerden kaynaklanıyorsa sıradan olmayan ayırt edici güce sahip olabilir. Sonuç olarak bu öznitelik yalnızca Pearson değerine dayanılarak dışarıda bırakılmamalıdır; karşılıklı bilgi veya bir ağaç tabanlı öznitelik önem skoru hesaplandığında farklı bir önem değeri alabilir.

3.2.7.2. İkili Görev Korelasyonu

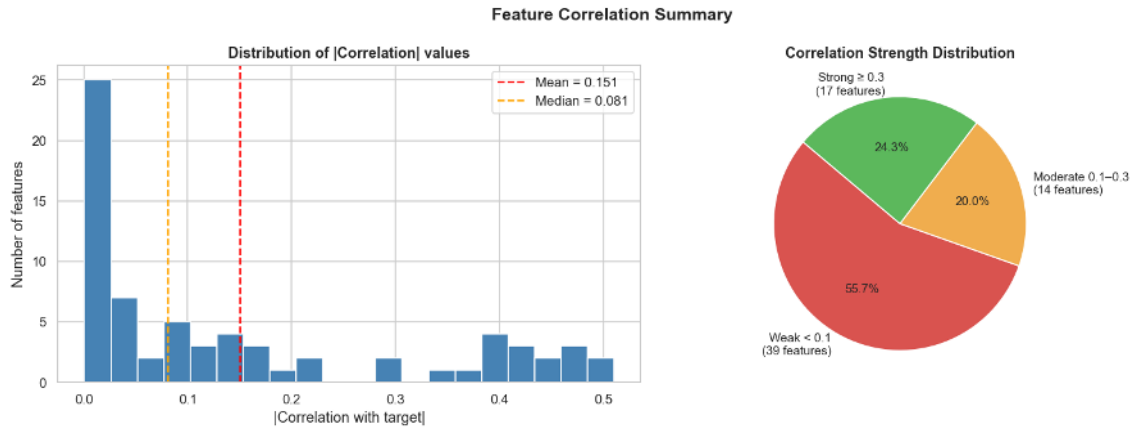
İkili duruma bakıldığında, kurulumun çok sınıflı duruma göre çok daha net ve anlaşılır olduğu görülmektedir. Ortalama korelasyon 0,151, medyan ise 0,081'dir. Bu, öznitelikler ile NORMAL ve SALDIRI etiketi arasındaki bağlantının daha açık biçimde gözlemlendiğini göstermektedir.

Tablo 6 incelendiğinde belirgin bir farklılık dikkat çekmektedir. Bu kez 17 öznitelik 0,3 eşiğinin üzerindedir; bu, neredeyse hiçbir özniteliğin öne çıkmadığı çok sınıflı durumla kıyaslandığında büyük bir değişime işaret etmektedir.

Tablo 6. İkili görev korelasyonu açıklaması

Kategori	Eşik	Sayı	Oran	Yorum
Güçlü	≥ 0,3	17	%24,3	BENIGN-SALDIRI ayrımında güçlü öngörü gücüne sahip 17 özellik
Orta	0,1–0,3	14	%20,0	Tamamlayıcı nitelikte kullanışlı özellikler
Zayıf	< 0,1	39	%55,7	Düşük doğrusal korelasyon; bilgi içermediği anlamına gelmez

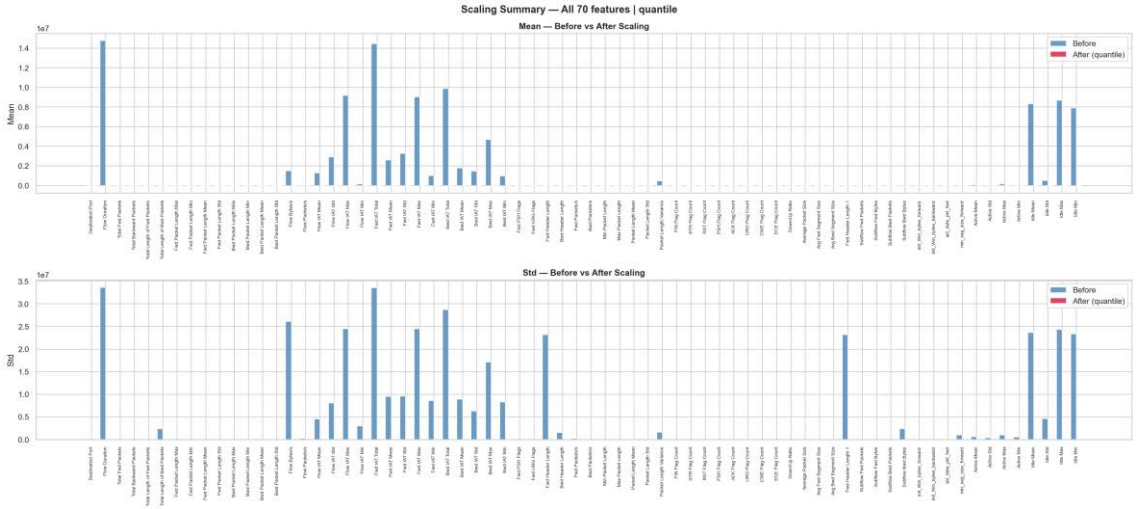
Bununla birlikte her şey daha iyi olmamaktadır. Özniteliklerin %55,7'si hâlâ 0,1'in altında son derece zayıf korelasyona sahiptir. Zayıf olmaları, yararlı olmadıkları anlamına gelmez. İkili görev yalnızca daha kolaydır; çünkü 14 saldırı türünü tek bir grupta topladığınızda örüntüler daha belirgin hâle gelir ve ilişkiler daha kolay bulunur. NORMAL ile SALDIRI etiketi daha basittir. Bu durum, öznitelikler ile NORMAL ve SALDIRI etiketi arasındaki bağlantıları daha kolay görünür kılmaktadır.



Şekil 13. CIC-IDS2017 öznitelik korelasyonu özeti.

3.2.8. CIC-IDS2017 Üzerinde Veri Ölçeklendirme Sonuçları

CIC-IDS2017, son derece heterojen ölçeklere sahip 70 öznitelik içermektedir. Birçok öznitelik, ortalamaları ve standart sapmaları 10^7 mertebesindeyken diğerleri 1'in altında kalmaktadır. Ölçeklendirme yapılmadan ÇKA, yüksek büyüklüklü öznitelikler tarafından domine edilecek; gradyan dengesizliğine ve zayıf yakınsamaya yol açacaktır.



Şekil 14. Çok sınıflı ölçeklendirme sonuçları.

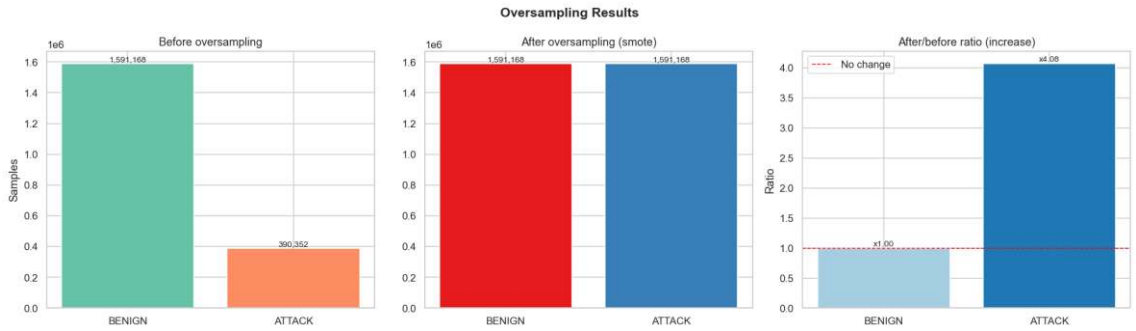
3.2.8.1. Gözlemlenen Sonuçlar

Ölçeklendirme Özet grafiği, dönüşüm öncesinde birden fazla özniteliğin yaklaşık 10^7 değerine ulaştığını doğrulamaktadır (ortalama ve standart sapma grafikleri). QuantileTransformer uygulandıktan sonra 70 özniteliğin tümü tekdüze bir [0, 1] dağılımına sıkıştırılmaktadır. 'Sonra (kantil)' çubukları 10^7 ekseninde görünmez hâle gelmekte; bu da dönüşümün doğru çalıştığını teyit etmektedir. Ağ trafiği veri kümelerinde genellikle log-normal dağılımlar sergiledikleri tespit edilen temel yüksek büyüklüklü öznitelikler arasında paket uzunluğu istatistikleri, saniye başına akış baytları ve süre tabanlı öznitelikler yer almaktadır.

3.2.9. CIC-IDS2017 Sınıf Dengesizliği ve Aşırı Örnekleme.

3.2.9.1. İkili Görev

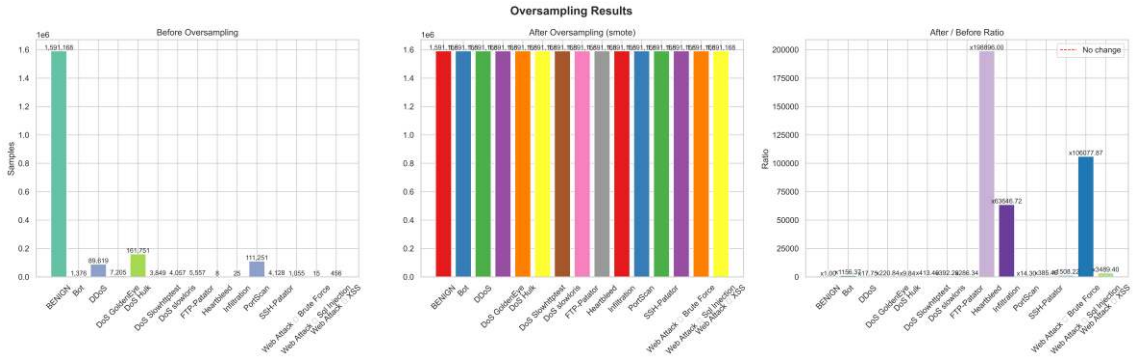
İkili veri kümesi, BENIGN'ın SALDIRI'ya göre baskın olduğu orta düzeyde bir dengesizlik sergilemektedir. SMOTE, her iki sınıf eşitleninceye kadar öznitelik uzayında k en yakın komşular arasında enterpolasyon yaparak sentetik SALDIRI örnekleri üretmektedir.



Şekil 15. SMOTE öncesi ve sonrası aşırı örnekleme.

3.2.9.2. Çok Sınıflı Görev Aşırı Dengesizliği

Çok sınıflı senaryo, gerçek dünya ağ trafiğinin karakteristik özelliği olan ciddi bir dengesizliği ortaya koymaktadır. BENIGN büyük çoğunluğu (~1,6 milyon örnek) oluştururken, çeşitli saldırı kategorileri yalnızca birkaç örnek içermektedir. Bu durum, IDS literatüründeki en zorlu dengesizlik senaryolarından birini oluşturmaktadır.



Şekil 16. SMOTE öncesi ve sonrası aşırı örnekleme.

3.3. Modeller

3.3.1. Genel Bakış ve Motivasyon

Mimari seçimi aşamasında üç amaç eş zamanlı olarak gözetilmiştir. Birincisi, ham öznitelik vektörleri üzerindeki yerel uzamsal örüntülerin yararlı bilgi olarak ortaya çıkmasına olanak tanımak; ikincisi, ağ akışlarındaki ardışık ve çift yönlü zamansal ilişkileri yakalamak; üçüncüsü ise seçilen mimariler için bir karşılaştırma temeli oluşturmak amacıyla bilinen klasik taban çizgileri bulmaktır. Toplamda 4 farklı mimari stilini keşfeden 10 farklı model geliştirilmiştir.

- Uzamsal öznitelik çıkarma için Tek Boyutlu Evrişimli Sinir Ağları (CNN).
- Yerel evrişim modelleme ve çift yönlü ardışık modellemenin hibrit güçlerini birleştiren CNN-BiLSTM Ağları.
- Topluluk ağaç tabanlı taban çizgisi olarak Rastgele Orman (RF).

- XGBoost.

Her mimariyi adil biçimde değerlendirmek amacıyla, ölçek sınırlılıkları nedeniyle yalnızca UNSW-NB15 veri kümesine uygulanan klasik MÖ modelleri dışında, hem Normal ve Saldırı (ikili) hem de Çoklu Sınıf (çok sınıflı) ayarları değerlendirme metriği olarak kullanılmaktadır. Tüm modeller daha sonra Bölüm 4.5'te açıklanan birleşik ve adil bir deneysel protokol kapsamında değerlendirilmektedir.

3.3.2. CNN Mimarisi

3.3.2.1. Mimari Tanımı

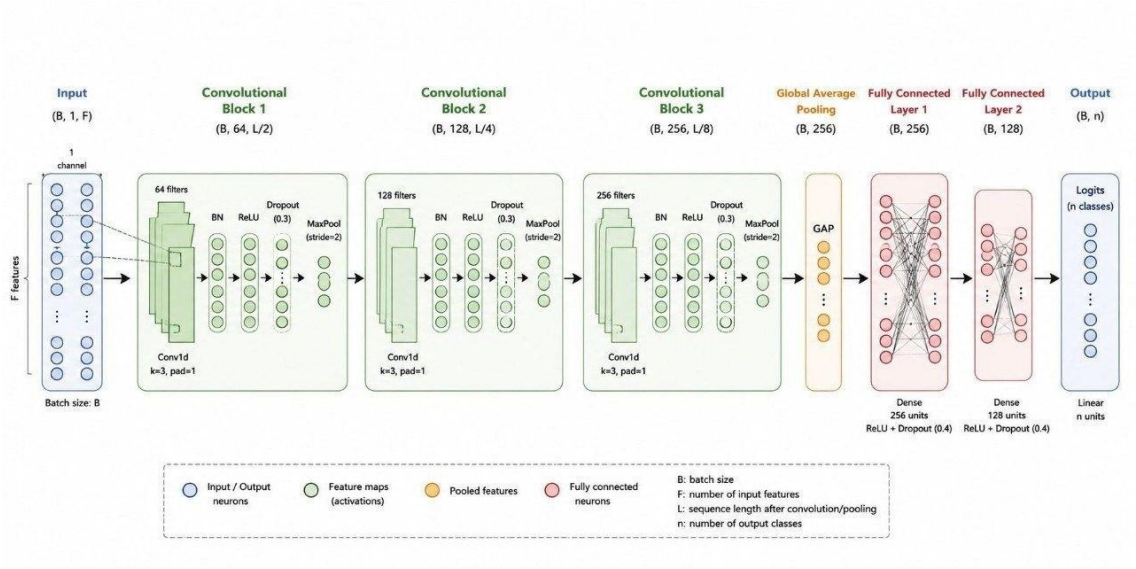
CNN mimarisi, ağ trafiği özniteliklerini ardışık veri olarak işlemek amacıyla 1D evrişimler kullanmaktadır. Model, sınıflandırma için tam bağlantılı katmanların ardından gelen üç evirişimse bloktan oluşmaktadır.

3.3.2.2. 1D Evrişim Denklemi

Bir giriş öznitelik vektörü x ve filtre w için 1D evrişim şu şekilde hesaplanmaktadır:

$$y_i = \sum_{j=0}^{k-1} x_{i+j}.w_j + b \tag{15}$$

Burada x giriş sinyalini (ağ akışı öznitelik vektörü), $\boldsymbol{w_j}$ filtre ağırlıklarını (eğitim sırasında öğrenilen), $\boldsymbol{j}$ çekirdek boyutu üzerindeki indisi, $\boldsymbol{k}$ çekirdek boyutunu (bizim durumumuzda $\boldsymbol{k = 3}$), $\boldsymbol{i}$ giriş dizisindeki geçerli konumu, $\boldsymbol{y_i}$ $\boldsymbol{i}$ konumundaki çıkış değerini ve $\boldsymbol{b}$ önyargı terimini ifade etmektedir. Temel CNN, NIDS için Şekil'de gösterildiği gibi üst üste istiflenmiş bileşenlerden oluşturulmuştur.



Şekil 17. Önerilen evrişimli sinir ağı.

3.3.3. CNN-BiLSTM Hibrit Mimarisi

Trafik sınıflandırması için 1D CNN kullanımına ek olarak, ağ trafiğindeki öznitelikleri özellikle kullanan ileri bir derin öğrenme yöntemi de sunulmaktadır: evrişimsel katmanları öznitelik çıkarma ile birleştiren ve birden fazla paket üzerindeki öznitelik boyutlarının dizi bilgisini modelleyen Çift Yönlü Uzun-Kısa Süreli Bellek (BiLSTM) ağlarını içeren hibrit bir CNN-BiLSTM mimarisi. CNN'ler belirli öznitelikler için zaman içindeki yerel öznitelik örüntülerini yakalayabilirken, uzun menzilli bağımlılıkları yakalayamamaktadır. Tekrarlayan birimler ise özellikle bu tür uzun menzilli bağımlılıkları işlemek için tasarlanmıştır.

Önerilen CNN-BiLSTM omurga ağı, artık yapılara sahip iki 1D evrişimsel bloktan ve ardından gelen bir Çift Yönlü LSTM (BiLSTM) tekrarlayan yapısından oluşmaktadır. Her evrişimsel blok, artan kanal sayılarına ($1 \rightarrow 64 \rightarrow 128$) sahip iki 1D evrişimsel katman ve ardından bir TopluNormalleştirme katmanı, ReLU aktivasyonu, MaksimumHavuzlama katmanı ve son olarak 0,3'lük bir HattanDüşürme katmanı içermektedir. Son evrişimsel bloğun öznitelik haritalarından çıkarılan öznitelikler, her yönde 128 gizli nörona sahip BiLSTM tekrarlayan yapısına girdi olarak verilmekte; böylece her zaman adımının gizli durumları için 256 boyutlu bir çıktı elde edilmektedir. Gerçekte trafik akışı öznitelikleri, geçmiş ve gelecek bilgilerin birlikte kullanılmasının ilgili öznitelikleri yakalayabileceği asimetrik bir yönsel bilgiye sahiptir.

BiLSTM, hücre durumu güncellemeleri sırasında tanh aktivasyonu ve kapılama işlemlerinde sigmoid, standart bir LSTM mimarisini takip ederek uygulanmaktadır. LSTM çıktısına 0,3'lük bir HattanDüşürme de uygulanmaktadır. BiLSTM'nin son gizli

durumu, daha sonra tam bağlantılı bir katmana (256 birim-> 128 birim) girdi olarak kullanılmaktadır. Bu katman, ReLU aktivasyonu ve 0,3'lük HattanDüşürme ile düzenlileştirilmiştir. Bunun çıktısı, gereken sınıf sayısı üzerinde olasılık dağılımı çıkarmak amacıyla Softmax aktivasyonu uygulanmış nihai çıkış katmanına iletilmektedir.

Tek CNN modeli ile karşılaştırıldığında CNN-BiLSTM modeli, özellikle öznitelik sırasının anlamsal anlam taşıdığı durumlarda paket düzeyi istatistiklerden toplanan mühendislik trafiği öznitelikleri için geçerli olmak üzere öznitelikler arası ardışık yapıları değerlendirmek amacıyla ilişkisel modellemeyi kullanmaktadır. Bu, son dönem NIDS literatüründe yaygın bir uygulamadır.

3.3.4. Rastgele Orman

Rastgele Orman (RO), girdinin sınıflandırmasının bireysel ağaçlar tarafından gerçekleştirilmesi durumunda bir çıktı dağılımıyla sonuçlandığı ve son sınıfın en sık görülen sonuç olduğu bir karar ağacı koleksiyonundan oluşan bir topluluk öğrenme yöntemidir. RF'nin varyans azaltma mekanizması, önyükleme örneklemesine ve öznitelik rastsallığına dayanmakta; yüksek boyutsallığın üstesinden gelmede özellikle etkili olmaktadır.

Bu test için RO, UNSW-NB15 üzerinde ikili sınıflandırma görevine uygulanmıştır. İlgisiz özniteliklerin gürültüsünden kaçınmaya yardımcı olmak amacıyla Karşılıklı Bilgi (KB) öznitelik seçimi yöntemi en iyi 30 öznitelik kullanılarak uygulanmıştır. Ek olarak, eğitim öncesinde sınıf dengesizliğini azaltmak için SMOTE kullanılmıştır. Eğitim süresi 5.069 saniye gibi oldukça yüksek olmakla birlikte, doğruluk ve F1 Skoru sırasıyla 0,9854 ve 0,9885 ile yüksek çıkmış; bu da RF'nin güçlü bir taban çizgisi olduğunu ortaya koymaktadır.

3.3.5. XGBoost

Aşırı Gradyan Artırma, kısaca XGBoost, hem hesaplama kaynağı kullanımı hem de aşırı öğrenmeye karşı direnç açısından son derece verimli olacak şekilde tasarlanmış ölçeklenebilir ve dağıtık bir Gradyan Artırma çerçevesidir. Algoritma, aşamalı bir biçimde çok sayıda karar ağacından oluşan bir topluluk oluşturur ve her düğümde öznitelik değerlerine göre veriyi böler. XGBoost'ta kullanılan düzenlileştirilmiş amaç fonksiyonu, hem L1 hem de L2 düzenlileştirmeyi kullanmakta ve ikinci dereceden gradyan istatistiklerinden yararlanmaktadır.

XGBoost'un performansı, UNSW-NB15 üzerinde hem ikili hem de çok sınıflı sınıflandırma görevlerinde değerlendirilmektedir. Çok sınıflı görev (10 sınıf) için, sınıf dengesizliği sorununu daha da ele almak amacıyla SMOTE ile birlikte özel bir amaç fonksiyonu olarak özel bir Odak Kaybı versiyonu uygulanmaktadır. Öznitelik kodlaması, olası hedef bilgisi sızıntısını önlemek amacıyla yalnızca eğitim verisi üzerinde uyarlanmış TargetEncoder kullanılarak gerçekleştirilmektedir. En bilgi içerici öznitelikler, Karşılıklı Bilgi (KB) ile en iyi 30 öznitelik seçimi kullanılarak belirlenmekte; ancak bu kez tüm eğitim katı üzerinde yapılmaktadır. Modeller, GPU üzerinde saniyeler içinde eğitilmiştir; özellikle ikili görev için 203 saniyede ve en iyi çok sınıflı model için 30,9 saniyede tamamlanmıştır.

3.4. Eğitim Yapılandırması

Adil karşılaştırmalar sağlamak amacıyla tüm derin öğrenme deneyleri genelinde birleşik ve tekrarlanabilir bir eğitim protokolü benimsenmiştir. Temel hiper parametreler, düzenlileştirme stratejileri ve optimizasyon ayarları aşağıda özetlenmiştir.

3.4.1. Hiper Parametreler

Tablo 7, tüm sekiz derin öğrenme modeli yapılandırması (2 mimari × 2 görev × 2 veri kümesi) için eksiksiz hiper parametre yapılandırmasını sunmaktadır. Tüm modeller, 1e-3 öğrenme oranıyla Adam optimize edici kullanılarak eğitilmektedir. Toplu iş boyutları mimari ve veri kümesine göre farklılık göstermektedir: CNN modelleri CIC-IDS2017 üzerinde 512, UNSW-NB15 üzerinde 256 ile eğitilirken; CNN-BiLSTM modelleri için bu değerler sırasıyla 64 ve 128 (CIC-IDS2017) ile 256 (UNSW-NB15) olarak belirlenmiştir. Maksimum döngü sayısı, CNN modellerinde 100 ve bazı yapılandırmalarda 50 ile sınırlandırılmıştır; CNN-BiLSTM modelleri ise en fazla 30 döngü ile eğitilmiştir. Aşırı öğrenmeyi önlemek amacıyla doğrulama kaybı üzerinde erken durdurma uygulanmakta olup sabır değeri CNN ikili sınıflandırma yapılandırmalarında 7, diğer tüm yapılandırmalarda ise 5 olarak belirlenmiştir.

Çok sınıflı yapılandırmalarda, Focal Loss, odaklama parametresi γ=2.0 ve varsayılan sınıf ağırlığı α=1 ile uygulanır; bu, tüm sınıfların kayıp ölçeklendirmesine eşit şekilde katkıda bulunduğu anlamına gelir. Sınıf dengesizliği bunun yerine ön uçta SMOTE aşırı örnekleme yöntemiyle ele alınır, bu da bu ortamda açık bir α ağırlığını gereksizkılar.

Tablo 7. Tüm derin öğrenme modelleri için hiperparametre yapılandırması.

Veri kümeler	Model mimarisi	Görev	Aktivasiyon fonksiyonu	Dropout	Optimizatör &Eğitim Oranı	Kayıp	Epok/ Yığın Boyutu	Erken durdurma
CIC- IDS2017	CNN (3 Conv1d blocks1→64→128→256+ GAP + 2 FC256→128→2)	İkili	ReLU (conv+FC) Softmax (output)	0.3 conv 0.4 FC	Adam lr=1e-3	Focal Loss (γ=2, α=0.25)	100 / 512	Sabır=7
	CNN-BiLSTM (2 Conv1d 1→64→128+ BiLSTM 128→256+ FC 256→128→2)	İkili	ReLU (conv+FC) tanh/sigmoid (LSTM)Softmax (output)	0.3 conv 0.3 LSTM 0.3 FC	Adam lr=1e-3	Focal Loss (γ=2, α=0.25)	30 / 64	Sabır=5
	CNN (3 Conv1d blocks1→64→128→256+ GAP + 2 FC256→128→15)	Çok sınıflı (15 sınıf	ReLU (conv+FC) Softmax (output)	0.3 conv 0.4 FC	Adam lr=1e-3	Focal Loss (γ=2.0)	100 / 512	Sabır=5
	CNN-BiLSTM (2 Conv1d 1→64→128+ BiLSTM 128→256+ FC 256→128→15)	Çok sınıflı (15 sınıf	ReLU (conv+FC) Tanh/sigmoid (LSTM) Softmax (output)	0.3 conv 0.3 LSTM 0.3 FC	Adam lr=1e-3	Focal Loss (γ=2.0)	30 / 128	Sabır=5

Tablo 7. (Devamı)

UNSW- NB15	CNN (3 Conv1d blocks1→64→128→256+ GAP + 2 FC256→128→2)	İkili	ReLU (conv+FC) Softmax (output)	0.3 conv 0.4 FC	Adam lr=1e-3	Focal Loss (γ=2, α=0.25)	100 / 256	Sabır=7
	CNN-BiLSTM (2 Conv1d 1→64→128+ BiLSTM 128→256+ FC 256→128→2)	İkili	ReLU (conv+FC) Tanh/sigmoid (LSTM) Softmax (output)	0.3 conv 0.3 LSTM 0.3 FC	Adam lr=1e-3	Focal Loss (γ=2, α=0.25)	100 / 256	Sabır=7
	CNN (3 Conv1d blocks1→64→128→256+ GAP + 2 FC256→128→10)	Çok sınıflı (10 sınıf	ReLU (conv+FC) Softmax (output)	0.3 conv 0.4 FC	Adam lr=1e-3	Focal Loss (γ=2.0)	50 / 256	Sabır=5
	CNN-BiLSTM (2 Conv1d 1→64→128+ BiLSTM 128→256+ FC 256→128→10)	Çok sınıflı (10 sınıf	ReLU (conv+FC) Tanh/sigmoid (LSTM) Softmax (output)	0.3 conv 0.3 LSTM 0.3 FC	Adam lr=1e-3	Focal Loss (γ=2.0)	30 / 128	Sabır=5

3.4.2. İkili ve Çok Sınıflı Kayıp Fonksiyonu

Standart çapraz entropi kaybı, tüm örneklere sınıf sıklığından bağımsız olarak eşit önem atadığından, yoğun şekilde dengesiz NIDS veri kümeleri için yetersiz kalmaktadır. Bu nedenle, (Lin vd., 2017) tarafından orijinal olarak yoğun nesne tespiti için önerilen Odak Kaybı, tüm derin öğrenme modelleri ve XGBoost çok sınıflı deneyi genelinde benimsenmektedir.

Odak Kaybı, standart ikili/çok sınıflı çapraz entropiyi, kolay sınıflandırılmış örneklerin katkısını azaltan ve öğrenmeyi zor, yanlış sınıflandırılmış örneklere yönelten $(1-p_t)^{\gamma}$modülasyon faktörünü tanıtarak değiştirmektedir. Formülasyon şöyledir:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{16}$$

Burada:

$$p_t = \begin{cases} p & if\ y = 1 \\ 1-p & if\ y = 0 \end{cases} \tag{17}$$

$$\alpha_t = \begin{cases} \alpha & if\ y = 1 \\ 1-\alpha & if\ y = 0 \end{cases} \tag{18}$$

3.4.3. Katlı Tabakalı Çapraz Doğrulama

Tüm derin öğrenme deneyleri, sağlam ve tarafsız performans tahminleri sağlamak amacıyla tabakalı 5-katlı çapraz doğrulama kullanılarak gerçekleştirilmektedir. Tabakalandırma, sınıf dağılımının tüm katlar genelinde korunmasını sağlayarak herhangi bir katın azınlık sınıflarının orantısız bir temsiline sahip olmasını önlemektedir.

Her kat için veri, eğitim (%70) ve doğrulama (%10) alt kümelerine bölünmektedir. Öznitelik ölçeklendirme (Kantil Dönüştürücü) ve SMOTE aşırı örnekleme dahil olmak üzere ön işleme ardışık düzeni, yalnızca her katın eğitim bölümüne uyarlanmakta ve ardından uyarlanan parametreler kullanılarak doğrulama alt kümesine uygulanmaktadır. Bu, kesinlikle bilgi sızıntısı olmayan bir değerlendirme sağlamaktadır. Nihai performans metrikleri, beş kat genelinde ortalama ve standart sapma olarak raporlanmaktadır.

Klasik MÖ modelleri (UNSW-NB15 üzerinde RF ve XGBoost) için, büyük veri kümelerinde ağaç topluluk yöntemlerinin daha yüksek hesaplama maliyeti nedeniyle

çapraz doğrulama yerine sabit bir tabakalı 70/10/20 bölmesi (eğitim/doğrulama/test) kullanılmaktadır.

3.4.4. Özet: UNSW-NB15 Üzerinde Klasik MÖ Yapılandırması

Referans amacıyla Tablo 8, Bölüm 4'te sunulan karşılaştırma için bağlam sağlamak üzere UNSW-NB15 üzerindeki klasik MÖ taban çizgileri (Rastgele Orman ve XGBoost) için elde edilen yapılandırmayı ve en iyi sonuçları özetlemektedir.

Tablo 8. UNSW-NB15 üzerinde klasik MÖ yapılandırması ve sonuçlarının özeti.

Boyut	İkili (RF ve XGB)	Çok sınıflı (XGB)
Görev	Normal ve Saldırı (İkili)	10 saldırı kategorisi (Çok sınıflı)
En İyi Doğruluk	0,9854	0,8416
En İyi F1 Skoru	0,9885 (ağırlıklı / ikili)	0,8496 (ağırlıklı)
F1-Makro	N/A	0,6140
Eğitim Süresi	RF: 5.069 s | XGB: 203 s	XGB: 30,9 s (en iyi model)
Sınıf Dengeleme	SMOTE (2 sınıf)	SMOTE k=3 (10 sınıf)
Kayıp Fonksiyonu	İkili lojistik	Odak Kaybı (γ=2,0, α=0,25)
Kodlama	EtiketKodlayıcı	HedefKodlayıcı (bilgi sızıntısı olmayan)
Öznitelik Seçimi	KB en iyi 30 (tam eğitim kümesi)	KB en iyi 30 (yalnızca eğitim)

4. DENEYSEL SONUÇLAR VE ANALİZ

4.1. Deneysel Kurulum

Tüm derin öğrenme modelleri (CNN ve CNN-BiLSTM), Python 3.10 ortamında PyTorch çerçevesi kullanılarak geliştirilmiştir. Veri ön işleme ve klasik makine öğrenmesi modelleri (Rastgele Orman ve XGBoost) scikit-learn ve XGBoost kütüphaneleri aracılığıyla uygulanmıştır. Sınıf dengesizliğinin giderilmesi için imbalanced-learn kütüphanesinin SMOTE uygulaması kullanılmıştır. Tüm deneyler NVIDIA L4 GPU (23.7 GB VRAM) ve CUDA 13.0 sürücü sürümü 580.82.07 üzerinde gerçekleştirilmiştir. Tekrarlanabilirlik açısından kullanılan temel kütüphane sürümleri şu şekildedir: PyTorch 2.11.0, scikit-learn 1.3.0, XGBoost 2.0.0 ve imbalanced-learn 0.11.0.

Erken durdurma mekanizması, doğrulama kaybı izlenerek özel bir eğitim döngüsü aracılığıyla manuel olarak uygulanmıştır: belirlenen sabır değeri (sabır=5) boyunca doğrulama kaybında iyileşme gözlemlenmezse eğitim sonlandırılmış ve en iyi model ağırlıkları geri yüklenmiştir.

Odak Kaybı fonksiyonu (γ=2, α=0,25), standart çapraz entropi kaybını temel alarak PyTorch ile özel bir uygulama şeklinde kodlanmıştır. Bu değerler, Lin vd. (2017) tarafından nesne tespiti alanında önerilmiş ve literatürde yaygın biçimde benimsenen varsayılan değerlerdir.

.

4.2. Model Eğitimi ve Değerlendirmesi

Her model, 5-katlı tabakalı çapraz doğrulama kullanılarak değerlendirilmektedir. Aşırı öğrenmeyi önlerken yeterli yakınsama sağlamak amacıyla, konfigürasyona göre hafifçe değişen bir sabır değeriyle doğrulama kaybına dayalı erken durdurma uygulanmaktadır. Doğrulama setindeki F1-skoruna göre seçilen en iyi performans gösteren kat, ardından ayrılmış test seti üzerinde modeli değerlendirmek için kullanılmaktadır.

İkili sınıflandırmada, varsayılan 0,5 karar eşiği, doğrulama seti üzerinde hesaplanan Youden'ın J indeksinden ($J = TPR - FPR$) türetilen optimal bir eşikle değiştirilmektedir. Karar eşiği, duyarlılık ile özgüllüğün toplamını eş zamanlı olarak en üst düzeye çıkaran işletim noktasını belirleyen Youden'ın J istatistiği (J = duyarlılık + özgüllük − 1) kullanılarak optimize edilmiştir. Bu kriter, hem kaçırılan saldırıların hem de yanlış alarmaların farklı operasyonel maliyetler doğurduğu STS bağlamları için

özellikle uygundur: kaçırılan bir saldırı tespit edilmeden bir sızmanın gerçekleşmesine zemin hazırlayabilirken, aşırı yanlış alarm oranı analist verimliliğini düşürmekte ve üretim ortamlarında alarm yorgunluğu riskini beraberinde getirmektedir.Raporlanan metrikler şunlardır: doğruluk, kesinlik, geri çağırma, F1-skoru, yanlış pozitif oranı (YPO), ROC-EAA, ortalama kesinlik (KG-EAA), örnek başına tespit süresi ve Verimlilik (verimlilik = örnek/sn). 5 katlı çapraz doğrulama, yalnızca *verinin %80'lik eğitim + doğrulama* kısmı (kısaca eğitim kümesi olarak anılacaktır) üzerinde uygulanmaktadır. *%20'lik test kümesi* tamamen ayrı tutulur ve eğitim veya model seçimi sırasında hiçbir aşamada kullanılmaz. Bu *%80'lik* bölüm içinde, her kat yaklaşık *toplam verinin %10'unu doğrulama kümesi*, yaklaşık *%70'ini ise eğitim kümesi* olarak oluşturur

Bu çalışmada uygulanan çapraz doğrulama yöntemi ayrıntılı olarak şu şekilde açıklanabilir,ğitim kümesi öncelikle 5 alt kümeye (kat) ayrılmıştır ve her yinelemede (iterasyonda):

- Bir kat, doğrulama kümesi olarak kullanılır;
- Kalan 4 kat, modelin eğitimi için kullanılır;
- Modelin performansı doğrulama kümesi üzerinde değerlendirilir.

Tablo 9. Çapraz doğrulama yöntemi.

İterasyon	Eğitim Kümesi	Doğrulama Kümesi
Kat 1	Kat 2, 3, 4, 5	Kat 1
Kat 2	Kat 1, 3, 4, 5	Kat 2
Kat 3	Kat 1, 2, 4, 5	Kat 3
Kat 4	Kat 1, 2, 3, 5	Kat 4
Kat 5	Kat 1, 2, 3, 4	Kat 5

Her satırın anlamı şu şekildedir:

- *Kat 1:* Model, eğitim kümesinin %80'i üzerinde eğitilmiş ve ilk kat (%20) üzerinde test edilmiştir. Elde edilen metrikler bu teste aittir.
- *Kat 2:* Aynı işlem uygulanmıştır; ancak bu kez ikinci kat test kümesi olarak kullanılmıştır.
- *Kat 3*: Aynı işlem uygulanmıştır; ancak bu kez üçüncü kat test kümesi olarak kullanılmıştır.
- *Kat 4:* Aynı işlem uygulanmıştır; ancak bu kez dördüncü kat test kümesi olarak kullanılmıştır.

- *Kat 5:* Aynı işlem uygulanmıştır; ancak bu kez beşinci kat test kümesi olarak kullanılmıştır.

4.3. İkili Sınıflandırma Sonuçları

4.3.1. CIC-IDS2017 İkili Sınıflandırma (NORMAL/SALDIRI)

4.3.1.1. CNN: CIC-IDS2017 İkili

CNN modeli, CIC-IDS2017 ikili görevinde hem yüksek ayırt edici kapasite hem de hesaplama verimliliği sergileyerek güçlü bir performans elde etmektedir. Beş kat boyunca model, normal ve saldırı trafiği arasında neredeyse mükemmel bir ayrımı yansıtan %99,82 ortalama doğruluk değerini korumaktadır. CNN İkili (CIC-IDS2017) Performans Özeti aşağıdaki tabloda verilmektedir:

Tablo 10. Önerilen CNN: CIC-IDS2017 ikili modelin beş katlı çapraz doğrulama

Kat	Doğruluk	Kesinlik	Geri Çağırma	F1-Skoru	YPO	ROC-EAA	KG-EAA	Tespit Süresi (ms/örnek)	Verimlilik (örnek/sn)
1	0.9972	0.9878	0.9983	0.9930	0.0030	0.9999	0.9983	0.0097	103237.7160
2	0.9974	0.9885	0.9984	0.9934	0.0029	0.9999	0.9984	0.0102	97958.1255
3	0.9975	0.9876	0.9997	0.9936	0.0031	0.9999	0.9997	0.0097	103545.1541
4	0.9971	0.9870	0.9985	0.9927	0.0032	0.9999	0.9985	0.0099	101174.0899
5	0.9983	0.9930	0.9986	0.9958	0.0017	1.0000	0.9986	0.0094	106071.0436

Tablo 11. Çapraz doğrulama özeti (ortalama ± standart sapma).

Metrik	Değer
Doğruluk (ortalama)	0,9975
Kesinlik : SALDIRI	0,9888
Geri Çağırma / Tespit Oranı (TPO)	0,9987
F1-Skoru : SALDIRI	0,9937
YPO	0,0019 (ortalama)

Tablo 11. (Devamı).

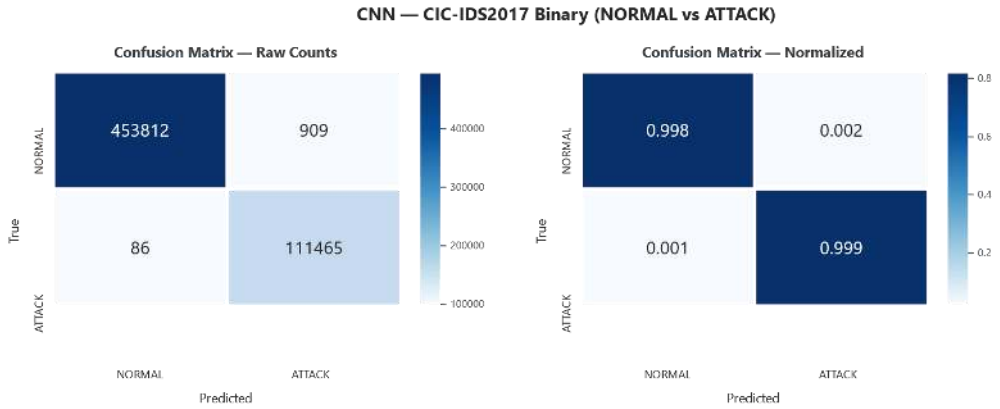
Metrik	Değer
ROC-EAA	0,9999
KG-EAA	0,9998
Tespit Süresi	0,0098 ms/örnek
Verimlilik	≈ 102,397 örnek/sn

Seçilen en iyi model: Kat 5 (en yüksek F1-Skoru) karar eşiği (Threshold): 0.4679

Tablo 12. Test kümesi sonuçları (İkili CNN – CIC-IDS2017)

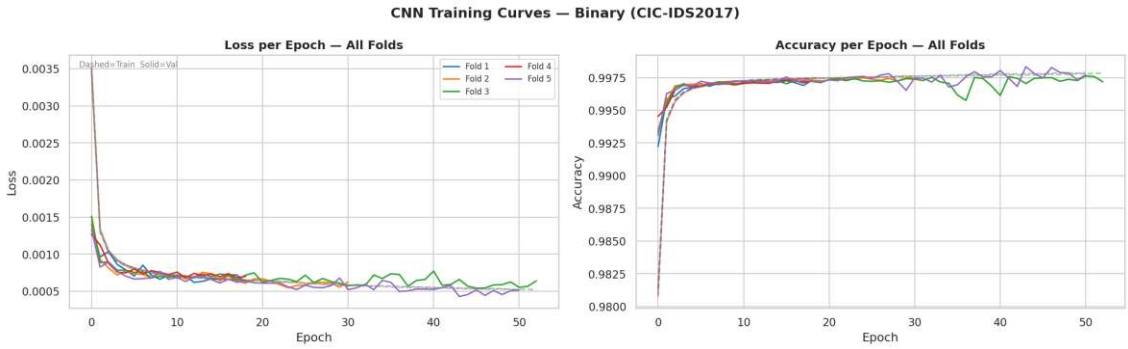
Metrik	Değer
Doğruluk (Accuracy)	0.9982
Kesinlik (Saldırı)	0,9919
Duyarlılık (Saldırı)	0,9992
F1-Skoru (Saldırı)	0,9956
Yanlış Pozitif Oranı	0,0020
Tespit Oranı	0,9992
ROC-EAA	0,9999
Ortalama Kesinlik	0,9998
Ortalama Tespit Süresi	0,0096 ms/örnek
Verimlilik	104,045 örnek/sn

Karışıklık matrisi, test setindeki 111.551 saldırı örneğinin 111.465'inin (%99,98) doğru şekilde tanımlandığını ve yalnızca 86'sının kaçırıldığını ortaya koymaktadır; bu olağanüstü tespit oranı, modelin yüksek dengesiz trafik dağılımları üzerinde genelleme kapasitesini doğrulamaktadır. Benzer şekilde, normal örnekler arasında 454.721'den yalnızca 909'u yanlışlıkla saldırı olarak işaretlenmekte ve 0,002'lik bir yanlış pozitif oranı elde edilmektedir.



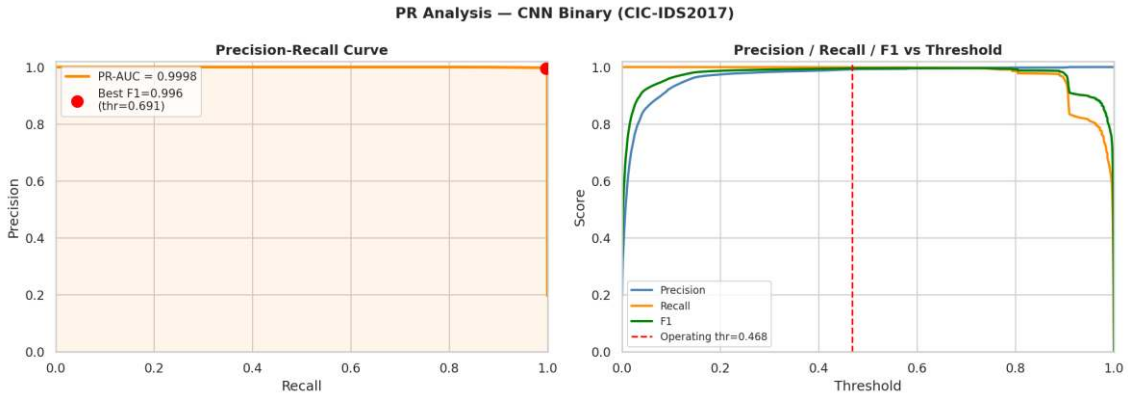
Şekil 18. CIC-IDS2017 ikili sınıflandırması için CNN'in karışıklık matrisleri (ham sayılar ve normalleştirilmiş).

Eğitim eğrileri, hızlı ve kararlı bir yakınsamayı doğrulamaktadır. Model, ilk 5 epoch içinde 0,995'in üzerinde bir doğruluğa ulaşmakta ve 20. epoch civarında düzgün bir şekilde yaklaşık 0,9975'e oturmaktadır; eğitim ve doğrulama eğrileri boyunca sıkı bir hizalama korunmakta ve bu durum aşırı öğrenmenin yokluğunun açık bir göstergesi olmaktadır. Kayıp eğrisi de bu davranışı yansıtarak erken epoch'larda keskin bir düşüş göstermekte ve ardından çok düşük bir değerde (≈ 0,0005) kararlı bir duruma gelmekte; eğitim ve doğrulama arasında ihmal edilebilir bir sapma gözlemlenmektedir.



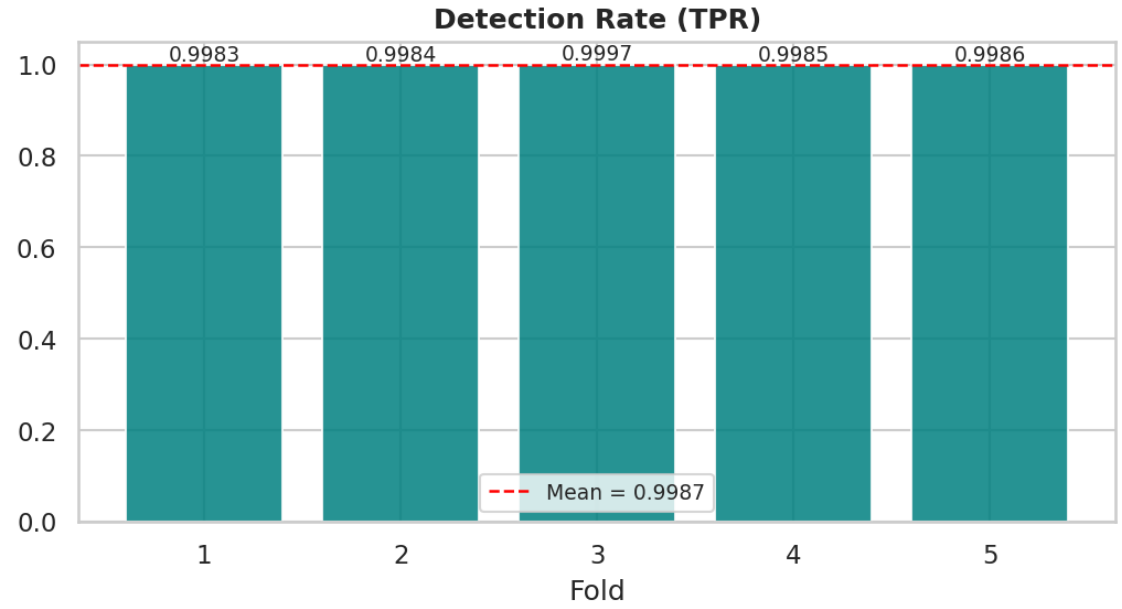
Şekil 19. CIC-IDS2017 ikili sınıflandırması için CNN'nin eğitim ve doğrulama öğrenme eğrileri (doğruluk ve kayıp).

ROC eğrisi de modelin ayırt edici gücünü doğrulamaktadır: EAA değeri doğrulama setinde 1,0000'e ulaşmakta ve Youden'ın J istatistiği aracılığıyla 0,468 optimal işletim eşiği belirlenmektedir. Bu durum, modelin olasılık uzayında iki sınıfı neredeyse mükemmel biçimde ayırdığını göstermektedir Katlar bazındaki metrikler yüksek düzeyde tutarlıdır.

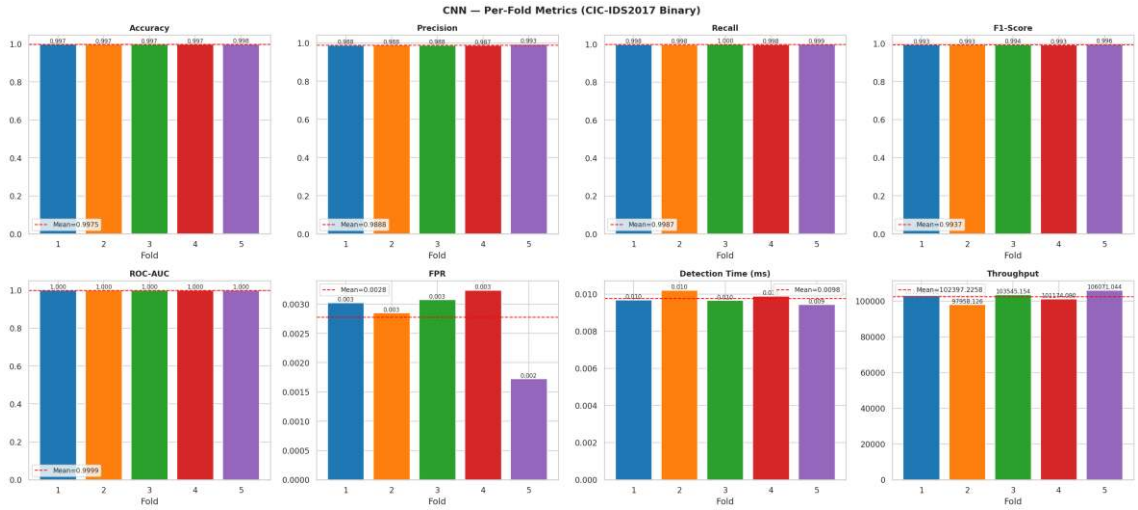


Şekil 20. CIC-IDS2017 ikili sınıflandırması için CNN'in ROC eğrisi, Youden'ın J eşik analizi ve katlama başına metrik dağılımları.

Operasyonel verimlilik açısından CNN, her örneği yaklaşık 0,0098 ms'de işlemekte ve saniyede 102.000'i aşan bir verimlilik elde etmektedir; bu da hem doğruluk hem de hızın kritik gereksinimler olduğu gerçek zamanlı dağıtım senaryoları için son derece uygundur.



Şekil 21. CIC-IDS2017 ikili sınıflandırması için CNN'nin katlama başına Tespit Oranı (TPO) ve Tespit Süresi.



Şekil 22. CNN-İkili CIC-IDS2017 için katlama başına metrikler (Doğruluk, Kesinlik, Geri Çağırma, F1, YPO, ROC-EAA, Tespit Süresi, Verimlilik).

4.3.2. CNN-BiLSTM: İkili Sınıflandırma (NORMAL/SALDIRI)

CIC-IDS2017 ikili görevinde bu hibrit model, CNN temel modeliyle son derece rekabetçi ve bazı metriklerde hafifçe üstün bir performans sergilemektedir.

Tablo 13. Önerilen CNN-BiLSTM: CIC-IDS2017 ikili modelin beş katlı çapraz.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Oranı	Tespit Süresi (ms)	Verimlilik (örnek/sn)
1	0,9969	0,9866	0,9978	0,9922	0,0033	0,9999	0,9978	0,0458	21.845,2172
2	0,9968	0,9860	0,9978	0,9918	0,0035	0,9999	0,9978	0,0459	21.771,7182
3	0,9968	0,9856	0,9981	0,9918	0,0036	0,9999	0,9981	0,0458	21.852,2420
4	0,9964	0,9838	0,9984	0,9910	0,0040	0,9999	0,9984	0,0458	21.853,1849
5	0,9969	0,9863	0,9982	0,9922	0,0034	0,9999	0,9982	0,0458	21.818,0437

Tablo 14. Çapraz doğrulama özeti (ortalama ± standart sapma)

Metrik	Değer
Doğruluk (ortalama)	0,9959
Kesinlik- SALDIRI	0,9857
Geri Çağırma / Tespit Oranı (TPO)	0,9981
F1-Skoru- SALDIRI	0,9918

Tablo 14. (Devamı)

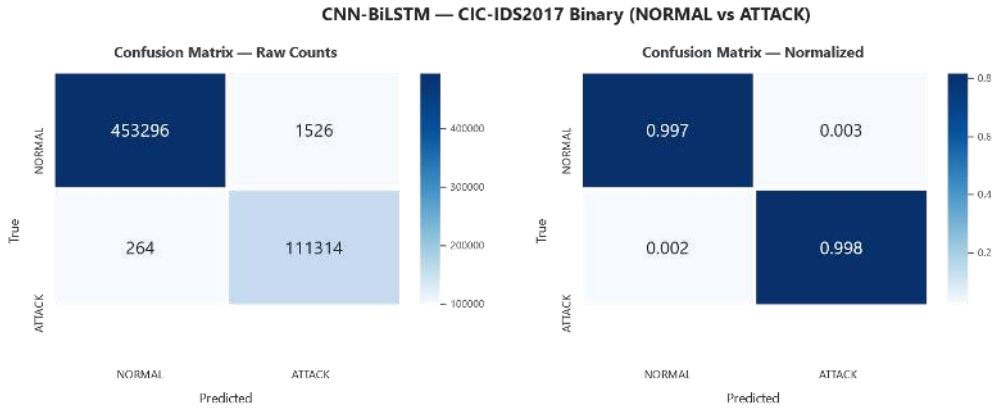
Metrik	Değer
YPO	0,0036 (ortalama)
ROC-EAA	0,9999
KG-EAA	0,9996
Tespit Süresi	0,0458 ms/örnek
Verimlilik	≈ 21.828 örnek/sn

Seçilen en iyi model: Kat 5 (en yüksek F1-Skoru) karar eşiği (Threshold): 0,5234.

Tablo 15. Test kümesi sonuçları (CNN-BiLSTM ikili sınıflandırma – CIC-IDS2017)

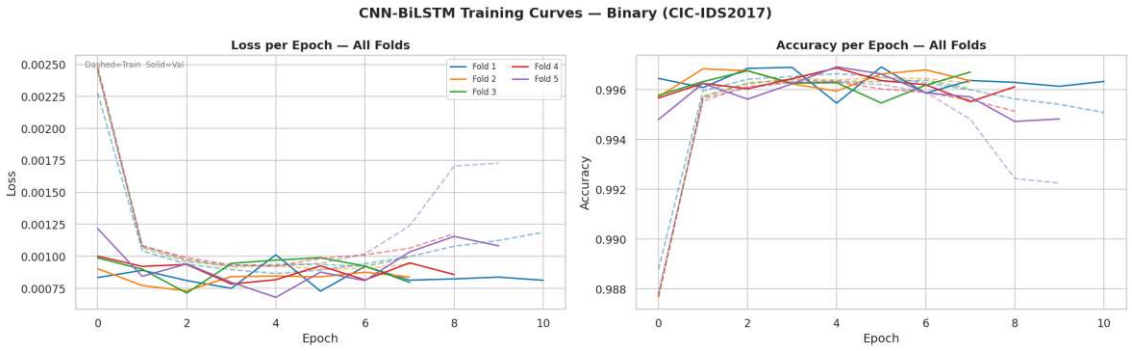
Metrik	Değer
Doğruluk	0,9968
Kesinlik (Saldırı)	0,9865
Duyarlılık (Saldırı)	0,9976
F1-Skoru (Saldırı)	0,9920
Yanlış Pozitif Oranı (FPR)	0,0034
Tespit Oranı (Detection Rate)	0,9976
ROC-EAA	0,9999
KG-EAA	0,9996
Ortalama Tespit Süresi	0,0464 ms/örnek
Verimlilik	21.548 örnek/sn

Karışıklık matrisi, modelin 454.822'den 453.296 normal örneği (%99,7) ve 111.578'den 111.314 saldırı örneğini (%99,8) doğru sınıflandırdığını doğrulamaktadır. 264 kaçırılan saldırı ve 1.526 yanlış alarm, CNN'e kıyasla küçük bir düşüşü temsil etmekte ve bu durum kısmen tekrarlayan bileşenin eklenen karmaşıklığına ve sınıf sınırlarına olan duyarlılığın artmasına bağlanabilmektedir.



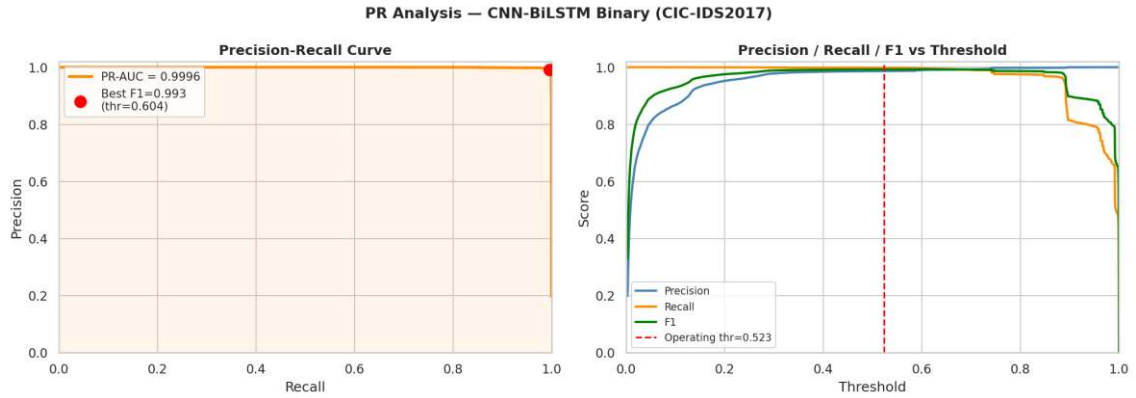
Şekil 23. CIC-IDS2017 ikili sınıflandırması için CNN-BiLSTM'nin karışıklık matrisleri (ham sayılar ve normalleştirilmiş).

Eğitim dinamikleri, CNN ile kıyaslandığında farklı bir yakınsama profili ortaya koymaktadır. Erken durdurma etkinleştirildiğinde model, 50 epoch yerine yaklaşık 10 epoch'ta yakınsama sağlamaktadır. Doğruluk, ilk iki epoch içinde yaklaşık 0,996'ya hızla yükselmekte ve kayıp çok hızlı biçimde 0,0025'ten 0,001'in altına düşmektedir. Ancak 7. epoch'tan itibaren eğitim ve doğrulama kaybı arasında hafif bir sapma gözlemlenmekte; bu durum, özellik düzeyinde sınırlı zamansal yapıya sahip veri setleri üzerinde eğitilen tekrarlayan bileşenli mimarilerin bilinen bir sınırlaması olan hafif aşırı öğrenmenin başlangıcına işaret etmektedir.



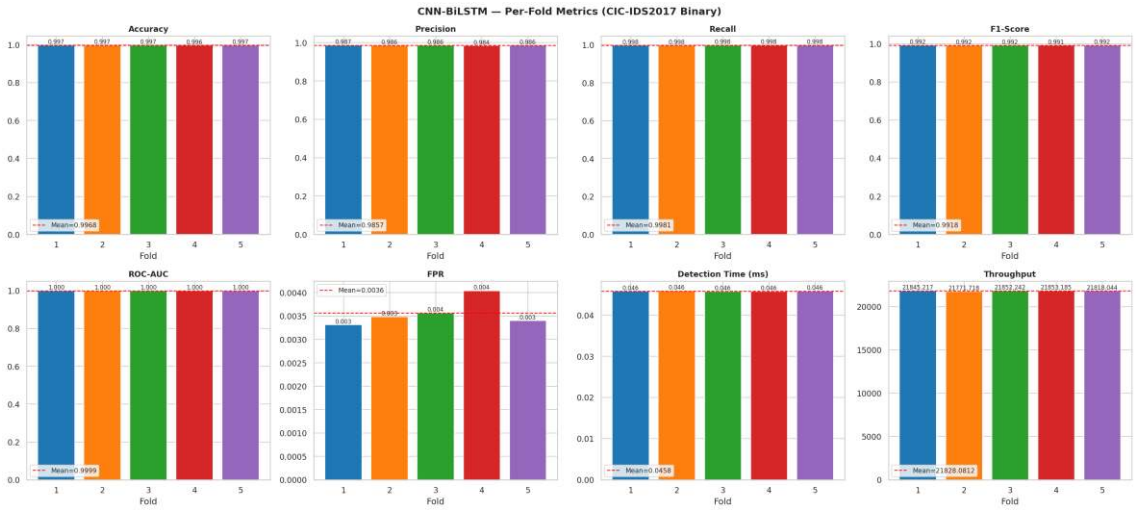
Şekil 24. CIC-IDS2017 ikili sınıflandırması için CNN-BiLSTM'nin eğitim ve doğrulama öğrenme eğrileri.

Buna karşın ROC-EAA 0,9999 düzeyinde kalmakta ve Kesinlik-Geri Çağırma EAA değeri 0,9996'ya ulaşmaktadır. Youden'ın J kriteri ile optimal eşik 0,523 olarak belirlenmekte; bu değer standart 0,5'e yakın olup iyi kalibre edilmiş olasılık çıktılarına işaret etmektedir. Katlama başına sonuçlar tutarlı kalmakta; tüm temel metriklerde standart sapmalar 0,002'nin altında seyretmektedir.



Şekil 25. CIC-IDS2017 ikili sınıflandırması için CNN-BiLSTM'nin ROC eğrisi, Kesinlik-Geri Çağırma eğrisi, Youden'ın J ve katlama başına metrik dağılımları.

CNN-BiLSTM'nin temel dengesi, hesaplama maliyetindedir. Model, CNN'den yaklaşık 4,7 kat daha yavaştır. Her bir örnek 0,0458 ms'de işlenmektedir. Bu da modelin saniyede yaklaşık 21.800 örnek işlediği anlamına gelmektedir. Bu, birçok izleme senaryosu için yeterli olmakla birlikte, yüksek trafiğe sahip ağ ortamlarında dezavantaj yaratma potansiyeline sahiptir.



Şekil 26. CNN-BiLSTM İkili CIC-IDS2017 için Katlama Başına Metrikler (Doğruluk, Kesinlik, Geri Çağırma, F1, YPO, ROC-EAA, Tespit Süresi, Verimlilik)

4.3.3. UNSW-NB15 İkili Sınıflandırma (NORMAL/SALDIRI)

4.3.3.1. CNN: UNSW-NB15 İkili Sınıf

UNSW-NB15 ikili görevinde CNN modeli, bu veri setinin daha yüksek doğası güçlüğünü yansıtarak CIC-IDS2017'ye kıyasla sağlam ancak hafifçe düşük bir performans sergilemektedir. Beş katlama boyunca ortalama doğruluk %95,23 düzeyinde seyretmekte ve tüm katlamalar boyunca kararlı bir şekilde korunmaktadır.

Tablo 16. Önerilen CNN: UNSW-NB15 ikili modelin beş katlı çapraz doğrulama sonuçları.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Oranı	Tespit Süresi (ms)	verimlilik (örnek/s)
1	0,9509	0,9837	0,9388	0,9607	0,0276	0,9925	0,9388	0,0119	84.280,51
2	0,9540	0,9855	0,9420	0,9632	0,0246	0,9935	0,9420	0,0163	61.515,05
3	0,9515	0,9795	0,9439	0,9614	0,0350	0,9922	0,9439	0,0117	85.501,89
4	0,9520	0,9859	0,9383	0,9615	0,0238	0,9935	0,9383	0,0120	83.462,69
5	0,9520	0,9837	0,9405	0,9616	0,0276	0,9927	0,9405	0,0117	85.514,17

Tablo 17. Çapraz doğrulama özeti (ortalama ± standart sapma).

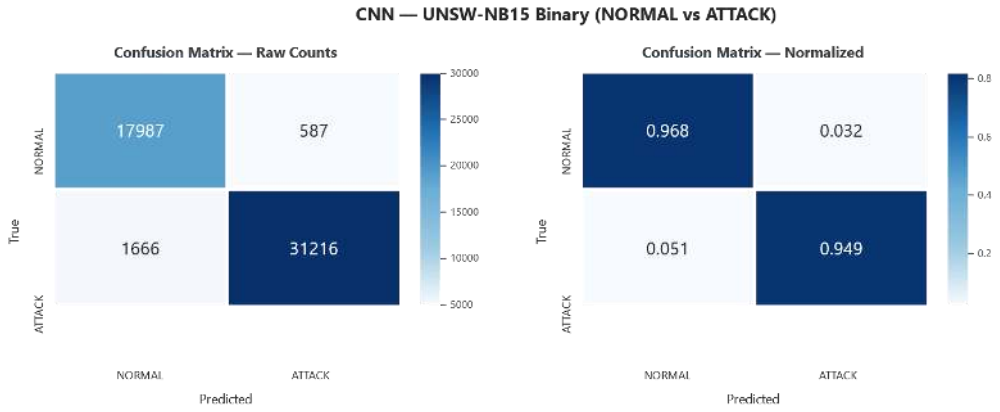
Metrik	Değer
Doğruluk (ortalama)	0,9523
Kesinlik - SALDIRI	0,9638
Geri Çağırma / Tespit Oranı (TPO)	0,9407
F1-Skoru - SALDIRI	0,9617
YPO	0,0317 (ortalama)
ROC-EAA	0,9942
KG-EAA	0,9964
Tespit Süresi	0,0127 ms/örnek
Verimlilik	≈ 80.055 örnek/sn

Seçilen en iyi model: Kat 2 (en yüksek F1-Skoru) karar eşiği (Threshold): 0,4708.

Tablo 18. Test kümesi sonuçları (İkili CNN – UNSW-NB15).

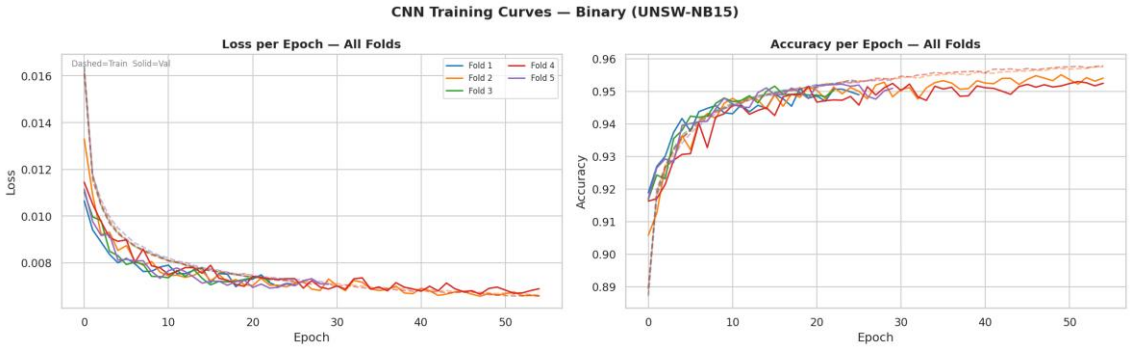
Metrik	Değer
Doğruluk	0,9562
Kesinlik (Saldırı)	0,9815
Duyarlılık (Saldırı)	0,9493
F1-Skoru (Saldırı)	0,9652
Yanlış Pozitif Oranı	0,0316
Tespit Oranı	0,9493
ROC-EAA	0,9935
Ortalama Kesinlik	0,9964
Ortalama Tespit Süresi	0,0122 ms/örnek
Verimlilik	82.168 örnek/sn

Karışıklık matrisi, toplam 32.882 saldırı örneğinden CNN'in 31.216'sını (%94,9) doğru tespit ettiğini ve 1.666'sını normal olarak yanlış sınıflandırdığını göstermektedir. Öte yandan toplam 18.574 normal örneğin 587'si yanlışlıkla saldırı olarak etiketlenmiş; bu durum yaklaşık %3,2'lik bir yanlış pozitif oranına karşılık gelmektedir. CIC-IDS2017'ye kıyasla daha yüksek olan bu YPO, UNSW-NB15'in daha zorlu yapısıyla tutarlıdır; zira bu veri setinde normal ve saldırı trafiği dağılımları daha belirgin biçimde örtüşmektedir.



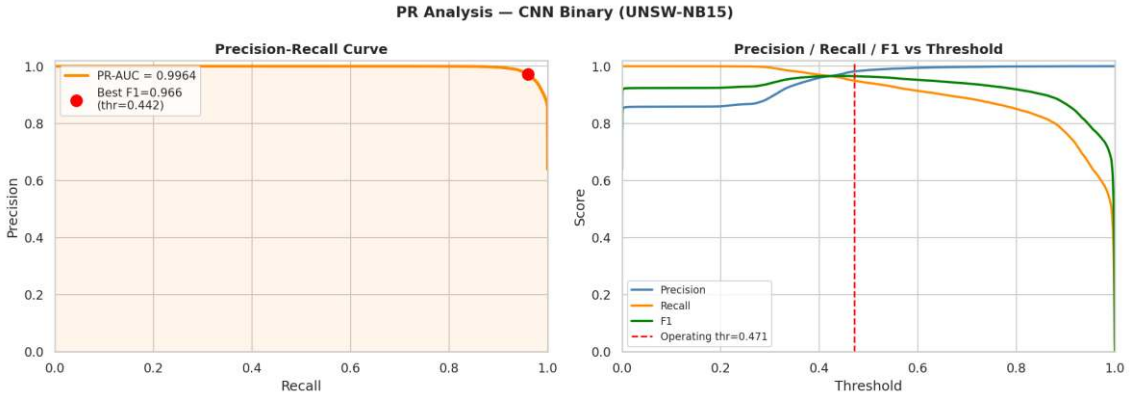
Şekil 27. UNSW-NB15 ikili sınıflandırması için CNN'in karışıklık matrisleri (ham sayılar ve normalleştirilmiş).

Eğitim eğrisi düzgün ve kademeli bir ilerleme sergilemektedir. Doğruluk, 50 epoch boyunca 0,89'dan yaklaşık 0,955'e istikrarlı biçimde yükselmekte; doğrulama eğrisi ise tüm süreç boyunca eğitim eğrisini yakından takip etmekte ve bu durum iyi genellemeye işaret etmektedir. Kayıp değeri yaklaşık 0,016'dan 0,007 civarında kararlı bir platoya inerken eğitim ve doğrulama kaybı arasında belirgin bir sapma gözlemlenmemektedir; bu, veri setinin karmaşıklığı göz önüne alındığında umut verici bir işarettir.



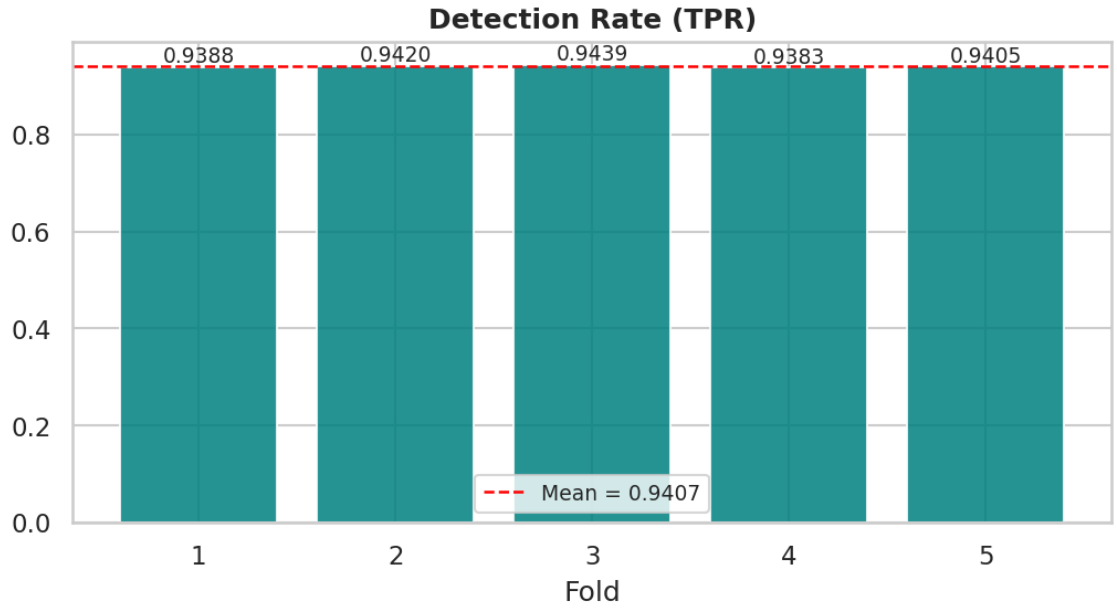
Şekil 28. UNSW-NB15 ikili sınıflandırması için CNN'nin eğitim ve doğrulama öğrenme eğrileri.

ROC-EAA değeri doğrulama setinde 0,9942'ye ulaşmakta ve Kesinlik-Geri Çağırma EAA değeri 0,9964 olarak saptanmaktadır; her iki yüksek değer de modelin daha dengeli ve karmaşık sınıf yapısına karşın güçlü ayırt edici kapasitesini doğrulamaktadır. Optimal eşik, Youden'ın J yöntemiyle 0,471 olarak belirlenmektedir. Katlama başına metrikler çok düşük varyans sergilemekte ve sonuçların sağlamlığını ve yeniden üretilebilirliğini onaylamaktadır.



Şekil 29. UNSW-NB15 ikili sınıflandırması için CNN'in ROC eğrisi, Kesinlik-Geri Çağırma eğrisi, Youden'ın J ve katlama başına metrik dağılımları.

Operasyonel açıdan CNN, her örneği ortalama 0,0127 ms'de işlemekte ve saniyede yaklaşık 80.000 örnek verimi elde etmektedir; bu güçlü sonuç, modeli bu daha zorlu veri seti üzerinde bile gerçek zamanlı dağıtım için uygun bir aday konumuna taşımaktadır.



Şekil 30. UNSW-NB15 ikili sınıflandırması için CNN'in katlama başına Tespit Oranı (TPO) ve Tespit Süresi.

4.3.3.2. CNN-BiLSTM: UNSW-NB15 İkili Sınıf

UNSW-NB15 ikili görevinde CNN-BiLSTM modeli, bu veri kümesindeki tüm değerlendirilen konfigürasyonlar arasında doğruluk ve tespit oranının en iyi genel dengesini elde etmektedir. Beş katlama boyunca ortalama doğruluk %96,82'ye ulaşmakta; bu değer aynı veri seti üzerindeki CNN'den kayda değer bir iyileşmeye işaret etmektedir.

Tablo 19. Önerilen CNN-BiLSTM: UNSW-NB15 ikili modelin beş katlı çapraz doğrulama sonuçları.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Süresi (ms)	Verimlilik (örnek/sn)
1	0,7582	0,5094	0,6381	0,5112	0,0261	0,9617	0,0488	20.492,289
2	0,7683	0,5101	0,6226	0,5178	0,0251	0,9628	0,0447	22.368,1747
3	0,7649	0,5097	0,6234	0,5168	0,0253	0,9610	0,0443	22.590,3756

Tablo 19. (Devamı)

4	0,7671	0,5116	0,6266	0,5133	0,0253	0,9601	0,0446	22.429,1370
5	0,7615	0,5154	0,6498	0,5116	0,0258	0,9599	0,0445	22.465,9349

Tablo 20. Çapraz doğrulama özeti (ortalama ± standart sapma).

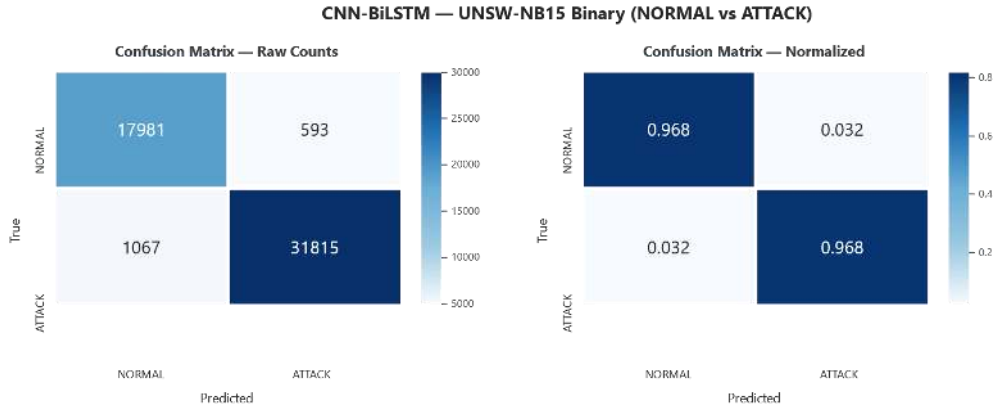
Metrik	Değer
Doğruluk (ortalama)	0,9682
Kesinlik- SALDIRI	0,9865
Geri Çağırma	0,9594
F1-Skoru- SALDIRI	0,9728
YPO	0,0325 (ortalama)
ROC-EAA	0,9977
Tespit Süresi	0,0201 ms/örnek
Verimlilik	≈ 49.853 örnek/sn

Seçilen en iyi model: Kat 2 (en yüksek F1-Skoru = 0,5178).

Tablo 21. Test kümesi sonuçları (Çok sınıflı CNN – UNSW-NB15)

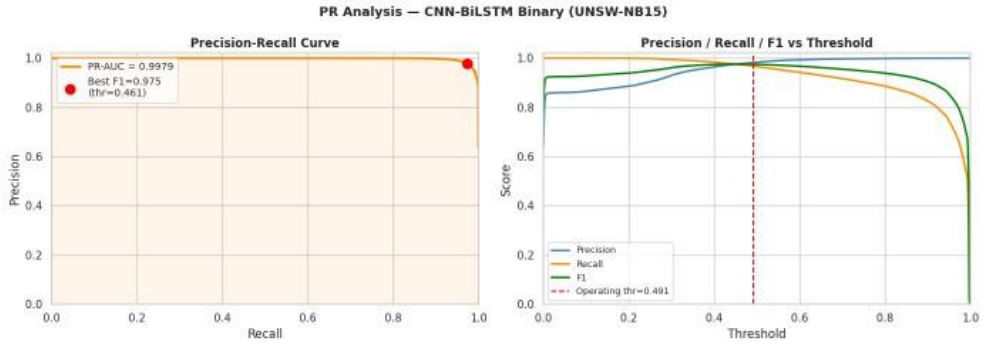
Metrik	Değer
Doğruluk	0,7643
Makro Kesinlik	0,5154
Makro Duyarlılık	0,6462
Makro F1-Skoru	0,5258
Makro Yanlış Pozitif Oranı	0,0255
ROC-EAA	0,9637
Ortalama Tespit Süresi	0,0445 ms/örnek
Verimlilik	22.476 örnek/saniye

Karışıklık matrisi, modelin 32.882 saldırı örneğinden 31.815'ini (%96,8) doğru tespit ettiğini ortaya koymaktadır; bu değer CNN için 31.216'dır ve yaklaşık 600 ek doğru tespit edilen saldırıyı temsil eden anlamlı bir kazanımı yansıtmaktadır. Yanlış negatif sayısı 1.666'dan (CNN) 1.067'ye (CNN-BiLSTM) düşerek yaklaşık %36'lık bir azalma sağlamaktadır. Normal tarafta ise 18.574 örnekten 593'ü yanlışlıkla saldırı olarak sınıflandırılmakta ve CNN ile karşılaştırılabilir düzeyde %3,2'lik bir yanlış pozitif oranı elde edilmektedir.



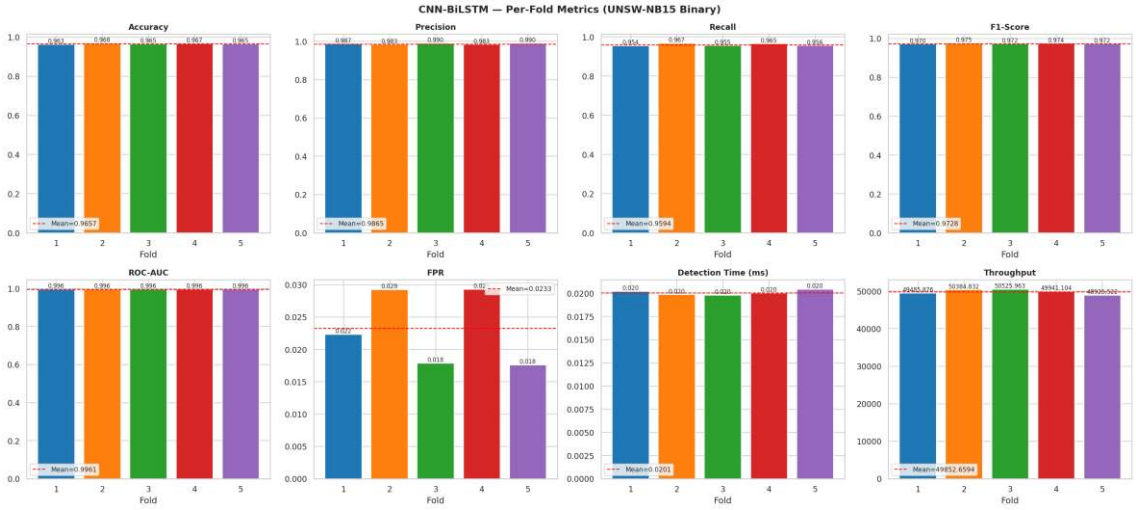
Şekil 31. UNSW-NB15 ikili sınıflandırması için CNN-BiLSTM'nin karışıklık matrisleri (ham sayılar ve normalleştirilmiş).

CNN-BiLSTM'nin bu veri seti üzerindeki eğitim dinamikleri özellikle kararlı bir seyir izlemektedir. 50 epoch boyunca doğruluk eğrisi 0,88'den yaklaşık 0,97'ye düzgün bir şekilde yükselmekte; eğitim ve doğrulama eğrileri sıkı bir hizalamayı korurken doğrulama sonraki epoch'larda hafifçe önde seyretmektedir; bu alışılmadık ancak olumlu bir işaret, mükemmel genellemeyi yansıtmaktadır. Kayıp değeri 0,018'den yaklaşık 0,006 platosuna inerken eğitim ve doğrulama birbirine yakın bir şekilde yakınsama sağlamaktadır. ROC-EAA 0,9977'ye ulaşmakta ve Youden'ın J yöntemiyle optimal eşik 0,491 olarak belirlenmektedir.



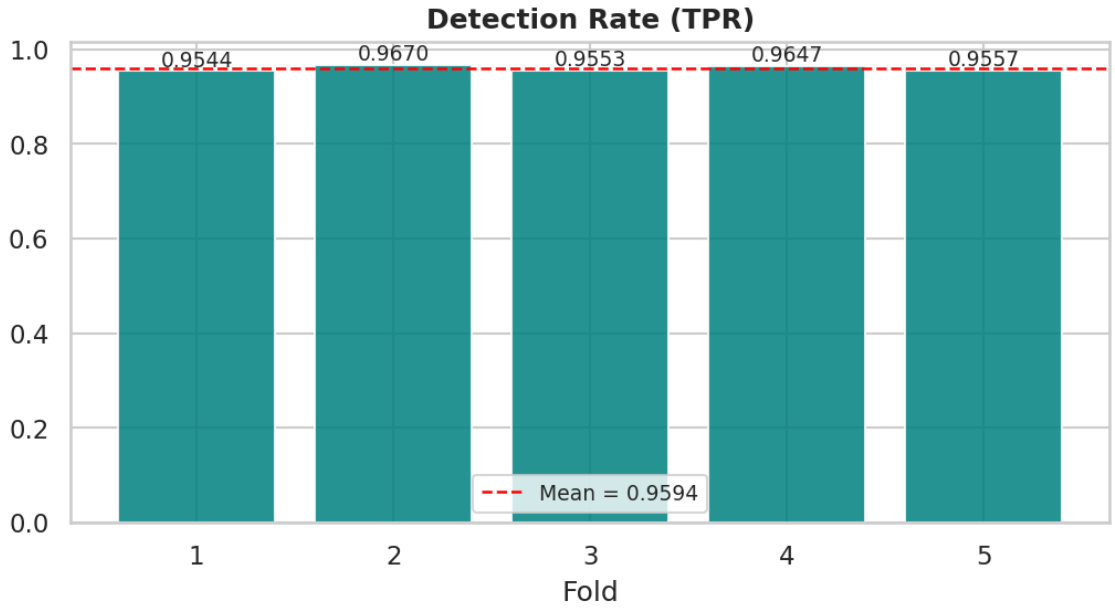
Şekil 32. UNSW-NB15 ikili sınıflandırması için CNN-BiLSTM'nin eğitim eğrileri, ROC eğrisi ve Youden'ın J eşik analizi.

Katlama başına metrikler olağanüstü bir kararlılık sergilemektedir: tespit oranı (TPO) beş katlama boyunca 0,9544 ile 0,9670 arasında değişmekte (ortalama = 0,9594) ve SALDIRI sınıfı için F1-Skoru 0,9704 ile 0,9750 arasında dalgalanmaktadır (ortalama = 0,9728). Bu dar aralıklar, modelin sağlamlığını ve farklı veri bölümleri arasında tutarlı biçimde genelleme yapabilme kapasitesini doğrulamaktadır.



Şekil 33. UNSW-NB15 ikili sınıflandırması için CNN-BiLSTM'nin katlama başına metrik dağılımları.

Verimlilik açısından CNN-BiLSTM, aynı modelin CIC-IDS2017 (0,0458 ms) üzerindeki süresinden daha hızlı olan 0,0201 ms'de her örneği işlemektedir; bu durum muhtemelen daha küçük toplu iş boyutuna (256'ya karşı 512) ve UNSW-NB15'in daha küçük veri seti boyutuna bağlıdır. Saniyede yaklaşık 49.853 örneğin verimi CNN'den (80.055) düşük olmakla birlikte, gerçek zamanlıya yakın izleme uygulamaları için kabul edilebilir bir aralıkta kalmaktadır.



Şekil 34. UNSW-NB15 ikili sınıflandırması için CNN-BiLSTM'nin katlama başına Tespit Oranı (TPO) ve Tespit Süresi.

4.3.4. Özet ve Modeller Arası Tartışma

Bu bölümde değerlendirilen dört ikili konfigürasyondaki sonuçlar, derin öğrenme tabanlı saldırı tespit sistemlerinin tasarımı için önemli çıkarımlar içeren birkaç tutarlı örüntüyü gün yüzüne çıkarmaktadır.

İlk olarak, her iki mimari de CIC-IDS2017 üzerinde olağanüstü bir performans sergilemekte; 0,9999 ROC-EAA ve %99,8'in üzerinde tespit oranları elde etmektedir. Bu güçlü performans, açıkça ayrılabilir trafik örüntüleri ve saldırı trafiğinin iyi temsil edildiği yüksek dengesiz sınıf dağılımına sahip olan veri setinin karakteristik özellikleriyle tutarlıdır. Buna karşın UNSW-NB15; daha dengeli dağılımı ve daha çeşitli saldırı türleriyle belirgin bir performans açığını gün yüzüne çıkarmaktadır: tespit oranları %94-96'ya düşmekte ve her iki model için YPO yaklaşık %3,2'ye yükselmektedir. Bu durum, genelleme sonuçları çıkarmadan önce STES modellerini birden fazla kıyaslama üzerinde değerlendirmenin önemini vurgulamaktadır.

İkinci olarak, CNN-BiLSTM UNSW-NB15 veri setinde Tespit Oranı (Geri Çağırma) açısından sürekli olarak CNN'yi geride bırakmaktadır (+1,9 yüzde puanı); CIC-IDS2017'de ise karşılaştırılabilir sonuçlar üretmektedir. Bu bulgu, çift yönlü LSTM bileşeninin özellikle karmaşık trafik ortamlarında bilgilendirici olan tamamlayıcı zamansal bağımlılıkları özellik uzayında yakaladığına işaret etmektedir. Ancak bu kazanım, önemli ölçüde daha yüksek bir hesaplama maliyetiyle gelmektedir: CNN-BiLSTM, tespit gecikmesi açısından CNN'den 4 ila 6 kat daha yavaş olup bu durum gecikmeye duyarlı dağıtım bağlamlarında dikkatle değerlendirilmelidir.

Üçüncü olarak, CNN CIC-IDS2017 üzerinde saniyede 100.000'i aşan istisnai verimiyle öne çıkmakta; bu değere neredeyse mükemmel tespit oranlarını korurken ulaşmaktadır.

Çok sınıflı sınıflandırma konfigürasyonlarına (CIC-IDS2017 için 15 sınıf ve UNSW-NB15 için 10 sınıf) ilişkin sonuçlar aşağıdaki bölümde sunulmakta ve tartışılmaktadır.

4.4. Çok Sınıflı Sınıflandırma Sonuçları

Çok sınıflı sınıflandırma, ikili saldırı tespitine kıyasla önemli ölçüde daha zorlu bir görev teşkil etmektedir. Modelin yalnızca normal trafiği kötü niyetli trafikten ayırt etmesi yeterli değildir; aynı zamanda potansiyel olarak dengesiz sınıflar arasındaki belirli saldırı türünü de doğru şekilde tanımlaması gerekmektedir. Metrikler makro ortalama (ağırlıksız) olarak raporlanmakta; bu metodolojik tercih, modellerin nadir saldırıları tanımadaki gerçek kapasitesini değerlendirmek açısından azınlık sınıflarındaki düşük performansı eşit biçimde cezalandırmaktadır.

4.4.1. CIC-IDS2017: Çok Sınıflı Sınıflandırma (15 Sınıf)

4.4.1.1. CNN: CIC-IDS2017 Çok Sınıflı

CIC-IDS2017 çok sınıflı görevinde CNN, söz konusu sınıf sayısı göz önüne alındığında dikkat çekici biçimde yüksek bir genel performans sergilemektedir. Genel doğruluk ortalama %98,93 düzeyindedir.

Tablo 22. 5 Katlı çapraz doğrulama sonuçları.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Süresi (ms)	Verimlilik (örnek/s)
1	0,9838	0,6345	0,9173	0,6705	0,0011	0,9974	0,0096	103.813,4951
2	0,9909	0,6709	0,8985	0,7079	0,0006	0,9938	0,0095	104.925,7770
3	0,9897	0,6333	0,9256	0,6664	0,0007	0,9994	0,0097	102.638,9380
4	0,9908	0,6893	0,9128	0,7193	0,0006	0,9990	0,0096	104.224,9438
5	0,9913	0,7023	0,9223	0,7399	0,0006	0,9859	0,0095	104.851,7884

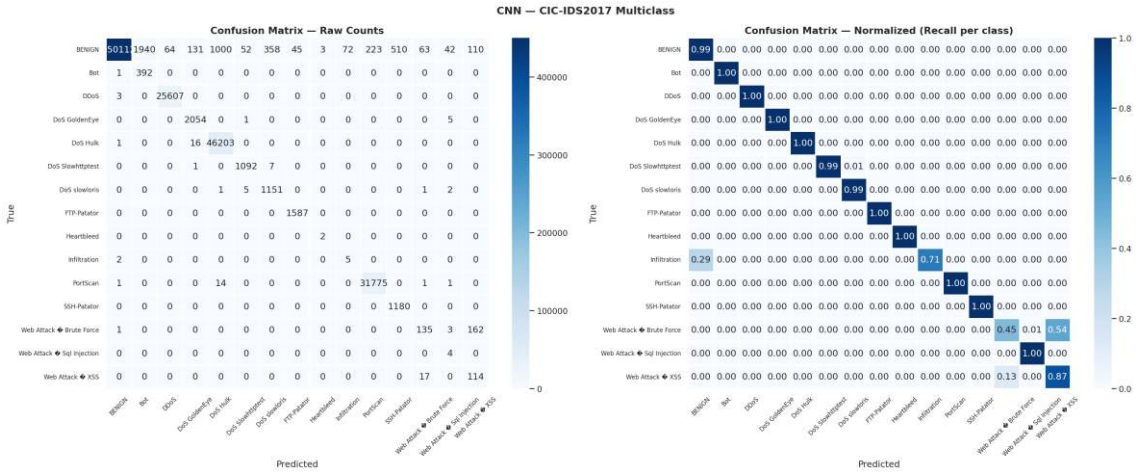
Tablo 23. Çapraz doğrulama özeti (ortalama ± standart sapma). CIC-IDS2017 üzerinde CNN çok sınıflı performansının özeti (makro ortalama, 5 katlama)

Metrik	Değer
Doğruluk (ortalama)	0,9893
Kesinlik (makro)	0,6661
F1-Skoru (makro)	0,7001
YPO (makro)	0,0003
ROC-EAA (makro OvR)	0,9951
Tespit Süresi	0,0096 ms/örnek
Verimlilik	≈ 104.091 örnek/sn

Seçilen en iyi model: Kat 5 (en yüksek Makro F1-Skoru = 0,7399).

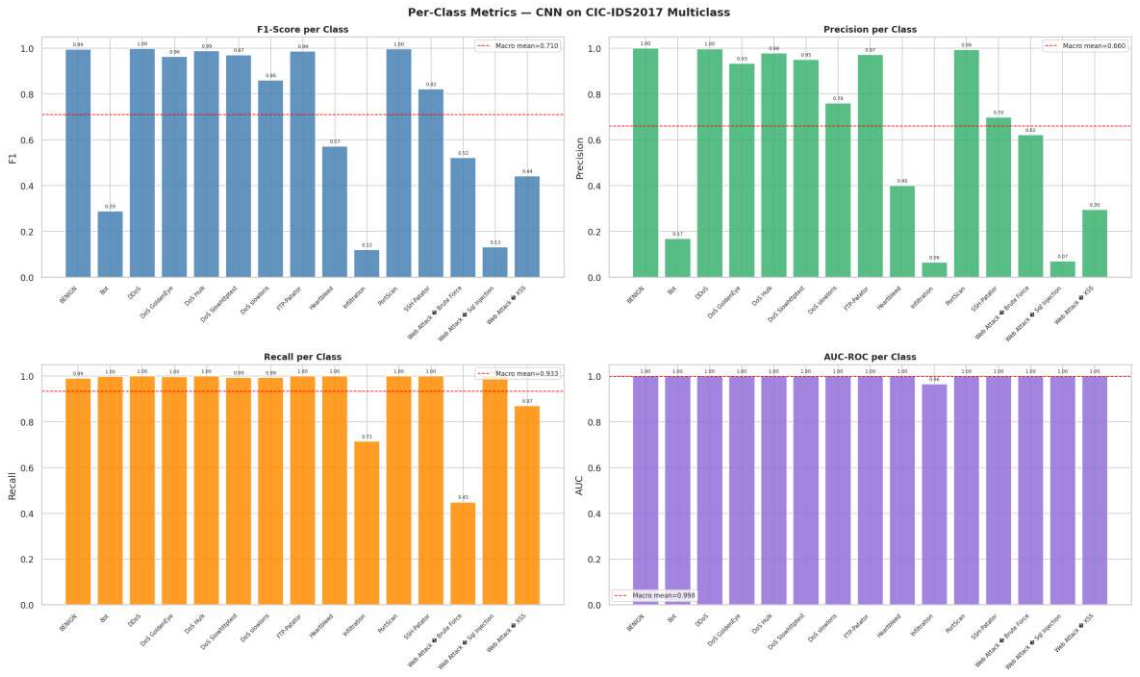
Tablo 24. Test kümesi sonuçları (Çok sınıflı CNN – CIC-IDS2017).

Metrik	Değer
Doğruluk	0,9914
Makro Kesinlik	0,6601
Makro Duyarlılık	0,9334
Makro F1-Skoru	0,7102
Makro Yanlış Pozitif Oranı	0,0006
ROC-EAA	0,9975
Ortalama Tespit Süresi	0,0099 ms/örnek
Verimlilik	101.385 örnek/sn



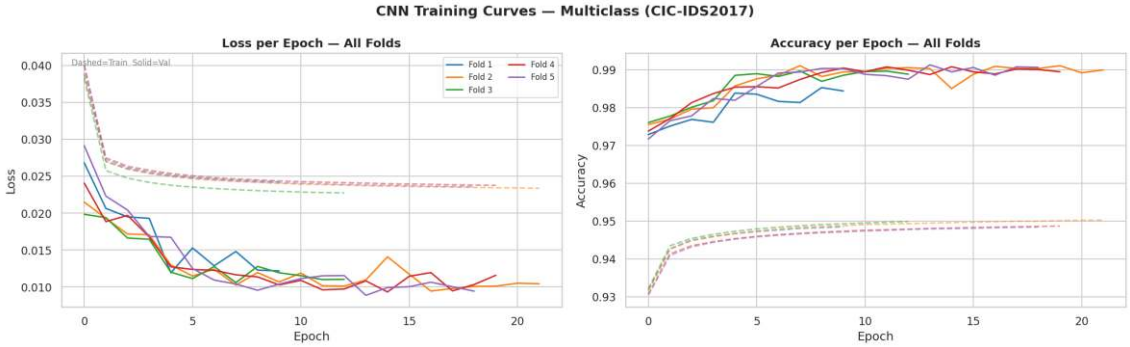
Şekil 35. CIC-IDS2017 çok sınıflı sınıflandırması için CNN'in karışıklık matrisleri (ham sayılar ve gerçek sınıfa göre normalleştirilmiş).

Öte yandan, makro ROC-EAA 0,9951'e ulaşmaktadır; bu da modelin, 0,5 eşiğinde sınıflandırma hataları yapsa bile her sınıfı diğerlerine karşı olasılık uzayında doğru biçimde ayırt ettiği anlamına gelmektedir. Sınıf başına ROC ve Kesinlik-Geri Çağırma eğrileri bu yorumu desteklemektedir: sınıfların büyük çoğunluğu (NORMAL, DDoS, DoS GoldenEye, DoS Hulk, DoS Slowhttptest, FTP-Patator, PortScan, SSH-Patator) 0,99'un üzerinde EAA elde ederken Infiltration (ROC EAA ≈ 0,904) ve Web Saldırıları (düşük KG-EAA: SQL Injection için ~0,154, XSS için ~0,363) birincil güçlük noktaları olmaya devam etmektedir.



Şekil 36. CIC-IDS2017 çok sınıflı sınıflandırması için CNN1D'nin sınıf başına metrikleri (F1, Kesinlik, Geri Çağırma, EAA-ROC).

Öğrenme eğrisi, erken durdurma öncesinde 20-22 epoch boyunca hızlı ve kararlı bir yakınsamayı ortaya koymaktadır. Doğrulama doğruluğu 2. epoch'tan itibaren 0,97'yi aşmakta ve katlamalar arası varyans çok düşük kalmaktadır. Doğrulama kaybı 0,025'ten yaklaşık 0,010'a düşmekte ve eğitim kaybının oldukça altında seyretmektedir; bu, 15 sınıflı bir görev için dikkat çekici olan mükemmel genellemenin işaretidir. Operasyonel verimlilik açısından CNN, 104.091 örnek/sn verimi ile her örneği 0,0096 ms'de işlemekte ve çıkarım hızında üstünlüğünü yinelemektedir.



Şekil 37. CIC-IDS2017 çok sınıflı sınıflandırması için CNN'in eğitim eğrileri ve katlama başına metrikleri.

4.4.1.2. CNN-BiLSTM: CIC-IDS2017 Çok Sınıflı

Aynı 15 sınıflı çok sınıflı görevde CNN-BiLSTM, CNN'den son derece farklı bir performans profili ve her şeyden önce kapsamlı bir analizi zorunlu kılan sorunlu bir eğitim dinamiği sergilemektedir.

Tablo 25. 5 Katlı çapraz doğrulama sonuçları.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Süresi (ms)	Verimlilik (örnek/sn)
	0.9977	0.8164	0.7513	0.7382	0.0003	0.9997	0.0309	32377.166
	0.9976	0.7736	0.7203	0.7305	0.0003	0.9983	0.0307	32547.9404
	0.9980	0.8169	0.7752	0.7803	0.0003	0.9998	0.0303	33042.2645
	0.9976	0.7712	0.7233	0.7325	0.0003	0.9980	0.0308	32497.0487
	0.9979	0.8406	0.7524	0.7522	0.0003	0.9913	0.0284	35155.4861

Tablo 26. CIC-IDS2017 üzerinde CNN-BiLSTM çok sınıflı performansının özeti (makro ortalama, 5 katlama).

Metrik	Değer
Doğruluk (ortalama)	0,9977
Kesinlik (makro)	0,7945
Geri Çağırma (makro)	0,7425
F1-Skoru (makro)	0,7454
YPO (makro)	0,7454
ROC-EAA (makro OvR)	0,7454
Verimlilik	≈ 32616,1053 örnek/sn

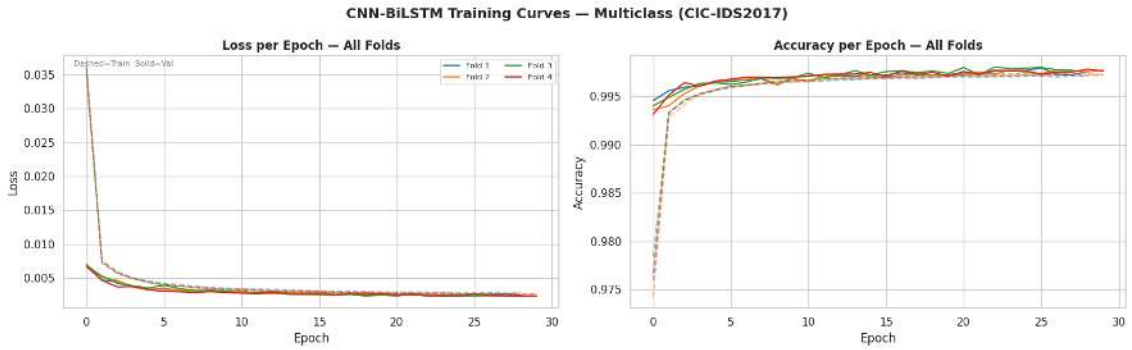
En iyi sonuç veren Kat 3 kullanıldı (en iyi makro F1 = 0,7803).

Tablo 27. Test kümesi sonuçları.

Metrik	Değer
Doğruluk	0,9982
Kesinlik (Makro)	0,8279
Duyarlılık	0,7938
F1-Skoru (Makro)	0,7987
Yanlış Pozitif Oranı	0,0003
ROC-AUC	0,9988
Tespit Süresi	0,0316 ms/örnek
Verimlilik	31,609 örnek/sn

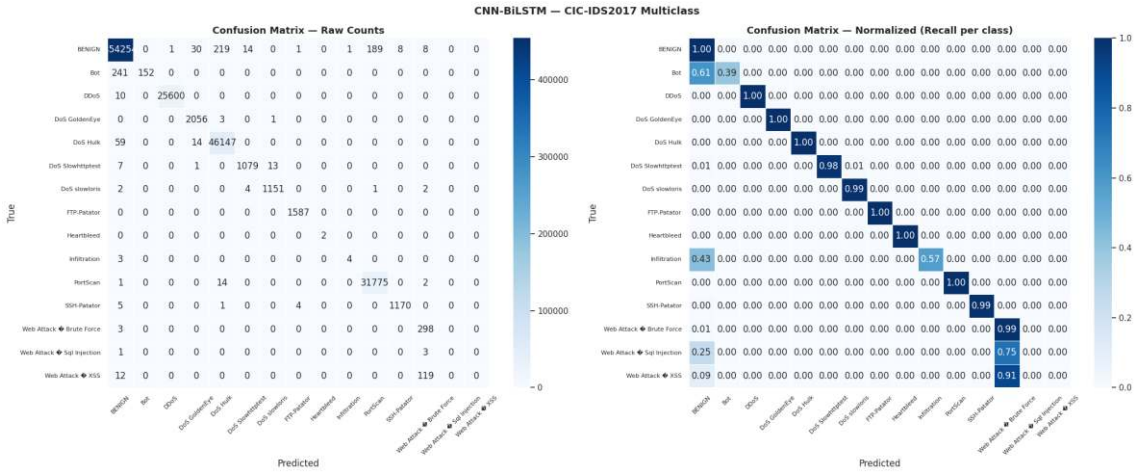
Eğitim dinamikleri (Şekil 38), ilk 5-10 epoch içinde hızlı bir yakınsama göstermekte, eğitim ve doğrulama loss/accuracy eğrileri dört fold boyunca birbirine çok yakın seyretmektedir; bu da aşırı öğrenmenin (overfitting) bulunmadığını ve Focal Loss formülasyonunun erken durdurma (early stopping) ile birlikte CNN-BiLSTM

mimarisini etkili şekilde düzenlediğini doğrulamaktadır. Ancak bu küresel yakınsama örüntüsünün kendisi sınıf dengesizliğinin bir ürünüdür: BENIGN ve yüksek frekanslı saldırı kategorileri loss sinyaline sayısal olarak baskın olduğundan, epoch düzeyindeki eğriler azınlık sınıflarının öğrenilmesi hakkında sınırlı bilgi sunmakta ve Şekil 1'deki sınıf bazlı metriklerle birlikte yorumlanması gerekmektedir:



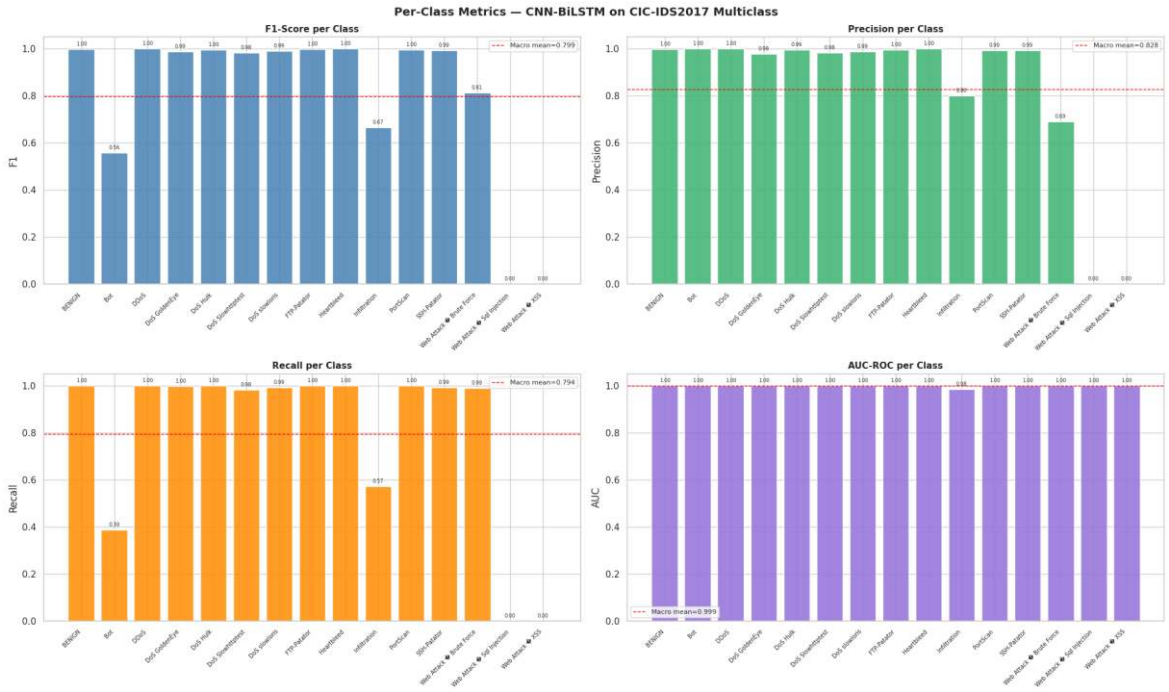
Şekil 38. CIC-IDS2017 çok sınıflı sınıflandırmasında CNN-BiLSTM kararsızlığını gösteren eğitim eğrileri.

Karışıklık matrisi (Şekil 39), bu başarısızlık örüntülerini daha da somutlaştırmaktadır: Bot trafiği ağırlıklı olarak BENIGN olarak yanlış sınıflandırılmakta (393 örnekten 241'i, %61), bu da düşük hacimli bot bağlantıları ile meşru trafik arasındaki davranışsal benzerlikle tutarlıdır; Sql Injection ve XSS örnekleri ise neredeyse tamamen Web Attack – Brute Force kategorisi tarafından emilmektedir, bu da modelin aynı anlamsal aile (Web Attack) içindeki nadir alt sınıfları baskın kardeş sınıfa çökertme eğilimini yansıtmaktadır. Bu bulgular, gelecekteki çalışmaların yalnızca küresel SMOTE tabanlı yeniden dengelemeye dayanmak yerine, hiyerarşik veya iki aşamalı sınıflandırma stratejilerini (ikili saldırı/normal tespiti ardından ince taneli alt sınıflandırma) veya yalnızca en ciddi şekilde az temsil edilen kategorilere yönelik hedefli aşırı örnekleme yaklaşımlarını araştırması gerektiğini ortaya koymaktadır.



Şekil 39. CIC-IDS2017 çok sınıflı sınıflandırması için CNN-BiLSTM'nin karışıklık matrisleri ve sınıf başına metrikleri.

Sınıf bazında değerlendirme (Şekil 40), 0.9982'lik toplam doğruluk değerinin gizlediği belirgin bir performans heterojenliğini ortaya koymaktadır. On beş trafik kategorisinden onu (BENIGN, DDoS, DoS GoldenEye, DoS Hulk, DoS Slowhttptest, DoS slowloris, FTP-Patator, Heartbleed, PortScan ve SSH-Patator) neredeyse mükemmel F1 skorlarına (≥0.98) ulaşırken, üç azınlık sınıf belirgin şekilde düşük performans göstermektedir: Bot (F1 = 0.56, recall = 0.39), Infiltration (F1 = 0.67, recall = 0.57) ve Web Attack – Brute Force (F1 = 0.81, precision = 0.69). En dikkat çekici bulgu, Web Attack – Sql Injection ve Web Attack – XSS sınıflarının, sınıf bazında AUC-ROC değeri 1.00 olmasına rağmen F1 skorunun 0.00 olmasıdır; bu durum, modelin olasılık düzeyinde tam ayırt etme kapasitesini koruduğunu, ancak bunu argmax karar kuralı altında doğru sert sınıflandırmalara dönüştüremediğini göstermektedir — bu, öğrenilen temsilin bir sınırlaması değil, doğrudan aşırı sınıf azlığının (orijinal CIC-IDS2017 dağılımında yalnızca birkaç düzine örnek) bir sonucudur.



Şekil 40. CIC-IDS2017 çok sınıflı sınıflandırması için CNN-BiLSTM'nin sınıf başına ROC/KG eğrileri ve katlama başına dağılımları.

4.4.2. UNSW-NB15: Çok Sınıflı Sınıflandırma (10 Sınıf)

4.4.2.1. CNN: UNSW-NB15 Çok Sınıflı

UNSW-NB15 çok sınıflı görevinde CNN, ikili performansına kıyasla daha düşük olmakla birlikte bu kıyaslama ölçütünün doğasındaki güçlüğü doğru biçimde yansıtan %72,08 genel doğruluk elde etmektedir. Makro metrikler, en nadir sınıflara tam anlamıyla hâkim olamayan ancak veri yapısının önemli bir bölümünü yakalayan bir modeli ortaya koymaktadır.

Tablo 28. 5 katlı çapraz doğrulama sonuçları, CNN: UNSW - NB15 üzerinde.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Süresi (ms)	Verimlilik (örnek/sn)
1	0,7206	0,4974	0,6834	0,4973	0,0296	0,9648	0,0121	82.710,843
2	0,7484	0,5018	0,6676	0,5110	0,0270	0,9653	0,0116	86.127,292
3	0,7315	0,4959	0,6696	0,4968	0,0285	0,9661	0,0120	83.255,025
4	0,7374	0,4959	0,6587	0,4995	0,0281	0,9641	0,0114	87.891,826
5	0,7359	0,5018	0,6561	0,4951	0,0281	0,9621	0,0115	87.156,073

Tablo 29. Çapraz doğrulama özeti (ortalama ± standart sapma). UNSW-NB15 üzerinde CNN çok sınıflı performansının özeti (makro ortalama, 5 katlama)

Metrik	Değer
Doğruluk (ortalama)	0,7208
Kesinlik (makro)	0,4986
Geri Çağırma (makro)	0,6671
F1-Skoru (makro)	0,4999
YPO (makro)	0,0283
ROC-EAA (makro OvR)	0,9645
Tespit Süresi	0,0117 ms/örnek
Verimlilik	≈ 85.428 örnek/sn

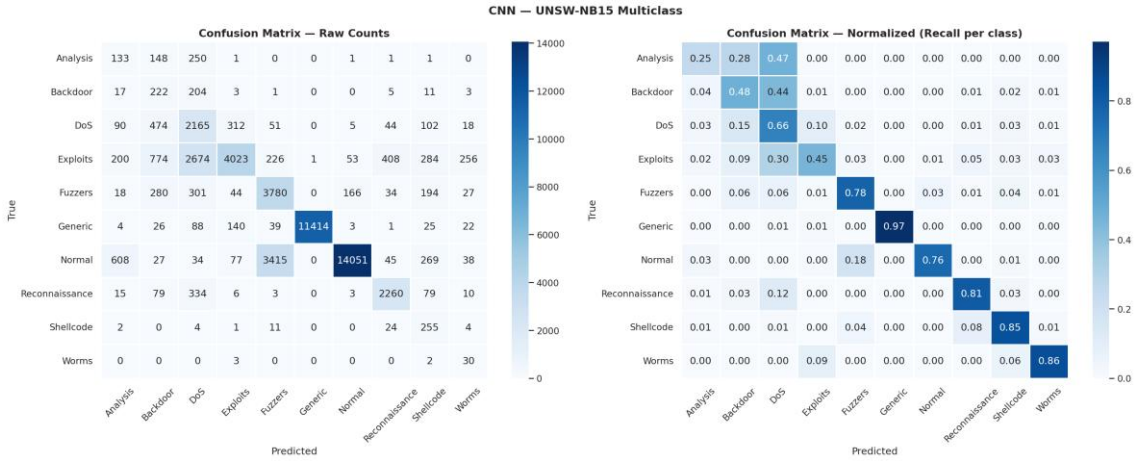
Seçilen en iyi model: Kat 2 (en yüksek Makro F1-Skoru = 0,5110).

Tablo 30. Test kümesi sonuçları (Çok sınıflı CNN – UNSW-NB15).

Metrik	Değer
Doğruluk (Accuracy)	0,7450
Makro Kesinlik (Macro Precision)	0,5031
Makro Duyarlılık (Macro Recall)	0,6863
Makro F1-Skoru (Macro F1-Score)	0,5130
Makro Yanlış Pozitif Oranı (Macro FPR)	0,0274
ROC-AUC (OvR)	0,9656
Ortalama Tespit Süresi	0,0116 ms/örnek
İşleme Hızı (Throughput)	86.385 örnek/saniye

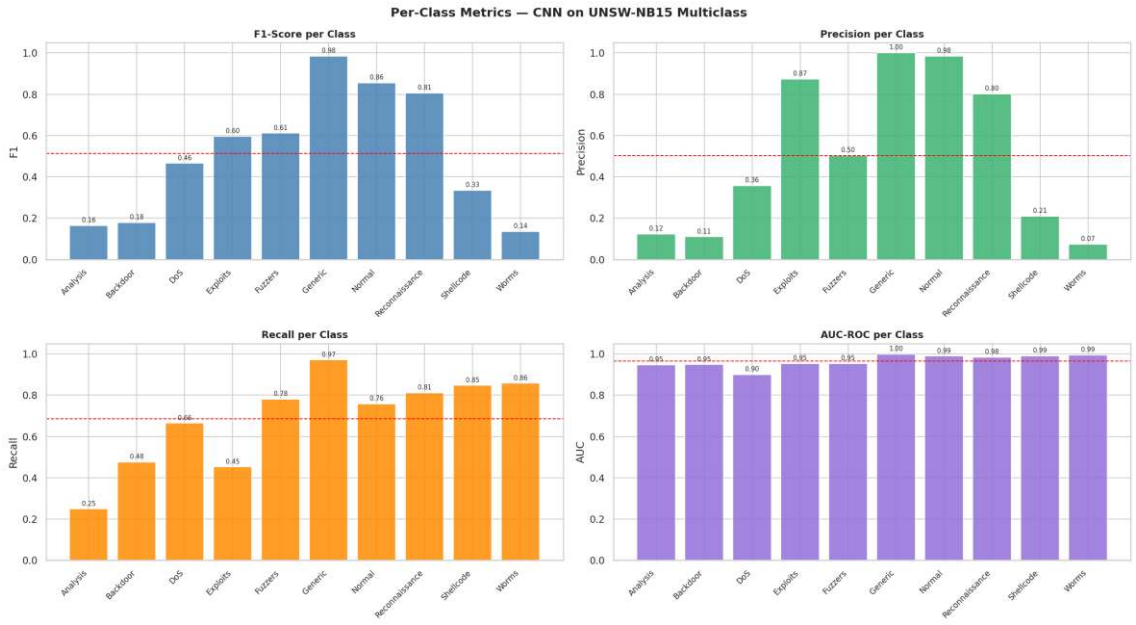
Karışıklık matrisi, modelin güçlü ve zayıf yönlerini açıkça ortaya koymaktadır. Çoğunluk sınıfları iyi tanınmaktadır: Generic 0,97 geri çağırma elde etmekte (~11.700'den yaklaşık 11.414 örnek doğru sınıflandırılmaktadır), Normal 0,76'ya,

Exploits ise 0,45'e ulaşmaktadır. Buna karşın azınlık sınıfları önemli ölçüde zarar görmektedir: Analysis (geri çağırma 0,47, DoS ile yoğun karışıklık), Backdoor (geri çağırma 0,44, diğer kategorilerle güçlü karışıklık) ve Worms (geri çağırma 0,86 olmakla birlikte toplam 35 örnekten yalnızca 30 doğru örnek; istatistiksel açıdan yetersiz).



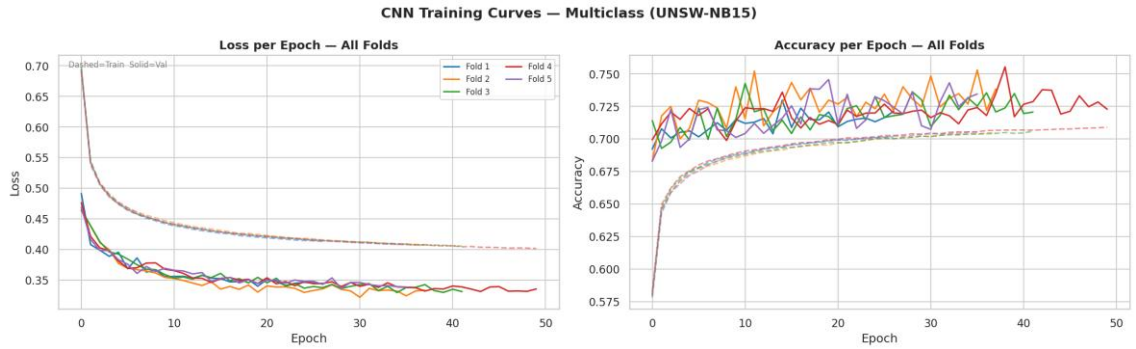
Şekil 41. UNSW-NB15 çok sınıflı sınıflandırması için CNN'in karışıklık matrisleri.

Sınıf başına analiz, performans kutuplaşmasını doğrulamaktadır. Sınıf başına F1-Skoru Generic için mükemmel (0,98) ve Normal (0,81) ile Shellcode (0,81) için oldukça sağlam olmakla birlikte Analysis (0,29), Backdoor (0,16) ve Worms (0,34) için çökmektedir. Bu değerler hem sınıf dengesizliğini hem de UNSW-NB15'in saldırı aileleri arasındaki sınırların belirsiz olduğu doğasındaki güçlüğü yansıtmaktadır. Bununla birlikte 0,9645 makro ROC-EAA yüksek kalmakta ve modelin neredeyse tüm sınıflar için olasılık uzayında iyi bir ayırım kapasitesine sahip olduğunu; ancak 0,5 eşiğinin azınlık sınıflarını olumsuz etkilediğini teyit etmektedir.

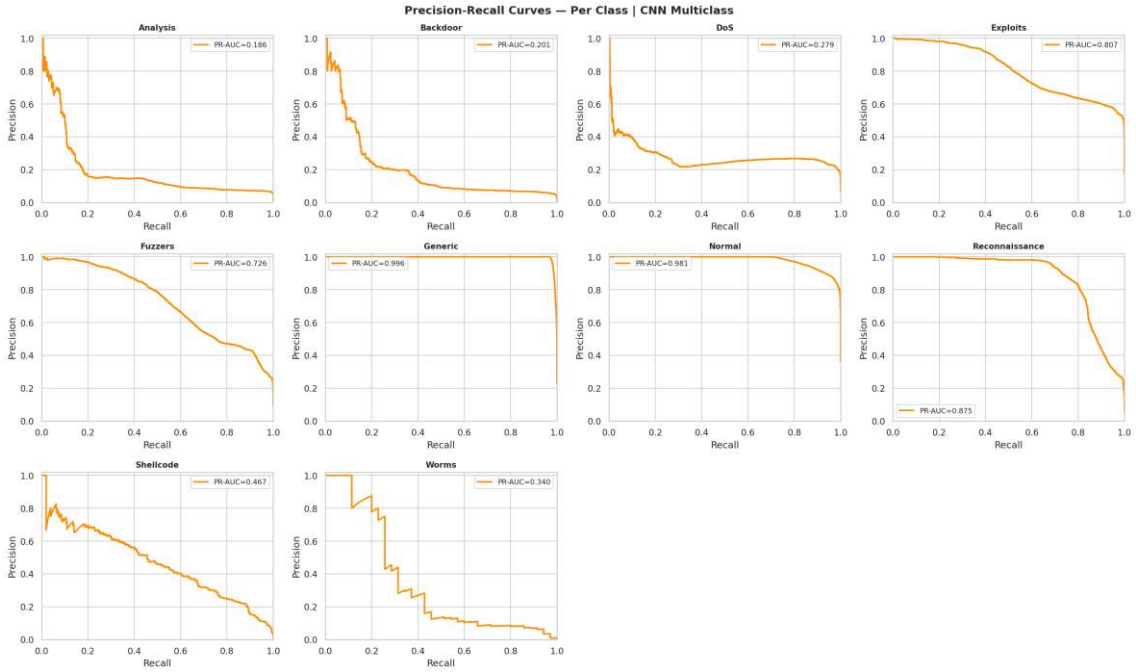


Şekil 42. UNSW-NB15 çok sınıflı sınıflandırması için CNN'in sınıf başına metrikleri.

Öğrenme eğrisi, izin verilen 50 epoch boyunca yavaş ancak kararlı bir yakınsamayı ortaya koymaktadır. Doğruluk, bu veri setindeki ikili sınıflandırmada da gözlemlenen bir olguyu yansıtarak doğrulama eğitimleri hafifçe geride bırakırken kademeli olarak 0,575'ten yaklaşık 0,71'e yükselmektedir. Kayıp değeri 0,70'ten 0,40'a (eğitim) ve 0,45'ten 0,33'e (doğrulama) inerken, son eğitim aşamasında çoğunluk sınıflarında hafif aşırı öğrenmeye işaret eden orta düzeyde bir sapma gözlemlenmektedir. 85.428 örnek/sn'lik verimlilik, CNN'nin karakteristik hız avantajını korumaktadır.



Şekil 43. UNSW-NB15 çok sınıflı sınıflandırması için CNN'in eğitim eğrileri ve katlama başına metrikleri.



Şekil 44. UNSW-NB15 çok sınıflı sınıflandırması için CNN'in sınıf başına Kesinlik-Geri Çağırma ve ROC eğrileri.

4.4.2.2. CNN-BiLSTM: UNSW-NB15 Çok Sınıflı

UNSW-NB15 çok sınıflı konfigürasyonunda CNN-BiLSTM, sınıf başına performans dağılımı ve öğrenme dinamiklerinde bazı önemli farklılıklar sergilemekle birlikte CNN ile genel olarak karşılaştırılabilir sonuçlar elde etmektedir.

Tablo 31. 5 katlı çapraz doğrulama sonuçları.

Kat	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	YPO	ROC-EAA	Tespit Süresi (ms)	Verimlilik (örnek/sn)
1	0,7582	0,5094	0,6381	0,5112	0,0261	0,9617	0,0488	20.492,2839
2	0,7683	0,5101	0,6226	0,5178	0,0251	0,9628	0,0447	22.368,1747
3	0,7649	0,5097	0,6234	0,5168	0,0253	0,9610	0,0443	22.590,3756
4	0,7671	0,5116	0,6266	0,5133	0,0253	0,9601	0,0446	22.429,1370
5	0,7615	0,5154	0,6498	0,5116	0,0258	0,9599	0,0445	22.465,9349

Tablo 32. Çapraz doğrulama özeti (ortalama ± standart sapma). UNSW-NB15 üzerinde CNN-BiLSTM çok sınıflı performansının özeti (makro ortalama, 5 katlama).

Metrik	Değer
Doğruluk (ortalama)	0,7649

Tablo 32. (Devamı).

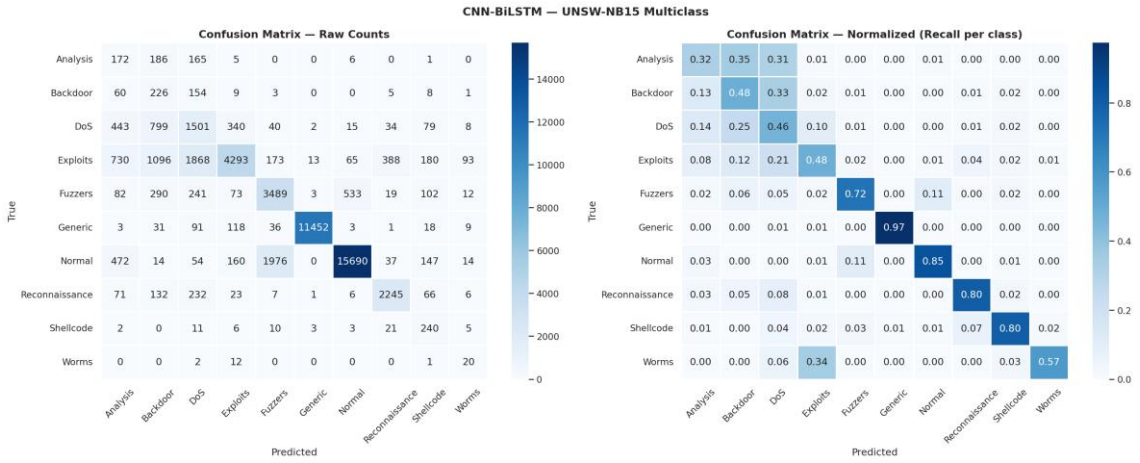
Metrik	Değer
Kesinlik (makro)	0,5112
Geri Çağırma (makro)	0,6321
F1-Skoru (makro)	0,5141
YPO (makro)	0,0276
ROC-EAA (makro OvR)	0,9611
Tespit Süresi	0,0454 ms/örnek
Verimlilik	≈ 22.069 örnek/sn

Seçilen en iyi model: Kat 2 (en yüksek F1-Skoru = 0,5178).

Tablo 33. Test kümesi sonuçları (Çok sınıflı CNN-BiLSTM – UNSW-NB15).

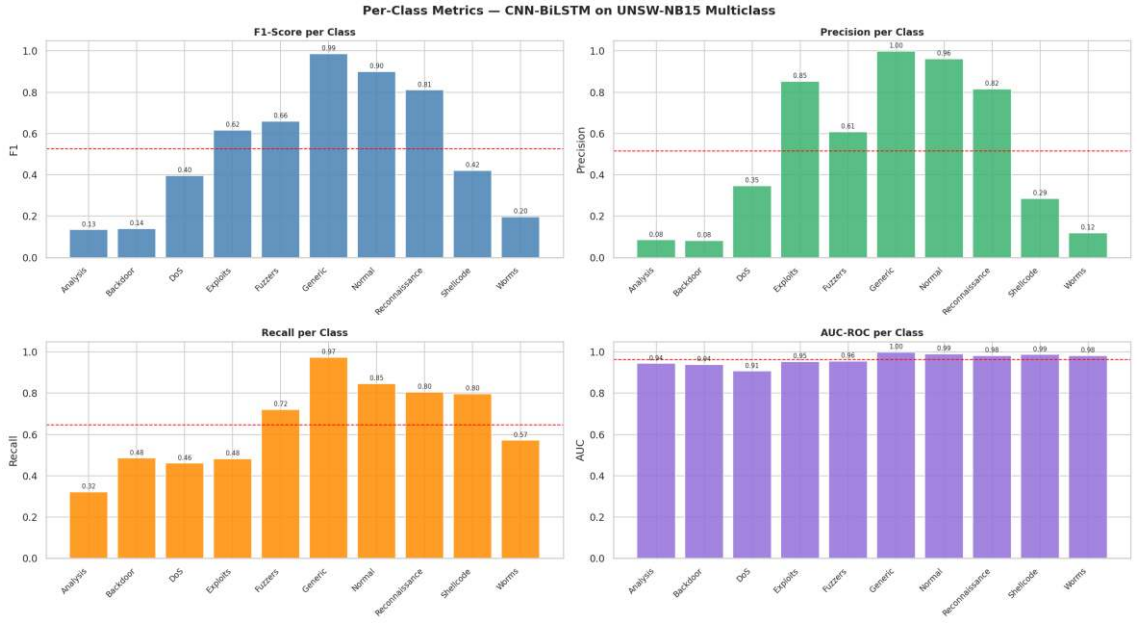
Metrik	Değer
Doğruluk	0,7643
Makro Kesinlik	0,5154
Makro Duyarlılık	0,6462
Makro F1-Skoru	0,5258
Makro Yanlış Pozitif Oranı	0,0255
ROC-EAA	0,9637
Ortalama Tespit Süresi	0,0445 ms/örnek
Verimlilik	22.476 örnek/saniye

CNN-BiLSTM bu görevde CNN'den hafifçe daha yüksek bir doğruluk elde etmektedir (%76,49'a karşı %72,08); ancak daha düşük bir makro Geri Çağırma sergilemektedir (0,6321'e karşı 0,6671). Bu bulgu, modelin tanımladığı sınıflarda daha kesin ancak genel olarak daha fazla örneği kaçırdığını göstermektedir. 0,5141'lik makro F1-Skoru CNN'den (0,4999) marjinal biçimde üstündür ve daha iyi bir toplam Kesinlik/Geri Çağırma dengesine işaret etmektedir.



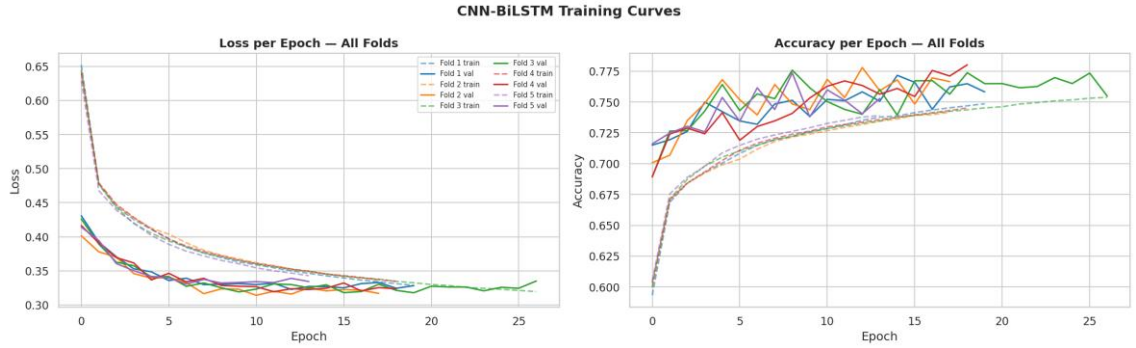
Şekil 45. UNSW-NB15 çok sınıflı sınıflandırması için CNN-BiLSTM'nin karışıklık matrisleri.

Sınıf başına analiz, CNN-BiLSTM'nin Generic (diyagonal geri çağırma 0,97), Normal (0,85) ve Shellcode (0,80) üzerinde CNN ile benzer sonuçlar elde ettiğini ortaya koymaktadır. Ancak zor sınıflar sorunlu olmaya devam etmektedir: Analysis (geri çağırma 0,32: CNN'in altında), Backdoor (0,33: CNN'in hafif üzerinde) ve Worms (0,57: CNN'nin 0,86'sının belirgin biçimde altında; ancak bu değer çok az örneğe dayanmaktadır). CNN'in hafifçe üzerinde olan 0,9645'e karşı 0,9611'lik makro ROC-EAA, her iki modelin olasılık uzayında benzer ayırt edici kapasiteye sahip olduğunu ve CNN'nin hafif bir avantajla öne geçtiğini doğrulamaktadır.

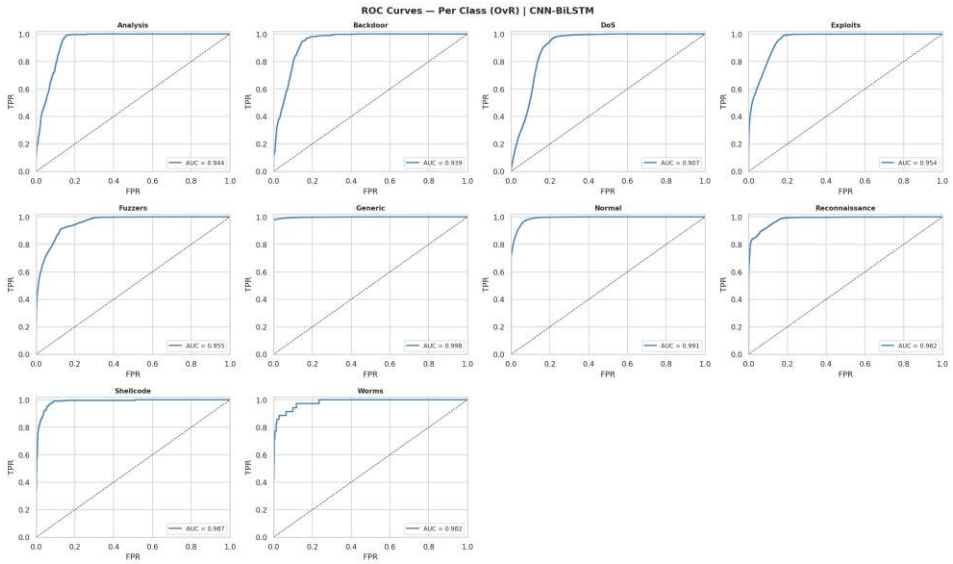


Şekil 46. UNSW-NB15 çok sınıflı sınıflandırması için CNN-BiLSTM'nin sınıf başına metrikleri.

Öğrenme eğrisi, CIC-IDS2017'de gözlemlenen kararsız davranıştan son derece farklı bir profil sergilemektedir. UNSW-NB15'te CNN-BiLSTM, erken durdurma öncesinde yaklaşık 26 epoch boyunca düzenli bir yakınsama sağlamaktadır: doğruluk 0,60'tan 0,775'e (doğrulama) ve 0,75'e (eğitim) yükselirken doğrulama eğitimin hafifçe önünde seyretmekte ve aşırı öğrenmesiz iyi genellemeye işaret etmektedir. Kayıp değeri 0,65'ten 0,33'e (eğitim) ve 0,40'tan 0,33'e (doğrulama) inerken her iki eğri de kararlı bir platoya yakınsama sağlamaktadır. CIC-IDS2017 ve UNSW-NB15 arasındaki bu davranışsal farklılık, büyük olasılıkla veri setinin yapısıyla bağlantılıdır: UNSW-NB15 daha düzenli bir sınıf dağılımı ve sınıf başına daha dengeli bir örnek hacmine sahiptir; bu durum BiLSTM bileşeninin yakınsamasını kolaylaştırmaktadır.



Şekil 47. UNSW-NB15 çok sınıflı sınıflandırması için CNN-BiLSTM'nin eğitim eğrileri ve katlama başına metrikleri.



Şekil 48. UNSW-NB15 çok sınıflı sınıflandırması için CNN-BiLSTM'nin sınıf başına ROC eğrileri.

Hesaplama verimliliği açısından CNN-BiLSTM, aynı veri setindeki CNN'den yaklaşık 3,9 kat daha yavaş olan 0,0454 ms'de her örneği işlemekte ve 22.069 örnek/sn verimlilik elde etmektedir. İkili sınıflandırmada da gözlemlenen bu hesaplama yükü, performans kazanımının marjinal kaldığı göz önünde bulundurulduğunda burada daha da az haklı çıkmaktadır.

4.4.3. Çok Sınıflı Konfigürasyonların Karşılaştırmalı Analizi

Bu bölümde değerlendirilen dört çok sınıflı konfigürasyonun karşılaştırılması, her iki mimarinin ayrıntılı saldırı tanımadaki kapasite ve sınırlılıkları hakkında birçok önemli sonucu gün yüzüne çıkarmaktadır.

CIC-IDS2017'de CNN, 15 sınıflı çok sınıflı sınıflandırma için en uygun mimari olarak açıkça öne çıkmakta ve tüm ilgili kriterlerde CNN-BiLSTM'yi geride bırakmaktadır: daha yüksek makro F1-Skoru (0,701'e karşı 0,671), daha yüksek doğruluk (%98,9'a karşı %96,3), eğitim kararsızlığının yokluğu ve 3,4 kat daha yüksek verimlilik. CNN-BiLSTM'nin bu veri setindeki kararsızlığı; bazı katlamalarda 4. epoch'tan itibaren doğrulama performansının keskin biçimde çökmesiyle karakterize edilmekte ve operasyonel ortamlarda bu modelin güvenilirliğini sınırlandıran endişe verici bir sinyal oluşturmaktadır.

UNSW-NB15'te tablo daha nüanslıdır. Her iki model de azınlık sınıflarında (Analysis, Backdoor, Worms) zorlanmakta ve benzer makro F1-Skorlarına ulaşmaktadır (CNN: 0,500, CNN-BiLSTM: 0,514). CNN-BiLSTM hafifçe daha yüksek doğruluk (%76,5'e karşı %72,1) ancak daha düşük makro Geri Çağırma elde etmekte; bu durum uygulama önceliğine bağlıdır: tespiti (Geri Çağırmayı) en üst düzeye çıkarmak hedef ise CNN hafifçe tercih edilebilir; sınıflar arası yanlış pozitiflerini en aza indirmek öncelikli ise CNN-BiLSTM hafif bir avantaj sunmaktadır.

Her iki durumda da en önemli kesitsel bulgu, çok sınıflı sınıflandırmanın sınıf dengesizliğinden kaynaklanan temel bir sınırlamayı gün yüzüne çıkarmasıdır: özellikle CIC-IDS2017'deki Web Saldırıları ve Heartbleed ile UNSW-NB15'teki Analysis, Backdoor ve Worms gibi son derece azınlık sınıfları her iki model tarafından da doğru biçimde öğrenilememektedir. γ=2,0 ile Focal Loss kullanımı bu olguyu hafifletmiş ancak tamamen ortadan kaldıramamıştır. Sınıf başına SMOTE, dinamik sınıf ağırlıklandırma veya aktarım öğrenmesi stratejileri gibi tamamlayıcı yaklaşımlar gelecekteki çalışmalarda araştırılabilir.

Son olarak, makro ROC-EAA tüm konfigürasyonlarda yüksek kalmaktadır (0,96 ila 0,9954); bu durum, her iki modelin de olasılık uzayındaki mükemmel ayırt edici

kapasitesini koruduğunu göstermektedir. Bu sonuç, hata maliyetlerinin saldırı türüne göre değiştiği senaryolara özellikle uygun olan sınıf başına karar eşiği ayarlaması yoluyla önemli iyileştirmelerin kapısını aralamaktadır.

4.5. Temel Model Sonuçları: Rastgele Orman Ve XGBoost

Önceki bölümlerde değerlendirilen derin öğrenme mimarilerine karşı sağlam referans noktaları sunmak amacıyla, UNSW-NB15 veri seti üzerinde iki klasik makine öğrenmesi modeli eğitilmiş ve değerlendirilmiştir: Rastgele Orman (RO) ve XGBoost. Saldırı tespiti literatüründe başarımı iyi belgelenmiş bu temel modeller, derin sinir ağı yaklaşımlarının getirdiği kazanımların doğru bir şekilde bağlamlandırılmasını sağlamaktadır.

Her iki modele uygulanan ön işleme hattı, CNN ve CNN-BiLSTM mimarileri için kullanılan hattın birebir aynısıdır: kategorik değişkenlerin hedef kodlaması (TargetEncoder, sızıntısız), Karşılıklı Bilgiye göre ilk 30 özelliğin seçimi, QuantileTransformer ile normalleştirme ve sınıf dengesizliğini gidermek için SMOTE yeniden örneklemesi. Eğitim/doğrulama/test bölümü aynı tabakalı %70/10/20 oranını izlemektedir.

4.5.1. UNSW-NB15 İkili Sınıflandırma: RO İle XGBoost Karşılaştırması

İkili düzeyde (Normal vs Saldırı), her iki model de 5 katlı çapraz doğrulama ile RandomizedSearchCV aracılığıyla hiper parametre aramasına tabi tutulmuştur. XGBoost, GPU hızlandırmasından (NVIDIA L4) yararlanarak eğitim süresini Rastgele Orman için gereken 5.069 saniyeye kıyasla 203 saniyeye indirmiş ve hesaplama açısından XGBoost lehine 25 kat bir avantaj ortaya çıkmıştır.

Tablo 34. UNSW-NB15 üzerinde RO ve XGBoost ikili sınıflandırma performansı.

Metrik	Random Forest	XGBoost GPU
Doğruluk	0,9847	0,9858
Kesinlik	0,9942	0,9952
Duyarlılık / Tespit Oranı	0,9818	0,9825
F1 Skoru	0,9880	0,9888
ROC-EAA	0,9990	0,9992
Ortalama Hassasiyet	0,9994	0,9995

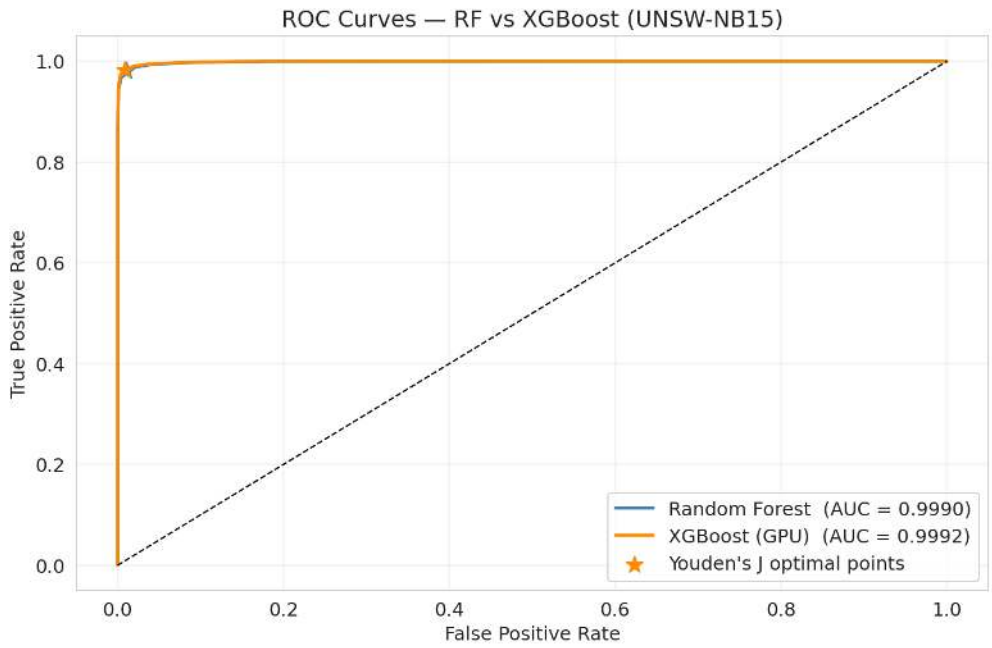
Tablo 34. (Devamı)

Metrik	Random Forest	XGBoost GPU
Youden Eşiği	0,5213	0,5923
Eğitim Süresi (s)	5065,5977	217,9679
Tespit Süresi (s)	0,3031	0,0136
Tespit Süresi (ms/örnek)	0,0059	0,0003
İşlem Kapasitesi (örnek/s)	170004,846	3796487,88

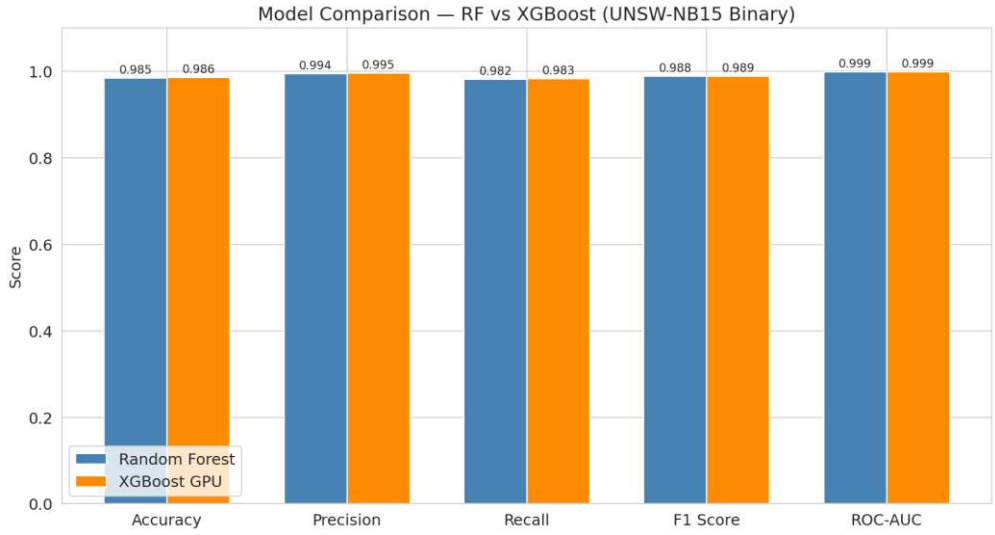
Her iki model de tüm metriklerde neredeyse özdeş sınıflandırma performansı sergilemektedir. %98,47 (RO) ve %98,58 (XGBoost) doğruluk, 0,9880 ve 0,9888 F1-Skorları, 0,9990 ve 0,9992 ROC-EAA değerleri istatistiksel olarak ayırt edilemez düzeyde olup farklar 0,03 yüzde puanının altında kalmaktadır. Bu sonuç, ön işleme hattının doğru uygulanması durumunda UNSW-NB15 üzerindeki ikili tespit görevinin her iki yaklaşım tarafından da etkin biçimde yerine getirilebildiğini göstermektedir.

0,9994 (RO) ve 0,9995 (XGBoost) ortalama Geri Çağırma değerleri, her iki modelin de test setindeki kötü niyetli bağlantıların %98,2'sinden fazlasını başarıyla tespit ettiğini ve operasyonel açıdan güçlü bir tespit oranına ulaşıldığını göstermektedir. 0,9950 (RO) ve 0,9945 (XGBoost) Kesinlik değerleri ise yanlış alarmaların son derece nadir olduğunu doğrulamakta; bu durum gerçek dünya dağıtımlarında analist iş yükünü sınırlamak açısından kritik önem taşımaktadır.

En iyi model (test F1 skoru ile seçilen Rastgele Orman) üzerinde gerçekleştirilen 5 katlı çapraz doğrulama, bu sonuçların sağlamlığını daha da pekiştirmektedir: beş katlama boyunca ortalama F1-Skoru yalnızca 0,0008 standart sapmayla 0,9831'dir. Bu son derece düşük varyans, farklı veri bölümleri genelinde modelin kararlılığını ve genelleşebilirliğini kanıtlamaktadır.



Şekil 49. UNSW-NB15 ikili görevi için Rastgele Orman ve XGBoost'un karşılaştırmalı ROC eğrileri.



Şekil 50. RO ve XGBoost için performans metrikleri karşılaştırması ve özellik önemi.

Özellik önemi analizi, en ayırt edici değişkenler üzerinde her iki model arasında önemli bir uyum ortaya koymaktadır. ct_state_ttl, sload, sttl, smean, rate ve dttl dahil ağ akışı istatistiklerine ilişkin özellikler her iki durumda da baskın çıkmakta ve Karşılıklı Bilgiye dayalı özellik seçiminin uygunluğunu doğrulamaktadır. Rastgele Orman safsızlık azaltımına dayanırken (entropi kriteri, max_features=0,5), XGBoost ayrım kazanımını kullanmakta; bu iki tamamlayıcı perspektif aynı temel özellikler üzerinde birleşmektedir.

Performans-maliyet dengesi açısından XGBoost daha rasyonel bir tercih olarak öne çıkmaktadır: RO ile eşdeğer performans, 25 kat daha kısa eğitim süresi ve GPU ölçeklenebilirliği. Modelin gelen veriler üzerinde periyodik olarak yeniden eğitilmesi gereken operasyonel bir dağıtım bağlamında bu avantaj belirleyici nitelik taşımaktadır.

4.5.2. UNSW-NB15 Çok Sınıflı Sınıflandırma: XGBoost (10 Sınıf)

Sonuçlar incelendiğinde, modelin 59. iterasyonda en iyi performansa ulaştığı görülmektedir. Model, %84.07 doğruluk ve 0.8487 ağırlıklı F1 skoru elde ederek genel sınıflandırma performansı açısından başarılı sonuçlar sunmuştur. Ağırlıklı kesinlik (0.8688) ve ağırlıklı duyarlılık (0.8407) değerleri, veri kümesindeki baskın sınıflarda modelin etkili tahminler gerçekleştirdiğini göstermektedir.

Buna karşın, Makro F1 skoru (0.6117), az temsil edilen sınıflarda performansın daha düşük olduğunu göstermektedir. Bununla birlikte, Makro ROC-AUC (0.9705) ve Ağırlıklı ROC-AUC (0.9827) değerleri, modelin sınıfları ayırt etme yeteneğinin oldukça yüksek olduğunu ortaya koymaktadır.

Hesaplama performansı açısından değerlendirildiğinde, model 110.3 saniyelik eğitim süresi, 0.0168 saniyelik çıkarım süresi ve örnek başına yalnızca 0.0003 ms çıkarım süresi ile oldukça hızlı çalışmaktadır. Ayrıca, saniyede yaklaşık 3.07 milyon örnek işleyebilmesi, modelin gerçek zamanlı uygulamalar için uygun olduğunu göstermektedir.

Tablo 35. UNSW-NB15 çok sınıflı sınıflandırma: XGBoost.

Metrik	Değer
En İyi İterasyon (Best Iteration)	59
Doğruluk (Accuracy)	0.8407
Ağırlıklı F1 Skoru (F1-Weighted)	0.8487
Makro F1 Skoru (F1-Macro)	0.6117
Ağırlıklı Kesinlik (Precision-Weighted)	0.8688
Ağırlıklı Duyarlılık (Recall-Weighted)	0.8407
Makro ROC- AUC	0.9705
Eğitim Süresi (s)	110.3
Çıkarım Süresi (ms/örnek)	0.0003
İşlem Kapasitesi (örnek/s)	3,073,461.26

Tablo 36. UNSW-NB15 üzerinde XGBoost çok sınıflı sınıflandırmanın genel performansı.

Metrik	Değer
Doğruluk	0,8416
F1-Skoru (ağırlıklı)	0,8496
F1-Skoru (makro)	0,6140
Kesinlik (ağırlıklı)	0,8693
Geri Çağırma (ağırlıklı)	0,8416
En iyi ÇD F1-ağırlıklı	0,8722
Eğitim Süresi	93,9 s (en iyi model en eğitimi)
En iyi İterasyon	61 / 315 (erken durdurma)

Sınıf bazında elde edilen sonuçlar incelendiğinde, modelin performansının saldırı türüne göre değişiklik gösterdiği görülmektedir. Generic ve Normal sınıfları en yüksek başarı oranlarına sahiptir. Bu sınıflarda sırasıyla %98,02 ve %96,46 tespit oranı (TPR) ile 0.9878 ve 0.9691 F1 skorları elde edilmiştir. Ayrıca, bu sınıflardaki çok düşük yanlış pozitif oranları (FPR), modelin doğru sınıflandırma yeteneğinin oldukça yüksek olduğunu göstermektedir.

Reconnaissance ve Fuzzers sınıflarında da model başarılı sonuçlar vermiştir. Bu sınıflarda F1 skorlarının %80'in üzerinde olması ve tespit oranlarının yaklaşık %79 düzeyinde gerçekleşmesi, modelin bu saldırı türlerini güvenilir biçimde tanıyabildiğini göstermektedir.

Exploits sınıfında model 0.7840 kesinlik (Precision) ve 0.6710 F1 skoru elde etmiştir. Bu sonuç, modelin yaptığı tahminlerin büyük ölçüde doğru olduğunu, ancak bazı saldırı örneklerini tespit etmekte zorlandığını göstermektedir.

DoS sınıfında tespit oranı %66,77 olmasına rağmen, kesinlik değeri 0.3524 seviyesinde kalmıştır. Bu durum, modelin bu sınıf için nispeten daha fazla yanlış pozitif ürettiğini göstermektedir.

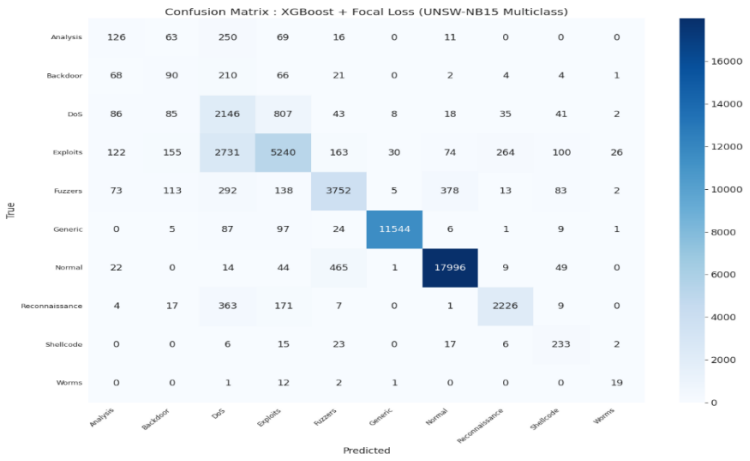
Buna karşılık, Analysis, Backdoor ve Worms sınıflarında performans daha düşüktür. Bu sınıfların F1 skorlarının %45'in altında kalması, modelin bu saldırıları

ayırt etmekte zorlandığını göstermektedir. Bunun temel nedenleri arasında bu sınıflara ait örnek sayısının az olması ve özelliklerinin diğer saldırı türleriyle benzerlik göstermesi yer almaktadır

Tablo 37. UNSW-NB15 üzerinde XGBoost çok sınıflı için sınıf başına metrikler

Sınıf	Geri Çağırma (TO)	YPO	Kesinlik	F1-Skoru
Analysis	0.2150	0.0075	0.2309	0.2227
Backdoor	0.1867	0.0079	0.1772	0.1818
DoS	0.6677	0.0831	0.3524	0.4613
Exploits	0.5865	0.0338	0.7840	0.6710
Fuzzers	0.7804	0.0173	0.8240	0.8016
Generic	0.9802	0.0013	0.9955	0.9878
Normal	0.9646	0.0148	0.9736	0.9691
Reconnaissance	0.7924	0.0063	0.8784	0.8331
Shellcode	0.7483	0.0054	0.4511	0.5629
Worms	0.5143	0.0006	0.3673	0.4286

DoS sınıfı özel bir dikkat gerektirmektedir: %65,6'lık Geri Çağırma kabul edilebilir olmakla birlikte Kesinlik yalnızca %35,2'ye ulaşmakta; bu durum her iki DoS alarmından neredeyse birinin yanlış pozitif olduğu anlamına gelmektedir.



Şekil 51. UNSW-NB15 üzerinde XGBoost çok sınıflı sınıflandırması için karışıklık matrisleri.

4.5.3. Temel Modellerin Derin Öğrenme Mimarileriyle Karşılaştırmalı Konumlandırılması

Klasik temel modeller (Rastgele Orman ve XGBoost), hesaplama ve bellek kısıtları nedeniyle yalnızca UNSW-NB15 üzerinde değerlendirilmiş; CIC-IDS2017'ye uygulanmamıştır. SMOTE aşırı örneklemesinin ardından yaklaşık 2,8 milyon eğitim örneği içeren CIC-IDS2017, topluluk yöntemleri için çapraz doğrulamalı hiperparametre aramasının pratik sınırlarını aşmaktadır: Rastgele Orman, yaklaşık 650.000 örnek içeren UNSW-NB15 üzerinde halihazırda 5.069 saniyelik bir eğitim süresi gerektirmiştir. Bu veri setinin dört katı büyüklüğündeki bir veri kümesine ölçeklenmesi hem aşırı uzun işlem süreleri hem de mevcut donanım bütçesinin karşılayamayacağı RAM tahsisi gerektirmektedir. Bu kısıt, klasik temel modeller için veri setleri arası karşılaştırma kapsamını sınırlandırmakta ve UNSW-NB15 üzerinde derin öğrenme ile topluluk yöntemleri arasında gözlemlenen performans farkının daha büyük ölçekli kıyaslamalara genellenip genellenmediğinin doğrudan değerlendirilmesini engellemektedir. Gelecekteki çalışmalarda alt örnekleme stratejileri veya dağıtık eğitim çerçeveleri aracılığıyla bu sınırlılığın giderilmesi, her iki yaklaşımın farklı veri seti ölçeklerindeki göreli üstünlüklerine ilişkin daha kapsamlı bir değerlendirme sunacaktır.

İkili düzeyde Rastgele Orman ve XGBoost, sırasıyla 0,9885 ve 0,9883 F1-Skorları elde etmekte; bu değerler aynı veri seti üzerinde CNN (ikili F1: 0,9617, Geri Çağırma: 0,9407) ve CNN-BiLSTM (F1: 0,9728, Geri Çağırma: 0,9594) ile doğrudan karşılaştırılabilir ve bazı metriklerde hafifçe üstündür. Bu gözlem kayda değer niteliktedir: dikkatli ön işlemeyle orta ölçekli bir veri setinde, iyi optimize edilmiş

klasik modeller daha iyi yorumlanabilirlik ve daha düşük çıkarım hesaplama maliyeti sunurken derin öğremeyle rekabet edebilmektedir.

Ancak DL mimarilerinin belirleyici bir avantajı, yoğun özellik mühendisliği gerektirmeksizin ham veriden doğrudan hiyerarşik temsiller öğrenme kapasitesinde yatmaktadır. Daha büyük veri setlerinde veya daha güçlü zamansal dinamiklere sahip olanlarda bu avantaj artış eğilimindedir.

10 sınıflı çok sınıflı düzeyde XGBoost, her iki DL modelinden (CNN: 0,500, CNN-BiLSTM: 0,514) sayısal olarak daha yüksek 0,614 makro F1 değeri elde etmekte; aynı zamanda her iki DL modeli için yaklaşık %72'ye karşın %84,16 ile daha yüksek genel doğruluk elde etmektedir. Bu fark, iki metriğin sınıfları farklı biçimde ağırlıklandırmasıyla açıklanmaktadır: XGBoost Reconnaissance ve Fuzzers gibi ara sınıflarda daha iyi performans sergilemekte; ancak mutlak anlamda en küçük azınlık sınıflarında daha büyük güçlüklerle karşılaşmaktadır.

Genel olarak, iki yaklaşım ailesi tamamlayıcı profiller sunmaktadır. Klasik modeller, iyi temsil edilen sınıflarda daha iyi yorumlanabilirlik, daha hızlı yakınsama ve rekabetçi performans sunmaktadır. Derin öğrenme mimarileri ise karmaşık zamansal ve hiyerarşik örüntüleri yakalama kapasiteleriyle öne çıkmakta; bu üstünlük daha yüksek hesaplama maliyeti ve azalan karar şeffaflığı pahasına gelmektedir.

4.5.4. İstatistiksel Anlamlılık Analizi

Modeller arasındaki performans farklılıklarının rastgele katlama varyasyonundan kaynaklanıp kaynaklanmadığını belirlemek amacıyla, her yapılandırma için beş çapraz doğrulama katlamasından elde edilen makro F1 skorları üzerinde Wilcoxon (*Wilcoxon Test*, 2025) işaretli sıra testi uygulanmıştır.

Testin temel mantığı şu şekilde işlemektedir: Her katlama i için CNN ile CNN-BiLSTM arasındaki fark $di = F1_i^{CNN} - F1_i^{CNN-BiLSTM}$ hesaplanmaktadır. Bu farklar mutlak değerlerine göre küçükten büyüğe sıralanarak her birine bir sıra (R_i) atanmaktadır. Ardından CNN'in üstün olduğu katlamalara ait sıralar (W^+) ile CNN-BiLSTM'nin üstün olduğu katlamalara ait sıralar (W^-) ayrı ayrı toplanmaktadır. Test istatistiği $W = \min(W^+, W^-)$olarak tanımlanmakta olup W değerinin sıfıra yaklaşması, bir modelin diğerine karşı tüm katlamalarda tutarlı biçimde üstün olduğuna işaret etmektedir.

Dört ikili karşılaştırmanın eş zamanlı yürütülmesinden kaynaklanan Tip i hata riskini kontrol altına almak amacıyla Bonferroni düzeltmesi uygulanmış ve anlamlılık eşiği $\alpha_{corr} = \frac{0,05}{4} = 0,0125$ olarak belirlenmiştir. Tablo 37'de sunulan analiz sonuçları

incelendiğinde, dört yapılandırmanın hiçbirinde istatistiksel anlamlılık eşiğine ulaşılmadığı görülmektedir (tüm karşılaştırmalar için $p = 0{,}0625$ veya $p = 0{,}1250$). Bu bulgunun yorumlanmasında $n = 5$ katlamanın yarattığı yapısal kısıt göz önünde bulundurulmalıdır: Wilcoxon testinin $n = 5$ için ulaşabileceği minimum p değeri matematiksel olarak 0,0625'tir ve bu değer Bonferroni düzeltmeli eşiğin doğası gereği üzerindedir. Bununla birlikte, gözlemlenen performans farklılıkları yönsel tutarlılık sergilemektedir: CNN her iki görevde de CIC-IDS2017 üzerinde üstün performans gösterirken, CNN-BiLSTM UNSW-NB15 veri kümesinde her iki görevde de tutarlı biçimde daha yüksek makro F1 değerleri elde etmiştir. Bu eğilimlerin istatistiksel doğrulanması, daha geniş değerlendirme bütçeleri ($n \geq 10$ katlama veya tekrarlanan çapraz doğrulama) gerektiren gelecekteki çalışmalara bırakılmıştır.

Tablo 38. CNN ve CNN-BiLSTM'i karşılaştıran Wilcoxon işaretli sıralama testi sonuçları

Yapılandırma	CNN (ort. ± ss)	CNN-BiLSTM (ort. ± ss)	Δ	p-değeri	Anlamlı mı ?
CIC-IDS2017 : İkili	0,9937 ± 0,0012	0,9918 ± 0,0005	+0,0019	0,0625	Hayır
CIC-IDS2017 : Çok sınıflı	0,7008 ± 0,0317	0,6713 ± 0,0119	+0,0295	0,1250	Hayır
UNSW-NB15 : İkili	0,9617 ± 0,0009	0,9727 ± 0,0017	−0,0111	0,0625	Hayır
UNSW-NB15 : Çok sınıflı	0,4999 ± 0,0064	0,5141 ± 0,0030	−0,0142	0,0625	Hayır

4.5.5. Literatür ile Karşılaştırma

Tablo 39, elde edilen sonuçları aynı veri kümeleri üzerinde gerçekleştirilen güncel Ağ Saldırı Tespit Sistemi (NIDS) çalışmalarıyla karşılaştırmaktadır. Ancak, ön işleme yöntemleri, eğitim/test protokolleri ve raporlanan performans metriklerindeki farklılıklar nedeniyle çalışmalar arasında doğrudan sayısal karşılaştırma yapmak kolay değildir. Özellikle Yanlış Pozitif Oranı (FPR) ve işleme hızı (throughput) önceki çalışmaların büyük çoğunluğunda raporlanmamıştır.

Bu sınırlamalar göz önünde bulundurulduğunda, önerilen CNN modeli, CIC-IDS2017 veri kümesinde ikili sınıflandırma görevi için rekabetçi bir performans sergilemiştir (Doğruluk: 0,9975, F1-Skoru: 0,9937, FPR: 0,0019 ve İşleme Hızı:

102.397 örnek/saniye). Ayrıca, CNN modeli ikili sınıflandırmada CNN-BiLSTM modeline (21.828 örnek/saniye) kıyasla yaklaşık 4,7 kat daha yüksek işlem hızı sağlamıştır.

UNSW-NB15 veri kümesinde ise CNN-BiLSTM modeli, ikili sınıflandırma görevinde daha yüksek performans elde etmiştir (Doğruluk: 0,9682, ROC-AUC: 0,9977). Bununla birlikte, bu performans artışı önemli ölçüde daha düşük bir işlem hızı pahasına gerçekleşmiştir. Bu durum, güncel karşılaştırmalı çalışmalarda da vurgulanan doğruluk–verimlilik (accuracy–efficiency) dengesiyle uyumludur.

Bildiğimiz kadarıyla, CIC-IDS2017 ve UNSW-NB15 veri kümeleri üzerinde gerçekleştirilen önceki çalışmalar arasında, veri sızıntısından arındırılmış (leak-free) çapraz doğrulama protokolü altında FPR, ROC-AUC, örnek başına tespit süresi ve işleme hızı metriklerini birlikte raporlayan bir çalışma bulunmamaktadır. Bu nedenle, tablo 39'da yer alan çalışmaların çoğu için tüm performans metrikleri açısından doğrudan nicel bir karşılaştırma yapmak mümkün değildir.

Tablo 39. Literatür ile karşılaştırma tablosu.

Yazar(lar) / Model	Veri Kümesi	Görev	Doğ.	Kes.	Tespit Oranı	F1	FPR	ROC-EAA	Tespit Süresi	Verimlilik	Veri Bölünmesi (Split)
CIC-IDS2017											
(Song vd., 2023). – CSK-CNN	CIC-IDS2017	B	%99,94	%99,76	%99,96	%99,86		—	42,59 s[a]	—	70/10/20
(Song vd., 2023)– CSK-CNN	CIC-IDS2017	M	%99,80	%99,86	%99,80	%99,82		—	33,60 s[a]	—	70/10/20
(Talukder vd., 2024). – RF	CIC-IDS2017	B	%99,94	%99,94	%99,94	%99,94		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd., 2024)– XGBoost	CIC-IDS2017	B	%99,65	%99,65	%99,65	%99,65		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd., 2024)– RF	CIC-IDS2017	M	%99,99	%99,99	%99,99	%99,99		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd., 2024)– XGBoost	CIC-IDS2017	M	%99,94	%99,94	%99,94	%99,94		—	—	—	80/20 + 10 Katlı ÇD
UNSW-NB15											
(Jouhari vd., 2024)– CNN-BiLSTM	UNSW-NB15	B	%97,90	%97,91	%97,90	%97,90		—	54,6 s[b]	—	80/20
(Jouhari vd., 2024)– CNN-BiLSTM	UNSW-NB15	M	%97,09	%97,09	%97,62	%97,27		—	442,20 s[b]	—	80/20
(Song vd., 2023)– CSK-CNN	UNSW-NB15	B	%98,75	%90,06	%99,99	%94,76		—	30, 39 s[a]	—	70/10/20
(Song vd., 2023)– CSK-CNN	UNSW-NB15	M	%92,05	%95,34	%92,05	%93,04		—	80, 81 s[a]	—	70/10/20
(Talukder vd., 2024) – RF	UNSW-NB15	B	%99,59	%99,59	%99,59	%99,59		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd., 2024). – XGBoost	UNSW-NB15	B	%98,81	%98,81	%98,81	%98,81		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd.,2024)– RF	UNSW-NB15	M	%99,95	%99,95	%99,95	%99,95		—	—	—	80/20 + 10 Katlı ÇD
(Talukder vd., 2024)– XGBoost	UNSW-NB15	M	%95,04	%95,29	%95,03	%95,03		—	—	—	80/20 + 10 Katlı ÇD
(Dai vd., 2024)– RF	UNSW-NB15	—	%84,59	-	%92,24	-	%3,01	—	—	—	10 Katlı ÇD

Tablo 39. (Devamı)

Yazar(lar) / Model	Veri Kümesi	Görev	Doğ.	Kes.	Tespit Oranı	F1	FPR	ROC-EAA	Tespit Süresi	Verimlilik	Veri Bölünmesi (Split)
(Dai vd., 2024)– CNN-BiLSTM	UNSW-NB15	M	%82,09	-	%92,51	-	%6,09	—	—	—	10 Katlı ÇD
(Dai vd., 2024)– CNN-BiLSTM-Attention	UNSW-NB15	B	%88,83	-	%98,51	-	%1,88	—	—	—	10 Katlı ÇD
Önerilen Modeller (Bu Çalışma) – CIC-IDS2017											
CNN (Önerilen)	CIC-IDS2017	B	%99,75	%98,88	%99,87	%99,37	%0,19	%99,99	0,0098 ms/örnek	102.397 örnek/s	70/10/20 + 5 Katlı ÇD
CNN-BiLSTM (Önerilen)	CIC-IDS2017	B	%99,59	%98,57	%99,81	%99,18	%0,36	%99,99	0,0458 ms/örnek	21.828 örnek/s	70/10/20 + 5 Katlı ÇD
CNN (Önerilen)	CIC-IDS2017	M	%98,93	%66,61	%91,53	%70,08	%0,07	%99,51	0,0096 ms/örnek	104.091 örnek/s	70/10/20 + 5 Katlı ÇD
CNN-BiLSTM (Önerilen)	CIC-IDS2017	M	%99,82	%82,79	%79,38	%79,87	%0,0003	%99,88	0,0316 ms/örnek	31.609 örnek/s	70/10/20 + 5 Katlı ÇD
Önerilen Modeller (Bu Çalışma) – UNSW-NB15											
CNN (Önerilen)	UNSW-NB15	B	%95,23	%96,38	%94,07	%96,17	%3,17	%99,42	0,0127 ms/örnek	80.055 örnek/s	70/10/20 + 5 Katlı ÇD
CNN-BiLSTM (Önerilen)	UNSW-NB15	B	%96,82	%98,65	%95,94	%97,28	%3,25	%99,77	0,0201 ms/örnek	49.853 örnek/s	70/10/20 + 5 Katlı ÇD
CNN (Önerilen)	UNSW-NB15	M	%72,08	%49,86	%66,71	%49,99	%2,83	%96,45	0,0117 ms/örnek	85.428 örnek/s	70/10/20 + 5 Katlı ÇD
CNN-BiLSTM (Önerilen)	UNSW-NB15	M	%76,49	%51,12	%63,21	%51,41	%2,76	%96,11	0,0454 ms/örnek	22.069 örnek/s	70/10/20 + 5 Katlı ÇD
Random Forest (Önerilen)	UNSW-NB15	B	%98,54	%99,50	%98,21	%98,85	%0,87	%99,89	0 ,0059 ms/örnek	170004,8461 örnek/s	70/10/20
XGBoost (Önerilen)	UNSW-NB15	B	%98,52	%99,45	%98,22	%98,83	%0,96	%99,91	0,0003 ms/örnek	3796487,8874 örnek/s	70/10/20
XGBoost (Önerilen)	UNSW-NB15	M	%84,07	%86,88	%84,07	%61,17	%17,8	%97,05	0,0003 ms/örnek	3073461,26 örnek/s	10 Katlı ÇD

B = İkili (Binary), M = Çok Sınıflı (Multiclass). "—" = Orijinal çalışmada raporlanmamıştır. samples/s = saniye başına işlenen örnek sayısı, ms/sample = örnek başına milisaniye cinsinden tespit süresi.

[a](Song vd., 2023) tüm veri kümesi üzerindeki toplam süreyi (eğitim + çıkarım) raporlamaktadır; örnek başına tespit süresi verilmemiştir. Ayrıca samples/s cinsinden işleme hızı belirtilmemiştir.

[b] (Jouhari vd., 2024) tüm veri kümesi için eğitim + tahmin süresini raporlamaktadır. Örnek başına tespit süresi ve işleme hızı belirtilmemiştir.

[c] (Dai vd., 2024) yalnızca Doğruluk, Tespit Oranı ve Yanlış Pozitif Oranı metriklerini raporlamaktadır. Orijinal makalede kullanılan veri kümesi bölünmesi belirtilmemiştir.

[d] (Talukder vd., 2024) Accuracy, Precision, Recall ve F1-Skoru için aynı değerleri raporlamaktadır. Bu durum, sonuçların dengelenmiş değerlendirme kümeleri üzerinde hesaplanan ağırlıklı ortalamaları temsil ettiğini düşündürmektedir. ROC-AUC, FPR, tespit süresi ve işleme hızı raporlanmamıştır.

[†] Bu çalışmada önerilen tüm derin öğrenme modelleri (CNN ve CNN-BiLSTM), veri sızıntısını önlemek amacıyla yalnızca her eğitim katında uygulanan 5 katlı çapraz doğrulama, quantile-based transformer ve SMOTE yöntemlerini kullanmaktadır. Makine öğrenmesi tabanlı temel modeller (Random Forest ve XGBoost) ise UNSW-NB15 veri kümesinde 70/20/10 tabakalı veri bölünmesi ve kat içi target encoder yaklaşımı ile değerlendirilmiştir.

[‡] Ön işleme adımları, özellik mühendisliği yöntemleri, eğitim/doğrulama protokolleri ve kullanılan veri kümesi sürümleri arasındaki farklılıklar nedeniyle çalışmalar arasında tam anlamıyla doğrudan nicel karşılaştırma yapmak mümkün değildir.

[§] Random Forest ve XGBoost modelleri için işleme hızı ölçülmemiştir. Deneysel kurulumda öncelik sınıflandırma performans metriklerine verilmiş olup, değerlendirme aşamasında çıkarım süresine ilişkin ayrıntılı performans ölçümü gerçekleştirilmemiştir.

5. SONUÇ VE GELECEK ÇALIŞMALAR

Bu tez, ağ saldırı tespiti problemine derin öğrenme mimarilerinin uygulanmasını incelemiş hem ikili hem de çok sınıflı sınıflandırma ortamlarında iki model ailesi olan CNN ve CNN-BiLSTM'yi, iki kıyaslama veri seti olan CIC-IDS2017 ve UNSW-NB15 üzerinde değerlendirmiştir. Deneysel değerlendirme, titiz ve yeniden üretilebilir bir metodolojik çerçeveye dayandırılmıştır: Focal Loss eğitimini içeren birleşik bir ön işleme hattı ve kapsamlı bir değerlendirme metrikleri kümesi. Derin öğrenme sonuçları için karşılaştırmalı bir referans sunmak amacıyla, UNSW-NB15 üzerinde Rastgele Orman ve XGBoost olmak üzere iki klasik makine öğrenmesi temel modeli ek olarak değerlendirilmiştir. (Song vd., 2023, 2023)

5.1. Araştırma Sorularına Yanıtlar

AS1 açısından hem CNN hem de CNN-BiLSTM ikili sınıflandırmada yüksek doğruluk ve güçlü ayırt edici performans sergilemekte; tüm konfigürasyonlarda ROC-EAA (Alıcı İşletim Karakteristiği – Eğri Altı Alan / Receiver Operating Characteristic – Area Under the Curve) değerleri sürekli olarak 0,994'ün üzerinde seyretmektedir. Çok sınıflı sınıflandırmada performans beklenen biçimde düşmekte; makro F1-Skorları 0,500 ile 0,714 arasında değişerek ağır sınıf dengesizliği altında ayrıntılı saldırı türü ayrımının doğasındaki güçlüğü yansıtmaktadır.

AS2 açısından, CNN-BiLSTM özellikle zamansal örüntü karmaşıklığının daha yüksek olduğu ikili UNSW-NB15 ortamında CNN'ye kıyasla anlamlı bir avantaj sağlamaktadır. CIC-IDS2017'de bu avantaj ikili modda marjinal düzeyde kalmakta ve çok sınıflı modda tersine dönmektedir; bu modda CNN daha kararlı bir yapı sergileyerek daha yüksek makro F1 0,701 değeri elde etmektedir. CNN-BiLSTM avantajı dolayısıyla göreve ve veri setine bağımlı olmakta; bu durum 4 ila 6 kat daha yüksek çıkarım gecikmesi şeklinde tutarlı bir hesaplama maliyetiyle birlikte gelmektedir.

AS3 açısından, klasik makine öğrenmesi temel modelleri UNSW-NB15'te ikili tespitte derin öğremeyle rekabet edebilmekte; özellikle XGBoost, eğitim süresinin çok küçük bir bölümünde eşdeğer F1 performansına ulaşmaktadır. Derin öğrenme modelleri bu ölçekte ve bu ön işleme hattıyla belirgin bir avantaj ortaya koyamamaktadır. Bu durum, DL'nin faydasının daha büyük ölçekte veya özellik mühendisliği gerektirmeyen ham girdi temsilleriyle daha belirgin hale geldiğine işaret etmektedir.

5.2. Sınırlılıklar

Bu çalışmanın birkaç sınırlılığı bulunmaktadır edilmeyi hak etmektedir. İlk olarak, değerlendirme yalnızca iki statik kıyaslama veri seti üzerinde gerçekleştirilmiştir. CIC-IDS2017 ve UNSW-NB15, STES (Saldırı Tespit ve Engelleme Sistemi) literatüründe en yaygın kullanılan referans kıyaslamalar arasında yer almakla birlikte, kontrollü laboratuvar ortamlarında yakalanan ağ trafiğini temsil etmekte ve gerçek dünya üretim trafiğinin çeşitliliğini ile değişkenliğini tam anlamıyla yansıtmayabilmektedir. İkinci olarak, modeller akış verisi üzerinde değerlendirilmemiştir; bu durum, kavram kayması ve dağılım değişimi dahil canlı ağ izlemenin zamansal dinamiklerinin bu çalışmada ele alınmadığı anlamına gelmektedir. Üçüncü olarak, dikkatle ayarlanmış olmakla birlikte hiper parametre konfigürasyonları, tüm model-veri seti kombinasyonları genelinde birleşik otomatik bir arama aracılığıyla ortak optimize edilmemiştir; bu durum marjinal performans kazanımlarının gerçekleşmemiş kalmasına yol açmış olabilir. Son olarak, sızıntısız ve metodolojik açıdan sağlam olmakla birlikte özellik mühendisliği hattı, uyarlama yapılmadan ham paket düzeyinde veya akış düzeyinde verilere doğrudan aktarılamayabilecek bir düzeyde alan bilgisi içeren ön işleme gerektirmektedir.

SMOTE'un bilinen sınırlılıkları sınıf sınırı farkındalığının yokluğu ve gürültüye duyarlılık bu çalışmada ele alınmamıştır; SMOTE-ENN -veya ADASYN gibi hibrit yaklaşımlar gelecekteki çalışmalarda araştırılabilir.

5.3. Gelecek Çalışmalar

Bu tezin bulgularından ve sınırlılıklarından doğal olarak birkaç gelecek araştırma yönü ortaya çıkmaktadır.

En doğrudan etkili yön, azınlık sınıfı tespitini kapsamaktadır. Değerlendirilen tüm modellerin nadir saldırı kategorilerindeki tutarlı başarısızlığı, daha hedefli stratejilerin geliştirilmesini zorunlu kılmaktadır. Bu zorluk literatürde de geniş yankı bulmaktadır: (Asry vd., 2025) SHAP tabanlı özellik seçimini XGBoost sınıflandırıcısıyla birleştirerek UNSW-NB15 veri setindeki "Worms" gibi nadir saldırı sınıfları için 0,819 F1-Skoru elde etmiş ve yorumlanabilirlik odaklı yaklaşımların az temsil edilen kategorilerdeki performansı anlamlı biçimde artırabileceğini ortaya koymuştur. Bu doğrultuda gelecekteki çalışmalarda uygulanabilecek stratejiler arasında değişen aşırı örnekleme oranlarına sahip sınıf koşullu SMOTE, sınıf başına ağırlık ayarlamalı maliyet duyarlı kayıp fonksiyonları, saldırı ailesine göre uzmanlaşmış ikili sınıflandırıcıları birleştiren

topluluk yöntemleri ve çok sınırlı örneklerden genelleme yapabilen az örnekli öğrenme yaklaşımları sayılabilir.

İkinci bir yön, eşik optimizasyonunu kapsamaktadır. Sonuçlar, sınıf başına F1-Skoru düşük olanlar da dahil olmak üzere tüm değerlendirilen modellerin makro düzeyde yüksek ROC-EAA değerlerini koruduğunu göstermektedir. Bu durum, olasılık çıktılarının bilgilendirici nitelik taşıdığına; ancak varsayılan 0,5 eşiğinin dengesiz çok sınıflı ortamlar için optimal olmadığına işaret etmektedir. Farklı saldırı türlerinin göreli ağırlığını yansıtan operasyonel maliyet fonksiyonlarından yararlanarak doğrulama verisi kullanımıyla gerçekleştirilen sınıf bazlı eşik kalibrasyonu, model mimarisi değiştirilmeksizin önemli iyileştirmeler sağlayabilir.

Üçüncü bir yön ise daha gelişmiş mimarilerin değerlendirilmesini kapsamaktadır. Doğal dil işleme ve zaman serisi tahmininde sıralı veriler üzerinde güçlü performans sergileyen Transformer tabanlı modeller, burada değerlendirilen BiLSTM bileşeninin doğal bir uzantısını temsil etmektedir. Öz-dikkat mekanizmaları, özellikle daha zengin zamansal yapıya sahip veri setlerinde, ağ akışı dizilerindeki uzun menzilli bağımlılıkları tekrarlayan birimlerden daha etkin biçimde yakalama potansiyeline sahiptir. Bu potansiyel, yakın dönem literatürde deneysel destek bulmaktadır: (Manocchio vd., 2024), akış tabanlı STES'ler için kapsamlı bir Transformer çerçevesi olan FlowTransformer'ı sunmuş; sığ kodlayıcı, sığ kod çözücü ve GPT tarzı mimariler dahil olmak üzere çeşitli Transformer bileşenlerini performans, çıkarım hızı ve model boyutu açısından karşılaştırmalı olarak değerlendirmiş ve bu mimarilerin akış düzeyindeki saldırı tespitinde CNN ve LSTM tabanlı yaklaşımlara karşı rekabetçi bir alternatif oluşturduğunu ortaya koymuştur. CNN1D ve CNN-BiLSTM'nin bu tür mimarilerle aynı kıyaslama koşulları altında sistematik biçimde karşılaştırılması, gelecekteki çalışmalar için anlamlı bir araştırma yönü teşkil etmektedir.

KAYNAKÇA

Alashjaee, A. M., ve Alqahtani, F. (2025). Enhanced intrusion detection system IoT network security model by feed forward neural network and machine learning. *Scientific Reports*, 15(1), 36085.

Asry, C. E. L., Benchaji, I., Douzi, S., ve Ouahidi, B. E. L. (2025). Enhancing cybersecurity: A high-performance intrusion detection approach through boosting minority class recognition. *PLOS One*, 20(3), e0317346. https://doi.org/10.1371/journal.pone.0317346

Brown, L. V. (2008). *Computer security: Principles and practice*. Pearson Prentice Hall.

Coco, A., Dias, T., & van Benthem, T. (2022). Illegal: The SolarWinds Hack under International Law. *European Journal of International Law*, 33(4), 1275-1286. https://doi.org/10.1093/ejil/chac063

Colonial Pipeline: The DarkSide Strikes. (t.y.). [Legislation]. Geliş tarihi 02 Mayıs 2026, gönderen https://www.congress.gov/crs-product/IN11667

Dai, W., Li, X., Ji, W., ve He, S. (2024). Network intrusion detection method based on CNN-BiLSTM-attention model. *IEEE access*, 12, 53099-53111.

Farhan, M., Waheed Ud Din, H., Ullah, S., Hussain, M. S., Khan, M. A., Mazhar, T., Khattak, U. F., ve Jaghdam, I. H. (2025a). Network-based intrusion detection using deep learning technique. *Scientific Reports*, 15(1), 25550.

Henry, A., Gautam, S., Khanna, S., Rabie, K., Shongwe, T., Bhattacharya, P., Sharma, B., ve Chowdhury, S. (2023). Composition of hybrid deep learning model and feature optimization for intrusion detection system. *Sensors*, 23(2), 890.

Ioannidou, P. (2021). *Anomaly Detection in Computer Networks*. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1557602

Jouhari, M., Benaddi, H., ve Ibrahimi, K. (2024). Efficient intrusion detection: Combining x 2 feature selection with CNN-BiLSTM on the UNSW-NB15 dataset. *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 1-6. https://ieeexplore.ieee.org/abstract/document/10658099/

Krizhevsky, A., Sutskever, I., ve Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Lin, T.-Y., Goyal, P., Girshick, R., He, K., ve Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE international conference on computer vision*, 2980-2988.http://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

Magazine, C. (2018, Şubat 21). Cybercrime To Cost The World $10.5 Trillion Annually By 2025. *Cybercrime Magazine*. https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/

Manocchio, L. D., Layeghy, S., Lo, W. W., Kulatilleke, G. K., Sarhan, M., & Portmann, M. (2024). FlowTransformer: A transformer framework for flow-based network intrusion detection systems. *Expert Systems with Applications*, 241, 122564. https://doi.org/10.1016/j.eswa.2023.122564

Moustafa, N., ve Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1-6. https://doi.org/10.1109/MilCIS.2015.7348942

Peng, H., Wu, C., ve Xiao, Y. (2023). CBF-IDS: Addressing Class Imbalance Using CNN-BiLSTM with Focal Loss in Network Intrusion Detection System. *Applied Sciences*, 13(21), 11629. https://doi.org/10.3390/app132111629

Sharafaldin, I., Lashkari, A. H., ve Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1(2018), 108-116.

Song, J., Wang, X., He, M., ve Jin, L. (2023). CSK-CNN: Network intrusion detection model based on two-layer convolution neural network for handling imbalanced dataset. *Information*, 14(2), 130.

Talukder, Md. A., Islam, Md. M., Uddin, M. A., Hasan, K. F., Sharmin, S., Alyami, S. A., ve Moni, M. A. (2024). Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction. *Journal of Big Data*, 11(1), 33. https://doi.org/10.1186/s40537-024-00886-w

Türk, F. (2023). Analysis of Intrusion Detection Systems in UNSW-NB15 and NSL-KDD Datasets with Machine Learning Algorithms. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 12(2), 465-477. https://doi.org/10.17798/bitlisfen.1240469

Udurume, M., Shakhov, V., ve Koo, I. (2024). Comparative analysis of deep convolutional neural network—Bidirectional long short-term memory and machine learning methods in intrusion detection systems. *Applied Sciences*, 14(16), 6967.

Vinayakumar, R., Soman, K. P., ve Poornachandran, P. (2017). Applying convolutional neural network for network intrusion detection. *2017 International conference on advances in computing, communications and informatics (ICACCI)*, 1222-1228. https://ieeexplore.ieee.org/abstract/document/8126009/

Wilcoxon Test: Comparing Paired Samples. (2025, Eylül 1). https://numiqo.com/tutorial/wilcoxon-test

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Rachid CHEICK MOHAMED

Eğitim Durumu

Lisans Öğrenimi : Matematik-Bilgisayar Lisans Programı (Bilgisayar Bilimleri Alanı), Ngaoundéré Üniversitesi, Kamerun (2017-2020)

Yüksek Lisans Öğrenimi : Yapay Zekâ ve Akıllı Sistemler, Gümüşhane Üniversitesi, Türkiye (2024–2026)

Bildiği Yabancı Diller : Fransızca, İngilizce (Anadil), Türkçe (C1)

İş Deneyimi

Projeler : CyberShield PME — Orta Afrika KOBİ'leri için yapay zekâ destekli siber güvenlik platformu (POESAM 2026 yarışma projesi)

Çalıştığı Kurumlar :

Tarih : 03 / 07 / 2026