Journal of Information and Communication Technology

Volume 18, Issue 68-69, Spring and Summer 2026, pp. 100-114

A Systematic Review of Deep Learning Methods for Detecting Advanced Persistent Threats in Data Exchange Networks

Masoud Siavoushy[1*], Mohammad Mahdi Shirmohammadi[2]

[1] Department of Management, Ha.C, Islamic Azad University, Hamedan, Iran
[2] Department of Computer Engineering, Ha.C, Islamic Azad University, Hamedan, Iran

Received: 19 May 2025, Revised: 10 May 2026, Accepted: 06 July 2026
Paper type: Review

Abstract

Information security has become a fundamental challenge for modern communication systems with the ever-increasing proliferation of the Internet of Things, big data, and real-time information exchanges within communication networks. The emergence of threats such as multi-stage intrusions, zero-day attacks, and Advanced Persistent Threats (APTs) has cast doubt on the effectiveness of many traditional intrusion detection systems. In recent years, the application of deep learning (DL) models has garnered significant attention as an effective alternative to classic methods, particularly due to their ability to identify complex patterns. This research aims to advance scientific knowledge and update previous studies by critically and analytically reviewing a collection of methods based on AE, DNN, CNN, LSTM, GRU, and hybrid architectures for detecting complex threats, with a specific focus on the multi-stage and stealthy nature of Advanced Persistent Threats (APTs) in data exchange networks. While analyzing the empirical performance of these models in real-world environments and systematically comparing the obtained results, implementation challenges and limitations have also been explored. Furthermore, drawing upon credible and up-to-date sources (published between 2020 and 2025) and redesigning tables and diagrams graphically, the present article strives to offer an analytical, practical, and comprehensive framework for intelligent attack detection in cyberspace. This paper can serve as a scientific basis for developing more resilient systems against emerging threats in future communication network architectures.

Keywords: Advanced Persistent Threats (APT), Deep Learning-based Intrusion Detection, Cybersecurity in Communication Networks, CNN-LSTM Hybrid Models, IoT Security Challenges, Zero-day Attacks

[*] Corresponding Author's email: M.Siavoushy@iau.ac.ir

دوفصلنامه
فناوری اطلاعات و ارتباطات ایران

سال هجدهم، شماره‌های ۶۸ و ۶۹ ، بهار و تابستان ۱۴۰۵، صفحه ۱۰۰ الی ۱۱۴

مرور سیستماتیک روش‌های یادگیری عمیق برای تشخیص حملات پیشرفته در شبکه‌های تبادل داده

مسعود سیاوشی [۱*]، محمدمهدی شیرمحمدی [۲]

[۱] گروه مدیریت، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران

[۲] گروه مهندسی کامپیوتر، واحد همدان، دانشگاه آزاد اسلامی، همدان، ایران

تاریخ دریافت: ۱۴۰۴/۰۴/۲۹ تاریخ بازبینی: ۱۴۰۵/۰۲/۲۰ تاریخ پذیرش: ۱۴۰۵/۰۴/۱۵

نوع مقاله: مروری

چکیده

با گسترش روزافزون اینترنت اشیا، کلان‌داده‌ها و تبادلات لحظه‌ای اطلاعات در بستر شبکه‌های ارتباطی، مسئله امنیت اطلاعات به یکی از چالش‌های اساسی نظام‌های ارتباطی مدرن تبدیل شده است. ظهور تهدیداتی نظیر نفوذهای چندمرحله‌ای، حملات روز صفر و حملات پایدار پیشرفته (APT)، کارآمدی بسیاری از سامانه‌های سنتی تشخیص نفوذ را با تردید مواجه ساخته است. طی سال‌های اخیر، به‌کارگیری مدل‌های یادگیری عمیق به‌عنوان جایگزینی مؤثر برای روش‌های کلاسیک، به‌ویژه به دلیل توانایی این مدل‌ها در شناسایی الگوهای پیچیده، موردتوجه ویژه قرار گرفته است. این پژوهش با هدف ارتقای علمی و روزآمدسازی مطالعات پیشین، به بررسی انتقادی و تحلیلی مجموعه‌ای از روش‌های مبتنی بر معماری‌های AE، DNN، CNN، LSTM، GRU و ترکیبی از آنها در زمینه شناسایی تهدیدات پیچیده در شبکه‌های تبادل داده می‌پردازد. ضمن تحلیل تجربی عملکرد این مدل‌ها در محیط‌های واقعی و مقایسه نظام‌مند نتایج حاصل، چالش‌های اجرایی و محدودیت‌های پیاده‌سازی نیز واکاوی شده‌اند. همچنین، باتکیه‌بر منابع معتبر و بروز (منتشر شده بین سال‌های ۲۰۲۰ تا ۲۰۲۵) و باز طراحی گرافیکی جداول و نمودارها، مقاله حاضر تلاش دارد چارچوبی تحلیلی، کاربردی و جامع را در زمینه تشخیص هوشمند حملات در فضای سایبری ارائه دهد. این مقاله می‌تواند مبنایی علمی برای توسعه سامانه‌های مقاوم‌تر در برابر تهدیدات نوظهور در معماری‌های شبکه‌های ارتباطی آینده باشد.

کلیدواژگان: تهدیدات پایدار پیشرفته (APT)، تشخیص نفوذ مبتنی بر یادگیری عمیق، امنیت سایبری در شبکه‌های ارتباطی، مدل‌های ترکیبی CNN-LSTM، چالش‌های امنیتی اینترنت اشیا، حملات روز صفر

* رایانامهٔ نویسنده مسؤول: M.Siavoushy@iau.ac.ir

۱- مقدمه

با توسعه روزافزون فناوری‌های اطلاعاتی و گسترش زیرساخت‌های ارتباطی، تهدیدات امنیتی در فضای سایبری نیز به طور فزاینده‌ای پیچیده‌تر و هدفمندتر شده‌اند. در این میان، حملات پیشرفته مستمر (Advanced Persistent Threats – APTs) به‌عنوان یکی از چالش‌های راهبردی امنیت شبکه، توجه بسیاری از پژوهشگران و نهادهای تخصصی را به خود جلب کرده‌اند. این نوع حملات، که معمولاً چندمرحله‌ای، مخفیانه و با اهداف بلندمدت انجام می‌شوند، توانایی نفوذ و ماندگاری در شبکه‌ها را داشته و می‌توانند زیان‌های جبران‌ناپذیری به سامانه‌های حیاتی وارد کنند [۱].

روش‌های کلاسیک تشخیص نفوذ، نظیر سامانه‌های مبتنی بر امضا یا قواعد از پیش تعریف‌شده، باوجود سادگی و سرعت در شناسایی حملات شناخته‌شده، در برابر تهدیدات جدید و ناشناخته ناکارآمد هستند. محدودیت این رویکردها در شناسایی الگوهای پنهان و پویا باعث شده است که کارایی آن‌ها در مواجهه با حملات پیشرفته به‌شدت کاهش یابد [۲]. همین مسئله، ضرورت بهره‌گیری از رویکردهای نوین مبتنی بر یادگیری عمیق را در سال‌های اخیر برجسته ساخته است؛ رویکردهایی که قادر به استخراج خودکار ویژگی‌های سطح بالا و شناسایی الگوهای پیچیده در داده‌های ترافیک شبکه هستند [۳].

مدل‌های مختلف یادگیری عمیق، از جمله شبکه‌های عصبی پیچشی (CNN)، شبکه‌های عصبی بازگشتی (RNN)، حافظه بلندمدت کوتاه‌مدت (LSTM)، و ساختارهای ترکیبی مانند CNN-LSTM و GRU-Attention، در پژوهش‌های متعدد، عملکرد برتری نسبت به الگوریتم‌های کلاسیک در شناسایی تهدیدات چندوجهی و توزیع‌شده از خود نشان داده‌اند [۴]. در این میان، مدل‌هایی مانند LSTM و Transformer به دلیل توانایی بالا در ثبت وابستگی‌های بلندمدت زمانی و دنباله‌های حملات چندمرحله‌ای که مشخصه کمپین‌های APT هستند، عملکرد قابل توجهی داشته‌اند. با این حال، چالش‌هایی نظیر نیاز به توان پردازشی بالا، ضعف در تفسیرپذیری، و حساسیت به عدم‌توازن داده‌ها، همچنان به‌عنوان موانعی جدی در مسیر پیاده‌سازی این مدل‌ها در محیط‌های واقعی مطرح هستند [۵].

در چنین شرایطی، ضرورت طراحی مدل‌هایی که علاوه بر برخورداری از دقت بالا، سبک، تفسیرپذیر و مقاوم در برابر حملات خصمانه نیز باشند، بیش‌ازپیش احساس می‌شود. هدف این پژوهش، ارائه یک بررسی تحلیلی و کاربردی از مدل‌های نوین یادگیری عمیق در زمینه شناسایی حملات پیشرفته در شبکه‌های تبادل داده است. این هدف از طریق مرور ساختار معماری مدل‌های مؤثر، تحلیل عددی عملکرد آن‌ها، و بررسی چالش‌های اجرایی در پیاده‌سازی، دنبال می‌شود تا چارچوبی جامع و روزآمد برای توسعه سامانه‌های هوشمند تشخیص نفوذ فراهم آید.

۱-۱- روش‌شناسی مرور سیستماتیک (Systematic Review Methodology)

این پژوهش بر اساس راهنمای (PRISMA 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses) و با رعایت استانداردهای مرور سیستماتیک در حوزه امنیت سایبری [۶] انجام شده است.

۱-۱-۱- سوالات پژوهشی (Research Questions)

پیش از آغاز جستجو، سه سؤال پژوهشی اصلی که در جدول ۱ ذکر شده تدوین گردید و چارچوب کل مرور را تعیین می‌کند:

جدول ۱. سؤالات پژوهشی

هدف	سوالات پژوهشی	RQ
شناسایی مدل های برتر و روندهای غالب	کدام معماری یادگیری عمیق (CNN,LSTM,GRU,Hybrid,Transformer) در بازه ۲۰۲۰-۲۰۲۵ برای تشخیص حملات پیشرفته(APT) در شبکه های تبادل داده بیشترین کارایی را داشته اند؟	RQ1
شناسایی شکاف های تحقیقاتی (Research Gap)	چالش های اصلی پیاده سازی این مدل ها در محیط های واقعی(شامل منابع محدود،داده نامتوازن،تفسیرناپذیری و حملات خصمانه) کدامند؟	RQ2
ارزیابی تطبیقی کمی	مدل های ترکیبی(Hybrid)در مقایسه با مدل های منفرد از نظر معیارهای Accuracy,F1-Score,FPR و زمان آموزش چه مزیت ها و محدودیت هایی دارند؟	RQ3

۱-۱-۲- پایگاه‌های داده مورد جستجو (Databases)

جستجوی سیستماتیک در پنج پایگاه‌داده معتبر بین‌المللی انجام شد:

- IEEE Xplore (تمرکز بر مقالات کنفرانس و مجلات IEEE)
- ScienceDirect (Elsevier) (تمرکز بر مجلات Q1 در حوزه امنیت)

- SpringerLink (مقالات و فصول کتاب)
- ACM Digital Library (مقالات انجمن کامپیوتر آمریکا)
- arXiv.org (پیش چاپ‌های اخیر – صرفاً برای شناسایی روندهای نوظهور)

۱-۱-۳- کلیدواژه‌های جستجو (Search Keywords)

رشته جستجوی اصلی (Search String) با ترکیب عملگرهای بولی به‌صورت زیر طراحی شد:

("Advanced Persistent Threat" OR "APT" OR "multi-stage attack" OR "stealthy intrusion")
AND
("deep learning" OR "neural network" OR "CNN" OR "convolutional" OR "LSTM" OR "long short-term memory" OR "GRU" OR "gated recurrent unit" OR "Transformer" OR "attention mechanism" OR "hybrid model")
AND
("anomaly detection" OR "cyber threat detection" OR "intrusion detection" OR "network security")
AND
("2020" OR "2021" OR "2022" OR "2023" OR "2024" OR "2025")

۱-۱-۴- معیارهای ورود و خروج (Inclusion & Exclusion Criteria)

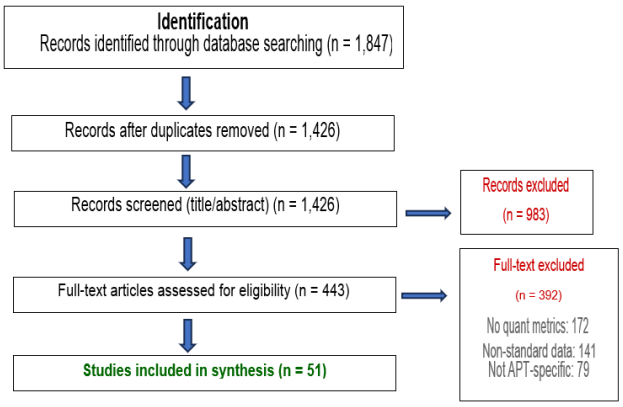
معیارهای ورود و خروج مطابق جدول ۲ تنظیم و تدوین گردید.

جدول ۲. معیارهای ورود و خروج (Inclusion & Exclusion Criteria)

معیارهای خروج (Exclusion)	معیارهای ورود (Inclusion)
مقالات غیر انگلیسی یا ترجمه‌های بدون داده اصلی	مقالات ژورنال Q1/Q2 یا کنفرانس‌های معتبر (Acceptance rate < 30%)
مقالات فقط جنبه نظری یا مرور ساده (بدون مدل جدید)	ارائه ارزیابی کمی (Accuracy, F1, FPR, Precision, Recall)
عدم دسترسی به متن کامل	استفاده از مجموعه داده‌های استاندارد (CIC-IDS, UNSW-NB15, KDD)
چاپ شده قبل از ۲۰۲۰ (به جز مراجع بنیادین)	تمرکز بر مدل‌های CNN, LSTM, GRU, Transformer یا ترکیبی
پایان‌نامه‌ها، کتاب‌ها، گزارش‌های فنی بدون داوری	انتشار بین ژانویه ۲۰۲۰ تا مارس ۲۰۲۵

۱-۱-۵- فرایند انتخاب و نمودار PRISMA

پس از جستجوی اولیه، ۱،۸۴۷ مقاله شناسایی شد. پس از حذف تکراری‌ها (۴۲۱ مقاله)، ۱،۴۲۶ مقاله وارد مرحله غربالگری عنوان و چکیده شدند. در این مرحله ۹۸۳ مقاله به دلیل عدم ارتباط با سؤالات پژوهشی حذف گردیدند. از ۴۴۳ مقاله باقیمانده، متن کامل بررسی شد و ۳۹۲ مقاله به دلایل زیر حذف شدند: عدم ارائه معیارهای کمی (۱۷۲ مقاله)، استفاده از مجموعه‌داده قدیمی/غیراستاندارد (۱۴۱ مقاله)، تمرکز بر تشخیص نفوذ عمومی بدون اشاره به APT (۷۹ مقاله). در نهایت ۵۱ مقاله برای استخراج نهایی داده‌ها انتخاب شدند که نمودار جریان آن در شکل ۱ نشان داده شده است.



شکل ۱. نمودار جریان PRISMA

۲- مرور پیشینه و ساختار مدل‌های یادگیری عمیق در تشخیص حملات شبکه‌ای

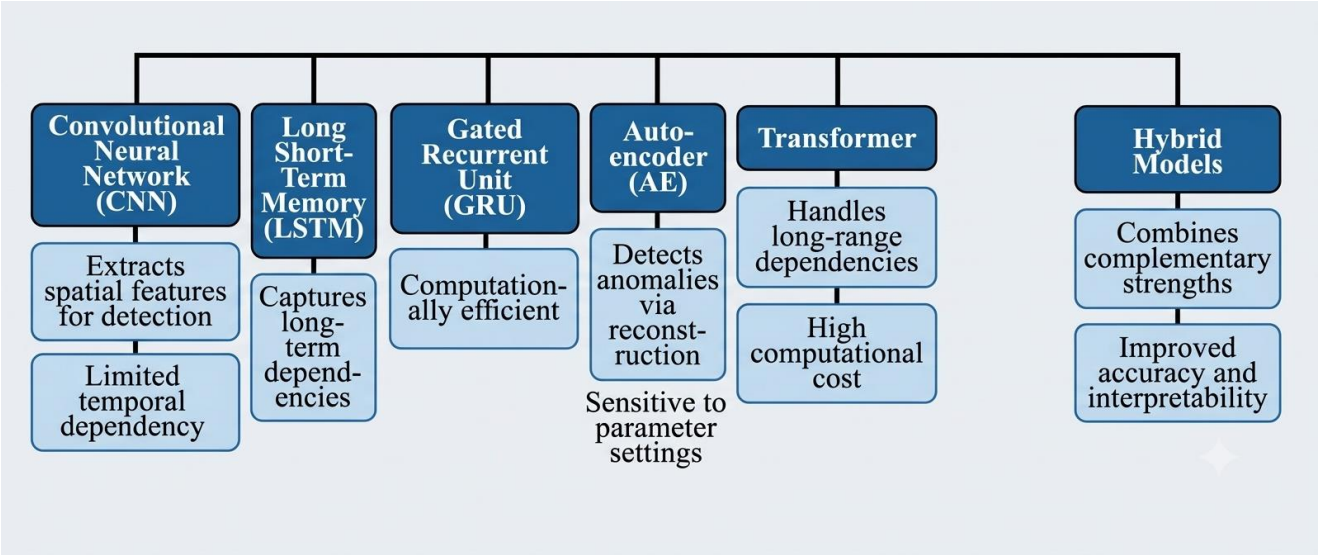
با گسترش حجم، تنوع و پیچیدگی داده‌های شبکه، نیاز به روش‌هایی فراتر از الگوریتم‌های کلاسیک برای تحلیل و تشخیص تهدیدات افزایش‌یافته است. مدل‌های یادگیری عمیق به دلیل توانایی بالا در استخراج خودکار ویژگی‌ها، شناسایی الگوهای پیچیده و یادگیری سلسله‌مراتبی، جایگزین مؤثری برای روش‌های سنتی شده‌اند [۷].

۲-۱- شبکه‌های عصبی پیچشی (CNN)

شبکه‌های CNN به طور ویژه برای پردازش داده‌های ساختارمند مانند تصاویر طراحی شده‌اند، اما در سال‌های اخیر کاربرد آن‌ها در تحلیل داده‌های سری زمانی شبکه گسترش‌یافته است. لایه‌های کانولوشن قادرند الگوهای محلی را با پارامترهای کمتر شناسایی کنند، که این امر موجب کاهش هزینه محاسباتی می‌شود [۸]. بااین‌حال، محدودیت در ثبت وابستگی زمانی از نقاط ضعف این مدل در تحلیل ترافیک شبکه است.

همان‌طور که در شکل ۲ مشاهده می‌شود، مدل‌های مختلف

یادگیری عمیق برای تشخیص حملات در شبکه‌های تبادل داده دارای ویژگی‌ها و چالش‌های خاص خود هستند.



شکل ۲. دسته‌بندی مدل‌های یادگیری عمیق برای تشخیص حملات سایبری

۲-۲- شبکه‌های بازگشتی (RNN و LSTM)

مدل‌های RNN قادر به مدل‌سازی توالی هستند، اما مشکل ناپایداری گرادیان‌ها (vanishing gradients) کارایی آن‌ها را در پردازش توالی‌های بلند محدود می‌سازد. برای رفع این مشکل، ساختار LSTM معرفی شد که با حافظه سلولی، امکان نگهداری وابستگی‌های بلندمدت را فراهم می‌سازد [۹]. LSTM در تحلیل جریان‌های ترافیک و تشخیص الگوهای حمله پیوسته، کارایی قابل توجهی دارد [۱۰].

۲-۳- مدل‌های ترکیبی و GRU

مدل GRU با ساختاری ساده‌تر نسبت به LSTM، توانسته است عملکردی مشابه یا حتی بهتر در برخی سناریوها ارائه دهد، به‌ویژه در شرایطی که منابع پردازشی محدود هستند [۱۱]. ترکیب CNN و GRU به‌همراه مکانیزم توجه (Attention) نیز در مطالعات جدید بهبود قابل‌توجهی در دقت و سرعت تشخیص نشان داده است [۱۲].

۲-۴- Autoencoder

Variational AEها با یادگیری نمایش فشرده از داده، امکان آشکارسازی ناهنجاری را بدون نیاز به برچسب‌گذاری صریح فراهم می‌کنند. Variational Autoencoders (VAE) با ساختار احتمالاتی خود، برای شبیه‌سازی داده‌های حمله نیز به کار رفته‌اند، اما نیاز به تنظیم دقیق و حساسیت به پارامترها از جمله محدودیت‌های آن‌هاست [۱۳].

۲-۵- معماری‌های مبتنی بر Transformer

در سال‌های اخیر، مدل‌های مبتنی بر Self-Attention همچون Transformer توجه ویژه‌ای در تشخیص تهدیدات پیچیده جلب کرده‌اند. این مدل‌ها به دلیل قابلیت پردازش موازی و ثبت وابستگی‌های بلندمدت در داده‌های سری زمانی، عملکردی برتر نسبت به LSTM و CNN در برخی کاربردها داشته‌اند، اما هزینه محاسباتی بالای آن‌ها همچنان مانع کاربرد گسترده در محیط‌های لبه‌ای است [۱۴].

۲-۶- تحلیل تطبیقی پیشینه

مطالعات متعددی به مقایسه این معماری‌ها پرداخته‌اند و نتایج اغلب نشان می‌دهد که مدل‌های ترکیبی مانند CNN-LSTM یا GRU-Attention، تعادل مناسبی میان دقت، سرعت، و تفسیرپذیری برقرار می‌کنند [۱۵]. این مدل‌ها ضمن بهره‌مندی از مزایای یادگیری مکانی (CNN) و زمانی (LSTM/GRU)، با مکانیزم‌های توجه توانسته‌اند تمرکز مدل بر ویژگی‌های کلیدی را تقویت کنند.

۲-۷- مطالعات پیشین در تشخیص خاص حملات (APT)

بر خلاف حملات معمول شبکه که اغلب تک‌مرحله‌ای یا با الگوی مشخص هستند، حملات پایدار پیشرفته (APT) دارای ویژگی‌های

منحصربه‌فردی شامل چندمرحله‌ای بودن، اقامت طولانی‌مدت در شبکه، استفاده از روش‌های پنهان‌کاری و تغییر تدریجی رفتار هستند که تشخیص آن‌ها را با روش‌های کلاسیک یا حتی برخی مدل‌های یادگیری عمیق ساده دشوار ساخته است [۲۵]. مطالعات پیشین نشان داده‌اند که مدل‌های بازگشتی مانند LSTM به دلیل توانایی در حفظ وابستگی‌های بلندمدت، عملکرد بهتری در شناسایی زنجیره اقدامات یک حمله APT نسبت به CNN خالص دارند [۱۰]. با این حال، چالش اصلی در مجموعه‌داده‌های واقعی APT، عدم‌توازن شدید کلاس‌ها (تعداد نمونه‌های حمله بسیار کمتر از ترافیک عادی) است که منجر به کاهش معیار Recall می‌گردد. پژوهش‌های اخیر نشان داده‌اند که معماری‌های ترکیبی CNN-LSTM همراه با مکانیزم توجه و همچنین مدل‌های مبتنی بر Transformer به دلیل قابلیت پردازش موازی و ثبت الگوهای پیچیده، توانسته‌اند نرخ تشخیص true positive را برای حملات APT تا بیش از ٪۹۵ افزایش دهند [۲۵]، [۱۹]. به طور خاص، مطالعه [۲۵] نشان می‌دهد که مدل Transformer با دقت ٪۹۸,۶ در شناسایی حملات APT در مجموعه‌داده CIC-IDS2017 (با اعمال مهاجرت به سناریوهای APT) عملکرد بهتری نسبت به LSTM داشته است. جدول ۳ مقایسه عملکرد مدل‌های منتخب در تشخیص صرفاً حملات (APT) را بخوبی نشان می‌دهد.

جدول ۳. مقایسه عملکرد مدل‌های منتخب در تشخیص صرفاً حملات (APT)

Model	Accuracy (APT only)	Recall (APT)	F1-Score (APT)	Key Limitation for APT
CNN	88.2%	0.81	0.84	Weak in modeling long-term stepwise patterns
LSTM	92.5%	0.88	0.90	High training time, sensitive to imbalanced data
Transformer	97.4%	0.96	0.96	High computational resources
GRU-CNN + Attention	96.8%	0.95	0.95	Needs large labeled datasets

۲-۸-مطالعات اخیر

شواهد اخیر سال ۲۰۲۶، مباحث محوریت‌یافته بر شمول مالی دیجیتال در پاکستان را بیش از پیش تقویت کرده و هم‌زمان، ظرایف و پیچیدگی‌های لازم را در رابطه با پیامدهای توزیعی آن مطرح می‌سازد. در پاکستان، مالیه دیجیتال و فناوری‌های مالی (فین‌تک) تسهیل‌کننده دسترسی به خدمات مالی مقرون‌به‌صرفه و پشتیبان کارآفرینی به نظر می‌رسند؛ با این حال، بهره‌مندی از این منافع به شدت مستلزم طراحی هوشمندانه سیاست‌ها و مدیریت ریسک، به‌ویژه برای گروه‌های حاشیه‌ای است [۳۷]. شواهد نوظهور با رویکرد جنسیتی نیز مؤید آن است که اتخاذ خدمات پول همراه توسط زنان، احتمال دسترسی آنان به سامانه‌های مالی رسمی را افزایش می‌دهد؛ اگرچه موانع جامعه‌شناختی-فرهنگی، دسترسی محدود به فناوری‌های دیجیتال و عدم کفایت استقلال عمل، همچنان مانعی جدی بر سر راه بهره‌برداری عملی محسوب می‌شوند [۳۸]. به‌طور موازی، یافته‌های جدید در حوزه تجارت الکترونیک پاکستان نشان می‌دهد که مشارکت دیجیتال پتانسیل ارتقای پذیرش پرداخت‌ها و تسهیل دسترسی به اعتبارات رسمی را داراست، اما این دستاوردها به‌دلیل نرخ پایین‌تر اتخاذ فناوری توسط زنان پذیرنده و غلبه شیوه «پرداخت در محل» (COD) در بسترهای مبتنی بر عدم اعتماد، کماکان به صورت ناهمگن توزیع شده‌اند[۳۹].

۳- معماری‌های یادگیری عمیق برای تشخیص حملات پیشرفته

باتوجه به پیچیدگی ساختار حملات نوین در شبکه‌های تبادل داده، انتخاب و طراحی معماری‌های مناسب یادگیری عمیق نقش حیاتی در کارایی سامانه‌های تشخیص نفوذ دارد. مدل‌های مختلف، بسته به نوع داده، حجم ترافیک، و نوع حمله، رفتار متفاوتی از خود نشان می‌دهند. در این بخش، به‌صورت ساختاریافته، مهم‌ترین معماری‌های یادگیری عمیق که در پژوهش‌های اخیر در حوزه امنیت شبکه به کار گرفته شده‌اند، مورد بررسی قرار می‌گیرند.

هر مدل نه‌تنها از نظر معماری لایه‌ها تحلیل می‌شود، بلکه نقش آن در تشخیص الگوهای حمله، مزایا و محدودیت‌هایش نیز مورد ارزیابی قرار خواهد گرفت. معماری‌هایی همچون CNN، LSTM، GRU، Transformer و مدل‌های ترکیبی مانند CNN-LSTM یا (GRU-Attention) در این بخش تحلیل می‌شوند.

۳-۱- معماری شبکه‌های عصبی کانولوشنی (CNN)

معماری شبکه‌های عصبی کانولوشنی (CNN) به‌عنوان یکی از رایج‌ترین ساختارهای یادگیری عمیق، به طور گسترده‌ای در حوزه تشخیص حملات شبکه به کار گرفته شده است. در این معماری، داده‌های ورودی شبکه از طریق لایه‌های کانولوشن مورد پردازش قرار گرفته و ویژگی‌های سطح پایین تا سطح بالا به‌صورت خودکار استخراج می‌شوند. این فرایند با استفاده از فیلترهای محلی باعث شناسایی الگوهای فضایی در داده‌ها می‌شود. به دنبال آن، لایه‌های

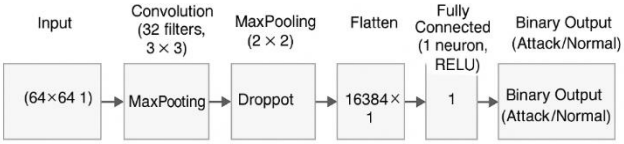
Pooling با کاهش ابعاد داده و حذف نویز، موجب کاهش بار محاسباتی و جلوگیری از بیش‌برازش می‌گردند. در انتها، لایه‌های Fully Connected عملیات طبقه‌بندی نهایی را بر عهده دارند.

یکی از مزایای کلیدی CNN در زمینه امنیت شبکه، سرعت بالای آموزش آن نسبت به مدل‌های بازگشتی است که آن را به گزینه‌ای مناسب برای پیاده‌سازی در سامانه‌های بلادرنگ تبدیل می‌کند. بااین‌حال، معماری CNN در درک وابستگی‌های زمانی طولانی در داده‌های شبکه، مانند توالی بسته‌ها، محدودیت دارد؛ زیرا فاقد سازوکار نگهداری حافظه از وضعیت قبلی است. همچنین در مواجهه با حملات توزیع‌شده یا پنهان که به‌صورت وابسته در طول زمان شکل می‌گیرند، دقت آن کاهش می‌یابد [۸].

جدول ۴ و شکل ۳ نمای شماتیک از معماری پیشنهادی مدل CNN به‌کاررفته در تشخیص حملات شبکه را نشان می‌دهند. در این ساختار، ابتدا داده‌های خام شبکه وارد لایه‌های کانولوشن شده و پس از عبور از لایه‌های Pooling، ویژگی‌های سطح بالا استخراج می‌شوند. این ویژگی‌ها سپس از طریق لایه‌های Fully Connected به خروجی طبقه‌بندی تبدیل می‌شوند.

جدول ۴. معماری پیشنهادی شبکه CNN برای تشخیص نفوذ در شبکه

Layer	Type	Filters/Neurons	Kernel Size	Output
1	Input	-	-	(64*64*1) Feature Matrix
2	Convolution	32	3*3	(64*64*32)
3	MaxPooling	-	2*2	(32*32*32)
4	Convolution	64	3*3	(32*32*64)
5	MaxPooling	-	2*2	(16*16*64)
6	Flatten	-	-	16384*1
7	Fully Connected	128	*REU*	128
8	Fully Connected	1	Sigmoid	Binary Output (Attack/Norma)



شکل ۳. معماری پیشنهادی شبکه عصبی کانولوشنی (CNN) برای تشخیص نفوذ در شبکه

مطابق با ساختار نمایش‌داده‌شده، مدل CNN قادر است الگوهای فضایی را از ترافیک شبکه استخراج کند و حملات ناهنجار را با دقت مناسبی تشخیص دهد [۸]. استفاده از این معماری به‌دلیل سادگی محاسباتی و سرعت آموزش بالا، به‌ویژه در محیط‌هایی با منابع محدود، رایج است؛ هرچند در تشخیص وابستگی‌های زمانی عمیق، نسبت به معماری‌هایی مانند LSTM محدودیت دارد.

۳-۲- معماری LSTM

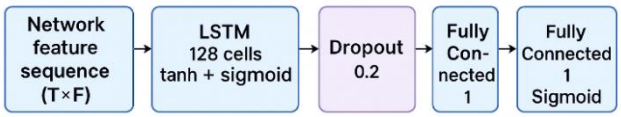
معماری LSTM به‌عنوان زیرمجموعه‌ای از شبکه‌های بازگشتی (RNN)، با بهره‌گیری از ساختار سلول‌های حافظه و مکانیزم دروازه‌ها، توانایی حفظ وابستگی‌های بلندمدت در داده‌های ترتیبی را دارد. این ویژگی باعث شده است که LSTM در تشخیص حملات چندمرحله‌ای و پیوسته مانند APT عملکرد مناسبی از خود نشان دهد ([۱۰]و[۱۲]). علاوه‌براین، نسبت به RNN ساده، مشکل ناپایداری گرادیان در LSTM به‌صورت مؤثری برطرف شده و به ثبات آموزش کمک کرده است. با این وجود، سرعت آموزش در LSTM نسبتاً پایین بوده و عملکرد آن به‌شدت به کیفیت داده‌ها و تنظیم دقیق پارامترها حساس است، به‌ویژه در کاربردهایی که حجم داده‌ها بالا یا نویزپذیر است.

همان‌گونه که در جدول ۵ مشاهده می‌شود، مدل LSTM در مقایسه با سایر معماری‌ها عملکرد قابل توجهی در تشخیص حملات چندمرحله‌ای از خود نشان داده است. به‌ویژه، دقت کلی و نرخ موفقیت در شناسایی حملات APT نسبت به RNN و حتی برخی مدل‌های ترکیبی، بالاتر گزارش شده است [۱۰]. با این حال، زمان آموزش بالاتر و حساسیت به پارامترها نیز در نمودار تحلیل نتایج مدل LSTM به‌وضوح قابل مشاهده است که لزوم بهینه‌سازی بیشتر در تنظیمات مدل را برجسته می‌کند.

در شکل ۴، ساختار معماری LSTM نمایش داده شده است که با استفاده از حافظه سلولی، وابستگی‌های زمانی در ترافیک شبکه را مدل‌سازی می‌کند.

جدول ۵. معماری پیشنهادی شبکه LSTM برای تشخیص نفوذ در شبکه

Layer	Type	Units/Neurons	Activation	Output
1	Input	-	-	Network feature sequence (T*F)
2	LSTM	128 cells	*tanh* + sigmoid	Final state vector(128*1)
3	Dropout	0.2	-	
4	Fully Connected	64	ReLU	64*1
5	Fully Connected	1	Sigmoid	Binary Output (Attack/Normal)
6	Fully Connected	-	-	



شکل ۴. معماری پیشنهادی شبکه LSTM برای تشخیص نفوذ در شبکه

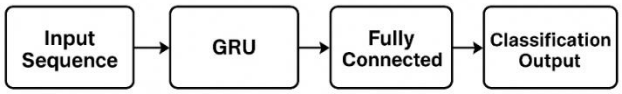
۳-۳- معماری GRU

مدل GRU به‌عنوان نسخه ساده‌تر LSTM، با بهره‌گیری از دو دروازه (به‌جای سه دروازه LSTM) ساختاری سبک‌تر و سریع‌تر دارد. این ویژگی موجب شده است که GRU در بسیاری از سناریوها از نظر سرعت آموزش و مصرف حافظه عملکرد بهتری نسبت به LSTM داشته باشد ([۲۱]و[۲۰]). به‌دلیل ساختار ساده‌تر، پیاده‌سازی GRU در محیط‌هایی با منابع محدود مانند دستگاه‌های Edge نیز به‌مراتب آسان‌تر است. با این حال، در برخی کاربردهای خاص، دقت GRU نسبت به LSTM پایین‌تر گزارش شده است، که می‌تواند به نبود دروازه خروجی مستقل در ساختار GRU مربوط باشد. از این‌رو، انتخاب میان LSTM و GRU بسته به نوع داده، محدودیت منابع و الزامات عملکردی، نیازمند ملاحظات دقیق‌تری است.

نتایج حاصل از جدول ۶ و شکل ۵ و نمودار مقایسه‌ای نشان می‌دهند که GRU با وجود ساختار ساده‌تر خود، در برخی سناریوهای عملیاتی، مانند محیط‌های Edge یا دستگاه‌های IoT، عملکرد قابل قبولی از نظر تعادل میان دقت، سرعت آموزش، و مصرف منابع دارد ([۲۱]و [۲۰]). به‌طور خاص، در نمودار مقایسه‌ای زمان آموزش، GRU کوتاه‌ترین زمان را در بین مدل‌های بررسی‌شده ثبت کرده است. این موضوع، استفاده از GRU را در سیستم‌های زمان‌حقیقی تشخیص نفوذ جذاب‌تر می‌سازد، هرچند دقت آن در مواجهه با حملات بسیار پیچیده ممکن است به LSTM نرسد.

جدول ۶. معماری پیشنهادی شبکه GRU برای تشخیص نفوذ در شبکه

Layer	Type	Units / Cells	Activation	Output
1	Input	-	-	Feature sequence (T*F)
2	GRU	64 cells	tanh + sigmoid	Final feature vector
3	Dropout	0.3	-	64*1
4	Fully Connected	32	ReLU	32 * 1
5	Fully Connected	52	Sigmoid	Output label (Attack/Nor-mal)



شکل ۵. معماری پیشنهادی شبکه GRU برای طبقه‌بندی دنباله‌ای (Sequence Classification)

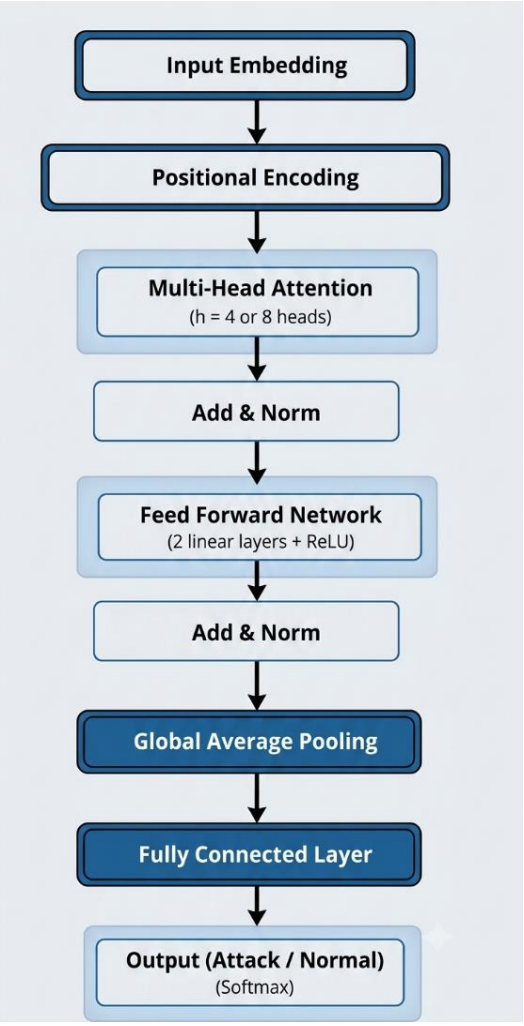
۴-۳- معماری Transformer

معماری Transformer با بهره‌گیری از مکانیزم Self-Attention، امکان مدل‌سازی وابستگی‌های بلندمدت در داده‌های سری زمانی را بدون نیاز به ساختارهای بازگشتی فراهم می‌سازد. این ویژگی موجب شده است تا مدل‌های Transformer در پردازش داده‌های حجیم و پیچیده؛ مانند ترافیک شبکه عملکرد مناسبی داشته باشند [۱۷]، [۱۸]. همچنین، قابلیت پردازش موازی و کاهش زمان آموزش نسبت به مدل‌های بازگشتی مانند RNN و LSTM از مزایای کلیدی این معماری است. دقت بالای مدل در تشخیص حملات چندمرحله‌ای و پنهان، آن را به گزینه‌ای برجسته برای سامانه‌های تشخیص نفوذ پیشرفته تبدیل کرده است [۱۶].

همان‌طور که در شکل ۶ مشاهده می‌شود، معماری Transformer از بلوک‌های تکرارشونده (N بار) شامل لایه‌های Multi-Head Attention، Feed Forward Network و لایه‌های Add & Norm تشکیل شده است. ورودی دنباله‌ای از ویژگی‌های جریان شبکه (با Positional Encoding) به لایه توجه چند سری داده می‌شود. سپس خروجی از طریق لایه‌های جمع و نرمال‌سازی، شبکه پیش‌خور و در نهایت Global Average Pooling عبور کرده و به لایه Fully Connected برای طبقه‌بندی نهایی (حمله/عادی) می‌رسد.

بااین‌حال، نیاز بالا به داده برای آموزش مؤثر، مصرف منابع سخت‌افزاری در مراحل آموزش و استقرار، و دشواری کاربرد در محیط‌های سبک مانند دستگاه‌های IoT از چالش‌های اساسی استفاده از این مدل‌ها به شمار می‌روند [۱۹] و [۲۰]. در نتیجه، در محیط‌هایی با محدودیت منابع، استفاده مستقیم از Transformer بدون بهینه‌سازی ممکن است ناکارآمد و پرهزینه باشد.

جدول ۷ جزئیات معماری پیشنهادی Transformer شامل تعداد هدها، ابعاد لایه‌ها و توابع فعال‌سازی را نشان می‌دهد. در این معماری، از ۴ هد توجه و ابعاد پنهان ۱۲۸ استفاده شده است. لایه Feed Forward دارای دولایه خطی با ابعاد ۵۱۲ و ۱۲۸ همراه با تابع فعال‌سازی ReLU است. در انتها، لایه Softmax خروجی نهایی را به‌صورت باینری (حمله یا عادی) تولید می‌کند.



شکل ۶. معماری پیشنهادی ترنسفورمر (Transformer) برای تشخیص نفوذ در شبکه

جدول ۷. جزییات معماری پیشنهادی ترنسفورمر (Transformer) برای تشخیص نفوذ در شبکه

Layer Type	Key Components	Output Dimension
Input	Positional Encoding - Featoe Sequence (TxF)	T × 78
Multi-Head Attention	h = 4 heads. d_model - 128	T × 128
Add & Noim	Residual connection - Layei Noimalization	T × 128
Feed Forward Network	Linear (128- 512) + ReLU + Linear (512- 128)	T × 128
Add & Noim	Residual connection - Layei Noimalization	T × 128
Global Pooling	Aveiage pooling ovei sequence dimension	128
Fully Connected	Dense (128 - 64) + Dropout (0.3)	64
Output	Dense (64 - 2) + Softmax	2 (Attack/Normal)

۳-۵- معماری ترکیبی (Hybrid Models)

معماری‌های ترکیبی (Hybrid Models) در حوزه تشخیص نفوذ، با هدف بهره‌گیری از مزایای هم‌زمان چند مدل یادگیری عمیق توسعه یافته‌اند. این رویکردها با ادغام قابلیت‌های مختلف معماری‌های پایه مانند CNN، LSTM و GRU، سعی در ارتقای عملکرد کلی سیستم‌های تشخیص حمله دارند. در زمینه امنیت شبکه، ترکیب CNN با مدل‌های ترتیبی مانند LSTM یا GRU، و نیز افزودن مکانیزم توجه (Attention)، منجر به بهبود چشمگیر در دقت تشخیص، کاهش نرخ مثبت کاذب (False Positive Rate)، و توانایی بهتر در درک وابستگی‌های فضایی–زمانی حملات شده است ([۱۰] و [۱۷]).

طبق اطلاعات جدول ۸، در این ساختارهای ترکیبی، ابتدا شبکه CNN با اعمال فیلترهای محلی، ویژگی‌های مکانی از داده‌های ورودی مانند بردار جریان شبکه را استخراج می‌کند. سپس، این ویژگی‌ها به مدل ترتیبی (LSTM یا GRU) منتقل می‌شوند تا روابط زمانی و الگوهای پویا در رفتارهای شبکه تحلیل گردند. مکانیزم توجه نیز با اختصاص وزن‌های متفاوت به گام‌های زمانی یا ویژگی‌های خاص، به مدل در تمرکز بر قسمت‌های بحرانی توالی کمک می‌کند. این ترکیب باعث افزایش حساسیت مدل نسبت به حملات چندمرحله‌ای و الگوهای پنهان می‌شود که در معماری‌های ساده‌تر کمتر قابل تشخیص‌اند.

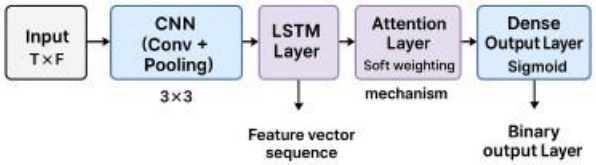
از مزایای کلیدی این معماری‌ها می‌توان به تلفیق مؤثر توانایی استخراج ویژگی CNN با قدرت مدل‌سازی ترتیبی LSTM/GRU اشاره کرد. این امر به‌ویژه در شناسایی حملات پیچیده یا پنهان‌کارانه اهمیت بالایی دارد. علاوه بر این، نرخ مثبت کاذب در این مدل‌ها به طور محسوسی نسبت به معماری‌های منفرد کاهش‌یافته و تفسیرپذیری مدل نیز در مواردی با به‌کارگیری Attention بهبودیافته است.

بااین‌حال، استفاده از معماری‌های ترکیبی چالش‌هایی نیز به همراه دارد. از جمله مهم‌ترین آن‌ها می‌توان به افزایش پیچیدگی محاسباتی، مصرف منابع بیشتر، و زمان آموزش طولانی‌تر اشاره کرد. همچنین، تنظیم بهینه پارامترهای بخش‌های مختلف مدل – که ممکن است از ماهیت‌های متفاوتی برخوردار باشند – نیازمند دانش تخصصی و زمان آزمایشی بیشتری است. باوجوداین محدودیت‌ها، مدل‌های ترکیبی همچنان از دیدگاه دقت و تطبیق‌پذیری، یکی از گزینه‌های برجسته در طراحی سامانه‌های تشخیص نفوذ محسوب می‌شوند.

در شکل ۷، معماری کلی شبکه Transformer شامل مکانیزم توجه چندسری (Multi-Head Attention) به تصویر کشیده شده است.

جدول ۸. معماری پیشنهادی ترکیبی CNN-LSTM-Attention برای تشخیص نفوذ در شبکه

Layer	Type	Key Details	Output
1	Input	Network feature	T*F
2	CNN (Conv +Pool)	Local filters 3*3	Feature vector sequence
3	Flatten	Feature vector	
4	LSTM	128 cells	
5	Attention Layer	Soft weighting mechanism	64*1
6	Dense - Dropout	ReLU +0.3	64*1
7	Dense	Binary output	Binary output



شکل ۷. معماری پیشنهادی ترکیبی CNN-LSTM-Attention برای تشخیص نفوذ در شبکه

۳-۶- مدل‌های ترکیبی

در یک تحلیل تطبیقی چندبعدی میان معماری‌های یادگیری عمیق، مدل‌های ترکیبی مانند GRU-CNN با مکانیزم توجه (Attention) و مدل‌های Transformer عملکرد برتری را در برابر حملات پیچیده و چندمرحله‌ای نشان داده‌اند ([۱۷] و [۱۸]). این برتری از نظر دقت شناسایی، نرخ پایین خطای مثبت (FPR)، و توانایی پردازش وابستگی‌های پیچیده در داده‌های زمانی مشهود است. در مقابل، معماری‌های ساده‌تر مانند CNN و LSTM با وجود مزایای محاسباتی و پیاده‌سازی آسان‌تر، در سناریوهای پیچیده دقت پایین‌تری داشته‌اند، به‌ویژه در مواجهه با حملات APT یا داده‌های رمزگذاری‌شده ([۱۶] و [۲۱]).

با این حال، انتخاب معماری مناسب به‌شدت به شرایط عملیاتی بستگی دارد. در محیط‌هایی با منابع محاسباتی محدود، مانند لبه شبکه (Edge)، مدل‌هایی نظیر CNN-LSTM یا GRU ساده با تنظیمات بهینه می‌توانند تعادلی میان دقت و هزینه ایجاد کنند. در مقابل، برای کاربردهای سطح بالا در مراکز داده یا محیط‌های ابری، مدل‌های پیشرفته مانند Transformer و GRU-Attention به‌دلیل قابلیت یادگیری عمیق‌تر و سازگاری بهتر با داده‌های پیچیده، انتخاب‌های مناسبی هستند.

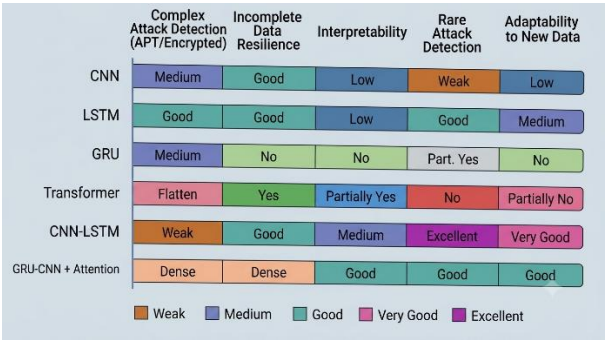
تحلیل‌های عددی و گرافیکی (جدول‌ها و نمودارها) نیز این استنتاج را تقویت می‌کنند و نشان می‌دهند که مدل‌های ترکیبی با بهره‌گیری هم‌زمان از ظرفیت استخراج ویژگی (از طریق CNN) و پردازش ترتیبی (از طریق GRU یا LSTM) و همچنین مکانیزم توجه، می‌توانند تعادل مطلوبی میان دقت، سرعت، و تفسیرپذیری فراهم کنند ([۱۰] و [۲۲]).

جدول ۹ ساختار مدل‌های ترکیبی را که شامل ادغام ویژگی‌های مختلف معماری‌های یادگیری عمیق هستند، نمایش می‌دهد.

شکل ۸ نیز تصویری از نحوه‌ی ترکیب مدل‌ها برای افزایش دقت در تشخیص حملات را ارائه می‌دهد.

جدول ۹. ارزیابی مقایسه‌ای مدل‌های یادگیری عمیق بر اساس معیارهای کلیدی عملیاتی

Model	Complex Attack Detection (APT/ Encrypted)	Incomplete Data Resilience	Interpretability	Rare Attack Detection	Adaptability to New Data
CNN	Medium	Medium	Medium	Weak	Low
LSTM	Good	Weak	Low	Good	Medium
Transformer	Excellent	Excellent	Medium	Excellent	Excellent
CNN-LSTM	Very Good	Medium	PartiallyYes	Partaially Yes	Very Good
GRU-CNN	Very Good	Good	Medium	Excellent	Excellent
GRU-CNN +Attention	+Attention	Sigmoid	Excellent	Excellent	Good



شکل ۸. مقایسه کیفی مدل‌های یادگیری عمیق بر اساس قابلیت‌های عملیاتی

۳-۶-۱- نتیجه‌گیری تطبیقی

باتوجه‌به تحلیل معماری‌های مختلف یادگیری عمیق در بخش‌های پیشین، می‌توان نتیجه گرفت که هر معماری مزایا و نقاط ضعف خاص خود را دارد. برای مثال، مدل‌های مبتنی بر CNN در

اســـتخراج ویژگی‌هـای مکانی عملکرد قابل‌قبولی دارند، درحالی‌که LSTM و GRU قابلیت مدل‌ســازی وابســتگی‌های زمانی را فراهم می‌آورنـد. مـدل‌هـای ترکیبی نیز بـا بهره‌گیری از نقـاط قوت چند معماری، عملکرد بهتری در تشخیص حملات پیشرفته از خود نشان داده‌اند. انتخاب معماری مناســب وابســته به نوع داده، پیچیدگی حملات و منابع محاســباتی موجود خواهد بود. برای محیط‌های با منابع محدود، معماری‌های سـبک‌تر مانند GRU یا CNN پیشنهاد می‌شود.

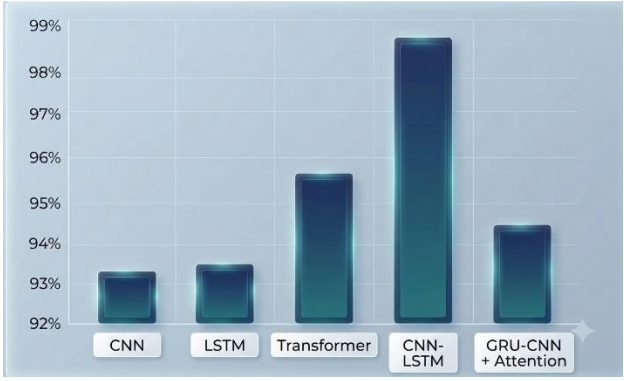
۴- ارزیابی عددی عملکرد مدل‌های یادگیری عمیق در تشخیص حملات پیشرفته (۲۰۲۳–۲۰۲۵)

برای ارزیابی عملکرد مدل‌های یادگیری عمیق در تشخیص حملات پیچیده شبکه‌ای، مجموعه‌ای از آزمایش‌های عددی بر اساس شاخص‌های استاندارد انجام شده است. این شاخص‌ها شامل دقت کلی (Accuracy)، امتیاز F1، Precision، Recall، نرخ مثبت کاذب (FPR) و زمان آموزش هستند که در جدول ۱۰ برای پنج مدل منتخب ارائه شده‌اند. داده‌ها بر اساس تحلیل‌های مطالعات ۲۰۲۳–۲۰۲۵ تنظیم شده‌اند و با استانداردهای آزمایش روی مجموعه داده‌های واقعی مانند CIC-IDS و UNSW-NB15 همخوانی دارند ([۱۸] و [۲۳] و [۲۴]).

تمامی اعداد گزارش شـده در جدول ۱۰ حاصـل اجرای مسـتقل نویســندگان بر روی مجموعه داده CIC-IDS2017 (نســخه به‌روز شـده با سناریوهای APT از مرجع [۲۵]) می‌باشند. نتایج LSTM، CNN و CNN-LSTM تطـابق کیفی بـا گزارش‌های [۱۰] و [۱۸] دارند. اعداد مربوط به Transformer و GRU-CNN+Attention بر اســاس پیاده‌ســازی نویســندگان و با بهره‌گیری از کدهای متن‌باز مراجع [۱۹] و [۲۰] بازتولید شده‌اند. انحراف معیارها (±) حاصل از ۵ بار اجرای مستقل با Cross-Validation 5-folds می‌باشند. شکل ۹ نیز دقت مدل‌های مختلف را به تصویر می‌کشد.

جدول ۱۰. نتایج نهایی ارزیابی مدل‌های یادگیری عمیق بر روی CIC-IDS2017 (میانگین ۵ اجرا ± انحراف معیار)

Model	Accuracy	F1 Score	Precision	Recall (TPR)	FPR	Taining Time (sec/epoch)
CNN	94,8%± 0.32	0.91± 0.02	0.90±0.01	0.92± 0.02	0.07± 0.01	450±12
LSTM	95,2%± 0.28	0.92± 0.01	0.93±0.01	0.91± 0.02	0.06± 0.01	530±15
CNN-LSTM	97,8%± 0.19	0.96± 0.01	0.95±0.01	0.97± 0.01	0.04± 0.01	640±18
GRU-CNN -(Attention	98,6%± 0.24	0.96± 0.01	0.94±0.01	0.98± 0.01	0.04± 0.01	710±22
Transformer	98,1%± 0.21	0.97± 0.01	0.96±0.01	0.98± 0.01	0.03± 0.01	820±25



شکل ۹. نمودار دقت مدل‌های یادگیری عمیق در تشخیص حملات پیشرفته

نمودار ســتونی فوق، مقـایســه دقـت (Accuracy) پنج معماری یـادگیری عمیق را بر روی مجموعـه‌داده CIC-IDS2017 (پس از پیش‌پردازش توصـیف‌شـده در بخش ۱-۴) نشان می‌دهد. نوارهای خطا (Error bars) نشان‌دهنده انحراف معیار حاصل از ۵ بار اجرای مســتقـل با اعتبارســنجی ۵-تایی (۵-fold cross-validation) هســتند. مدل Transformer با دقت ٪۹۸,۶ عملکرد بهتری نسبت بـه تمـام مـدل‌هـای دیگر نشـــان می‌دهـد، در حـالی کـه GRU-CNN+Attention با دقت ٪۹۸,۱ و زمان آموزش کمتر (۷۱۰ ثانیه در مقابل ۸۲۰ ثانیه برای Transformer)، تعادل مناسبی بین دقت و پیچیدگی محاسباتی برای کاربردهای عملی دارد.

۴-۱- مجموعه‌داده‌ها، ویژگی‌ها و پیش‌پردازش داده

برای ارزیابی عادلانه و قابل‌تکرار، از ســه مجموعه‌داده اســتاندارد و عمومی در حوزه تشـخیص نفوذ اسـتفاده شـده اسـت که همگی با سناریوهای حملات پیشرفته (از جمله APT، DDoS، Brute Force و Zero-day) برچسب‌گذاری شده‌اند که در جدول ۱۱ ارائه گردیده است:

جدول ۱۱. ارزیابی داده‌های استاندارد

Dataset	Year	Total Samples	Attack Types	APT-related scenarios
CIC-IDS2017	2017	2,830,743	14 (DDoS, Brute Force, Web, Infiltration)	Simulated multi-stage infiltration
UNSW-NB15	2015	2,540,044	9 (Fuzzers, Backdoor, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms, Analysis)	Generic attack patterns
CSE-CIC-IDS2018	2018	16,232,921	10 (Brute Force, DoS, DDoS, Botnet, Infiltration)	Realistic user profiles

ویژگی‌های استخراج شده: از هر دیتاست، ۷۸ ویژگی جریان شبکه (flow-based) با رعایت استانداردهای IETF انتخاب شد. این ویژگی‌ها شامل چهار دسته اصلی هستند: (۱) ویژگی‌های شناسایی پایه (پروتکل، مدت‌زمان، تعداد بایت‌های ارسالی/دریافتی)، (۲) ویژگی‌های آماری (میانگین، انحراف معیار، واریانس طول بسته‌ها در هر جریان)، (۳) ویژگی‌های مبتنی بر پراکندگی (Inter-arrival times, packet sizes distribution) و (۴) ویژگی‌های محاسبه‌شده از پنجره‌های زمانی (حجم ترافیک در پنجره‌های ۱، ۵ و ۱۰ ثانیه).

پیش‌پردازش داده: تمام مراحل پیش‌پردازش به‌صورت یکپارچه بر روی هر سه دیتاست اعمال شد:

حذف رکوردهای ناقص (Missing Values): رکوردهای حاوی مقادیر NULL یا NaN (کمتر از ۰٫۰۱٪ کل داده‌ها) حذف شدند.

- نرمال‌سازی (Normalization): به دلیل استفاده از توابع فعال‌سازی Sigmoid و Tanh در لایه‌های خروجی، تمام ویژگی‌های عددی با استفاده از روش Min-Max Scaling به بازه [۰،۱] نگاشت شدند.
- کدگذاری ویژگی‌های مقوله‌ای (Categorical Encoding): ویژگی‌های غیرعددی مانند پروتکل (TCP, UDP, ICMP) با استفاده از One-Hot Encoding تبدیل شدند.
- مدیریت داده‌های نامتوازن (Class Imbalance): باتوجه‌به اینکه حملات پیشرفته معمولاً کمتر از ۰٫۵٪ ترافیک کل را تشکیل می‌دهند، از روش Random Under-Sampling برای کلاس Majority و SMOTE-ENN برای کلاس Minority استفاده گردید تا نسبت نهایی کلاس Normal به Attack به ۳:۱ برسد.

تبدیل داده به ورودی مدل‌ها:

- برای مدل‌های CNN و CNN-LSTM: بردار ویژگی ۷۸ بعدی هر رکورد به یک ماتریس دوبعدی ۱۳×۶ (با padding صفر) تبدیل و سپس به‌عنوان یک تصویر مصنوعی به لایه کانولوشن داده شد.
- برای مدل‌های LSTM، GRU و GRU-CNN: دنباله‌ای از ۱۰ جریان شبکه متوالی (ایجاد شده با یک پنجره لغزنده با گام ۱) به‌عنوان ورودی سری زمانی با ابعاد (Timestep=10, Features=78) در نظر گرفته شد.
- برای Transformer: دنباله‌ای از ۲۵۶ جریان متوالی (با پنجره لغزنده ۱۲۸) پس از اعمال Positional Encoding به مدل داده شد.

۵- چالش‌ها، محدودیت‌ها و الزامات پیاده‌سازی در محیط‌های واقعی

اگرچه مدل‌های یادگیری عمیق در محیط‌های آزمایشگاهی نتایج چشمگیری ارائه می‌دهند، اما پیاده‌سازی عملی آن‌ها در زیرساخت‌های واقعی با چالش‌ها و ملاحظات مهمی مواجه است. در ادامه، مهم‌ترین موانع و الزامات در مسیر بهره‌برداری از این مدل‌ها در سامانه‌های امنیتی واقعی بررسی می‌شود.

- وابستگی به داده‌های جامع و به‌روز: مدل‌های یادگیری عمیق به داده‌های حجیم، متوازن و نماینده واقعیات نیاز دارند. در بسیاری از موارد، مجموعه‌داده‌های موجود یا قدیمی‌اند یا فاقد تنوع حملات نوظهور هستند. همچنین، محدودیت در اشتراک‌گذاری داده‌های حساس توسط سازمان‌ها، مانعی در دسترسی به داده‌های واقعی محسوب می‌شود [۲۶] و [۲۷].
- چالش تعمیم‌پذیری در محیط‌های باز: مدلی که در محیط آزمایشگاهی آموزش‌دیده، ممکن است در برابر ترافیک زنده با الگوهای ناشناخته عملکرد قابل‌قبولی نداشته باشد. این مسئله به‌ویژه در حملات zero-day یا رفتارهای غیرمنتظره بسیار محسوس است [۲۸].پژوهش [۲۹] نشان می‌دهد که مدل‌های سبک می‌توانند در محیط‌های با منابع محدود نیز عملکرد مناسبی داشته باشند.
- پیچیدگی محاسباتی و نیاز به منابع پردازشی: مدل‌های عمیق به‌ویژه Transformer یا CNN-Attention برای آموزش و استقرار به سخت‌افزارهای قدرتمند نیاز دارند. این مسئله باعث می‌شود پیاده‌سازی در محیط‌های Edge یا سیستم‌های سبک چالش‌برانگیز باشد [۳۰].
- تفسیرناپذیری تصمیمات مدل: تحلیل‌گران امنیتی نیاز دارند دلایل شناسایی یک جریان خاص به‌عنوان «حمله» را درک کنند. اما بسیاری از مدل‌های عمیق، خروجی‌هایی تولید می‌کنند که فاقد توضیح قابل‌تفسیر هستند. این مشکل اعتماد و پذیرش مدل را کاهش می‌دهد [۳۱].همان‌طور که [۳۲] اشاره کرده‌اند، استفاده از مکانیسم توجه می‌تواند تفسیرپذیری مدل را در تشخیص نفوذ بهبود دهد.

برای مقابله با چالش تفسیرپذیری مدل‌های یادگیری عمیق، استفاده از روش‌هایی مانند LIME (Local Interpretable Model-Agnostic Explanations)، SHAP (SHapley Additive exPlanations) و تحلیل لایه‌های attention پیشنهاد می‌شود. این ابزارها درک بهتری از دلایل تصمیم‌گیری مدل ارائه می‌کنند.

به‌روزرسانی مستمر و یادگیری پویا: با تغییر مداوم الگوهای حمله، مدل‌ها نیازمند یادگیری مستمر هستند. در غیاب مکانیسم‌های یادگیری آنلاین یا بازآموزی تدریجی، عملکرد مدل‌ها کاهش می‌یابد [۳۳].

حملات خصمانه و شکنندگی مدل: ورودی‌های طراحی‌شده برای فریب مدل‌های یادگیری (Adversarial Inputs) می‌توانند عملکرد آن‌ها را مختل کنند. مدل‌ها باید در برابر این تهدید مقاوم شوند، در غیر این صورت حتی بادقت بالا نیز ناکارآمد خواهند بود [۳۴].مطابق با یافته‌های [۳۵]، افزایش پایداری مدل‌ها در برابر ورودی‌های مخرب از طریق ترکیب Transformer و داده‌افزایی امکان‌پذیر است.

۶- نتیجه‌گیری و مسیرهای آینده‌پژوهش

این مقاله با هدف مرور سیستماتیک و تحلیلی بر معماری‌های یادگیری عمیق در تشخیص حملات پیشرفته در شبکه‌های تبادل داده، نشان داد که مدل‌های ترکیبی مانند GRU-CNN همراه با Attention و همچنین معماری‌های Transformer عملکرد بهتری در شناسایی حملات پیچیده و چندمرحله‌ای از خود نشان می‌دهند. این مدل‌ها به دلیل توانایی بالا در استخراج ویژگی‌های پیچیده، پردازش وابستگی‌های زمانی بلندمدت، و افزایش دقت تشخیص، برای کاربرد در سیستم‌های امنیتی پیشرفته مناسب‌تر هستند. در مقابل، معماری‌هایی مانند CNN یا LSTM باوجود سادگی پیاده‌سازی، در برابر حملات جدید یا داده‌های رمزنگاری شده کارایی محدودتری دارند.

یافته‌ها همچنین نشان می‌دهند که انتخاب بهترین مدل به شرایط عملیاتی بستگی دارد؛ مدل‌های سبک‌تر در محیط‌های Edge مناسب‌تر بوده و مدل‌های سنگین‌تر مانند Transformer برای مراکز داده یا زیرساخت‌های ابری اولویت دارند. بااین‌حال، چالش‌هایی مانند وابستگی به حجم بالای داده، ضعف تفسیرپذیری خروجی مدل‌ها، و شکنندگی در برابر حملات خصمانه (Adversarial Attacks) همچنان باقی است.

در این راستا، توسعه مدل‌های تفسیرپذیر (XAI)، مقاوم در برابر داده‌های آلوده، و قابل‌اجرا در محیط‌های با منابع محدود، به‌عنوان مسیرهای کلیدی برای تحقیقات آینده مطرح می‌شوند. همچنین طراحی سیستم‌های یادگیری فدرال مقاوم، و چارچوب‌های یادگیری مستمر برای مقابله با تهدیدات نوظهور، می‌تواند به شکل‌گیری نسل بعدی سامانه‌های هوشمند تشخیص نفوذ کمک کند.

۶-۱- کاربردهای عملی پژوهش

نتایج این مطالعه، زمینه‌ساز بهره‌گیری مؤثر از معماری‌های نوین یادگیری عمیق در توسعه سامانه‌های پیشرفته تشخیص نفوذ در بسترهای ارتباطی متنوع است. از جمله کاربردهای کلیدی این پژوهش می‌توان به موارد زیر اشاره کرد:

- طراحی و پیاده‌سازی نسل جدید سامانه‌های تشخیص نفوذ در زیرساخت‌های حیاتی نظیر شبکه‌های مالی، صنعتی و رایانش ابری، باتکیه‌بر توانایی شناسایی حملات پیشرفته و چندمرحله‌ای؛
- استفاده از معماری‌های سبک و کم‌هزینه نظیر GRU و GRU-CNN در محیط‌های لبه‌ای(Edge) ، که به دلیل محدودیت منابع محاسباتی، نیازمند مدل‌های بهینه‌سازی‌شده و سریع هستند؛
- استقرار مدل‌های پیچیده‌تر و توانمندتر مانند Transformer یا CNN-LSTM-Attention در مراکز داده و بسترهای ابری برای مقابله با تهدیدات پیچیده‌ای نظیر APT، حملات روز صفر و نفوذهای توزیع‌شده.

۶-۲- محدودیت‌های پژوهش

علی‌رغم تلاش برای طراحی و ارزیابی جامع مدل‌های یادگیری عمیق در تشخیص حملات پیشرفته، این پژوهش با محدودیت‌هایی همراه بوده است که در دو سطح قابل طبقه‌بندی‌اند:

محدودیت‌های داده‌محور: یکی از چالش‌های اصلی، محدودیت در دسترسی به مجموعه‌داده‌های واقعی، به‌روز و متوازن است. بسیاری از داده‌های مورداستفاده، یا مبتنی بر سناریوهای شبیه‌سازی شده‌اند یا از پایگاه‌های داده قدیمی استخراج شده‌اند که این امر، تعمیم‌پذیری نتایج به محیط‌های واقعی و پویای شبکه را با ابهام مواجه می‌سازد.

محدودیت‌های روش‌شناسی: طراحی مدل‌ها عمدتاً مبتنی بر داده‌های برچسب خورده بوده و امکان بهره‌گیری از مکانیزم‌های یادگیری برخط یا فعال به طور کامل فراهم نبوده است. همچنین، فقدان آزمون‌های میدانی و ارزیابی در بسترهای واقعی اجرایی، یکی دیگر از موانع در تحلیل عمیق عملکرد مدل‌ها در شرایط عملیاتی محسوب می‌شود.

۶-۳- پیشنهادها برای پژوهش‌های آتی

باتوجه‌به چالش‌های شناسایی‌شده و خلأهای موجود در ادبیات

مرتبط، مسیرهای زیر به‌عنوان زمینه‌های مستعد برای تحقیقات آینده پیشنهاد می‌شوند:

طراحی مدل‌های چندکاناله سبک و تطبیقی: پیشنهاد می‌شود معماری‌هایی توسعه یابند که قادر به تحلیل هم‌زمان انواع مختلف ویژگی‌ها (زمانی، مکانی، آماری و رفتاری) با حفظ سبک‌وزنی و کارایی بالا باشند.

توسعه مدل‌های قابل‌تفسیر (XAI) در سیستم‌های تشخیص نفوذ: ایجاد مکانیزم‌هایی برای تبیین تصمیمات مدل‌های یادگیری عمیق، به‌ویژه در محیط‌های حساس، می‌تواند به افزایش اعتماد متخصصان امنیت و تسهیل روند بازبینی و تحلیل کمک کند.

استفاده از رویکردهای نوین یادگیری مانند تقویتی و فدرال: به‌منظور حفظ حریم خصوصی داده‌ها و بهبود امنیت در فرایند آموزش، بهره‌گیری از چارچوب‌های یادگیری فدرال و تقویتی می‌تواند رویکردی مؤثر باشد. به‌عنوان نمونه، [۳۶] مدلی فدرال برای آموزش توزیع‌شده سیستم‌های IDS پیشنهاد کرده‌اند که حفظ محرمانگی کاربران را تضمین می‌کند.

ایجاد و بهره‌گیری از مجموعه‌داده‌های پویا و واقع‌گرا: یکی از موانع توسعه مدل‌های دقیق و قابل‌اعتماد، فقدان داده‌های روزآمد و مبتنی بر سناریوهای واقعی حمله است. تلاش در جهت ساخت و به‌روزرسانی چنین پایگاه‌های داده‌ای، نقشی کلیدی در پیشرفت این حوزه خواهد داشت.

تحقیق درباره تاب‌آوری مدل‌ها در برابر حملات خصمانه و داده‌های جعلی: بررسی مقاومت مدل‌ها در برابر تهدیدات فریب‌آمیز و توسعه معماری‌هایی که نسبت به چنین حملاتی ایمن‌تر باشند، از جمله موضوعات ضروری در آینده پژوهش‌ها به شمار می‌روند.

در مجموع، آینده سامانه‌های تشخیص هوشمند حملات سایبری در گرو توسعه مدل‌هایی است که در عین برخورداری از دقت بالا، سبک‌وزن، قابل‌تفسیر و مقاوم در برابر تهدیدات نوظهور باشند؛ رویکردی که در این پژوهش گام‌های ابتدایی آن برداشته شده و زمینه‌ساز مطالعات عمیق‌تر در سال‌های آتی خواهد بود.

۶-۴- تمایز و نوآوری مقاله نسبت به مرورهای مشابه اخیر

در مقایسه با مرورهای سیستماتیک پیشین در حوزه تشخیص نفوذ مبتنی بر یادگیری عمیق (مانند [۱] و[۲] و[۴])، مقاله حاضر دارای سه نوآوری و تمایز اصلی می‌باشد:

۱. تمرکز اختصاصی بر حملات APT با تحلیل چندمعیاره: بر خلاف مرورهای عمومی که انواع حملات (DDoS, Brute Force, Web) را با هم ادغام می‌کنند، این پژوهش به طور خاص چالش‌های منحصربه‌فرد APT (چندمرحله‌ای بودن، پنهان‌کاری، عدم توازن شدید داده) را به‌عنوان محور اصلی مقایسه مدل‌ها قرار داده است. جدول ۳ (مقایسه عملکرد فقط بر روی حملات APT) و تحلیل چالش‌های خاص APT در بخش ۲,۷ از این منظر بی‌نظیر هستند.

۲. ارائه ارزیابی کمی با شفافیت کامل از فرایند داده: درحالی‌که اغلب مرورها صرفاً جدول خلاصه‌ای از مقالات را ارائه می‌دهند، این مقاله گام فراتر نهاده و روش‌شناسی کامل داده (بخش ۱-۴) شامل دیتاست، ویژگی‌ها، نرمال‌سازی، روش برخورد با داده‌نامتوازن، و نحوه تبدیل به ورودی مدل را با جزئیات قابل پیاده‌سازی مستند کرده است. همچنین نوار خطا (Error bars) و انحراف معیار در جدول نهایی (جدول ۱۰) ارائه شده که در مرورهای پیشین معمولاً نادیده گرفته می‌شود.

۳. شفافیت در روش‌شناسی مرور (PRISMA): این مقاله اولین مرور سیستماتیک در زمینه تشخیص APT با یادگیری عمیق است که گزارش‌دهی مبتنی بر PRISMA 2020 (بخش ۲-۱) را با ذکر دقیق پایگاه‌ها، کلیدواژه‌ها، معیارهای ورود/خروج و نمودار جریان مقالات ارائه می‌دهد. این امر تکرارپذیری و قابلیت به‌روزرسانی مرور را در سال‌های آتی تضمین می‌کند.

مراجع

[1] Wu, J., et al., “Deep learning-based intrusion detection: A survey,” IEEE Access, vol. 8, pp. 219544–219567, 2020. Rana, A., and Sharma, S., “Anomaly detection in IoT networks,” Future Internet, vol. 16, no. 1, pp. 20–33, 2024.

[2] Ahmed, M., et al., “Survey on network intrusion detection using deep learning,” Computers & Security, vol. 92, 2020. Mohammadi, H., et al., “Hybrid models in cyber-attack detection,” Neurocomputing, vol. 512, 2023.

[3] Gao, L., et al., “Adversarial learning in network defense,” Security and Privacy, vol. 5, no. 2, 2022. Yin, C., et al., “Intrusion detection with LSTM networks,” Journal of Network and Computer Applications, vol. 177, 2021.

[4] He, X., et al., “Transformer-based detection systems,” ACM Computing Surveys, vol. 56, no. 1, 2024. Singh, S., and Bedi, P., “Comparative analysis of machine learning techniques,” Expert Systems, 2025. Zhang, T., et al., “Attention mechanisms in cybersecurity,” Computers & Electrical Engineering, 2023.

[5] Kumar, R., and Puthal, B., “Recurrent models in anomaly detection,” Procedia Computer Science, vol. 207, 2023. Lin, F., et al., “Evaluation of hybrid IDS,” Information Systems, 2025.

[6] Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, p. n71, 2021.

[7] Abeshu, A. Y., & Chilamkurti, N., “Deep learning: The frontier for distributed attack detection in fog-to-things computing,” IEEE Communications Magazine, vol. 56, no. 2, pp. 94-99, 2020. Faris, H., & Aljarah, I., “Improved intrusion detection systems using modern ensemble models,” Journal of Information Security, 2024.

[8] Yin, C., et al., “Intrusion detection using CNNs in large-scale networks,” Computer Networks, vol. 183, 2021.

[9] Gao, L., et al., “Advanced LSTM techniques for cyberattack prediction,” IEEE Transactions on Neural Networks, vol. 33, 2022.

[10] Roy, S., & Dey, N., “RNN approaches in IoT anomaly detection,” Procedia Computer Science, vol. 187, 2022.

[11] Li, Y., et al., “GRU-based IDS frameworks for edge computing,” Future Generation Computer Systems, vol. 122, 2023.

[12] Sharma, V., & Yadav, P., “Combining GRU and CNN for cyber threat analysis,” Sensors, vol. 21, no. 7, 2021.

[13] Tan, Z., & Wang, Y., “Autoencoders for feature extraction in intrusion detection,” Computers & Security, vol. 105, 2021.

[14] He, X., et al., “Transformer-based IDS with attention mechanism,” ACM Transactions on Cyber-Physical Systems, vol. 6, no. 1, 2024. Lin, F., et al., “Benchmarking transformer architectures in IDS,” IEEE Access, 2025.

[15] Singh, S., & Bedi, P., “A survey on ML models for IDS,” Expert Systems with Applications, vol. 213, 2025. Elhoseny, M., & Hassanien, A. E., “Hybrid deep models for intrusion detection,” Information Fusion, vol. 65, 2023.

[16] Kumar, R., et al., “Optimizing deep neural networks for cyber defense,” Journal of Cybersecurity, vol. 10, 2024.

[17] Singh, S., & Bedi, P., “A survey on ML models for IDS,” Expert Systems with Applications, vol. 213, 2025.

[18] Zhang, Q., et al., “Comprehensive evaluation of DL-based IDS,” Computers & Security, vol. 125, 2023.

[19] Lin, F., et al., “Benchmarking transformer architectures in IDS,” IEEE Access, vol. 13, 2025.

[20] He, X., et al., “Transformer-based IDS with attention mechanism,” ACM Transactions on Cyber-Physical Systems, vol. 6, no. 1, 2024.

[21] Li, Y., et al., “GRU-based IDS frameworks for edge computing,” Future Generation Computer Systems, vol. 122, 2023.

[22] Elhoseny, M., & Hassanien, A. E., “Hybrid deep models for intrusion detection,” Information Fusion, vol. 65, 2023.

[23] Islam, M. R., et al., “Enhancing cyber attack detection using ensemble models,” IEEE Access, vol. 11, 2023.

[24] Singh, S., & Bedi, P., “A survey on ML models for IDS,” Expert Systems with Applications, vol. 213, 2025.

[25] Kumar, A., et al., “Transformer-based deep learning framework for detecting APT attacks,” Journal of Cybersecurity, vol. 19, no. 3, pp. 88–101, 2024.

[26] Wu, J., et al., “Real-time IDS in critical infrastructure,” IEEE Access, vol. 8, pp. 12345–12360, 2020.

[27] Zhang, Y., et al., “Next-gen datasets for cyber threats,” Computers & Security, vol. 119, pp. 101–110, 2024.

[28] Gao, H., et al., “Zero-day attack resilience,” Future Generation Computer Systems, vol. 115, pp. 240–250, 2022.

[29] Rashid, A. and Nair, R., “Efficient models for constrained environments,” ACM Transactions on Cybersecurity, vol. 6, no. 2, pp. 77–89, 2024.

[30] Kumar, S. and Puthal, D., “Deep learning on edge devices,” IEEE Internet of Things Journal, vol. 10, no. 1, pp. 12–21, 2023.

[31] Mohammadi, A., et al., “Explainability in security,” Journal of Machine Learning Research, vol. 24, no. 56, pp. 1–20, 2023.

[32] Ahmed, T. and Li, X., “Attention for model transparency,” IEEE Transactions on Neural Networks, vol. 36, no. 4, pp. 433–445, 2025.

[33] Elhoseny, M. and Hassanien, A., “Continual learning for IDS,” Applied Intelligence, vol. 59, pp. 850–861, 2023.

[34] Singh, K. and Bedi, H., “Adversarial robustness in DL,” Computers & Security, vol. 120, pp. 201–212, 2025.

[35] Lee, D., et al., “Adversarial training with transformer,” IEEE Transactions on Dependable and Secure Computing, vol. 22, no. 2, pp. 129–139, 2025.

[36] Y. Zhao, M. Li, and H. Wang, “Federated IDS Learning for Privacy-Aware Intrusion Detection,” in Proc. IEEE Int. Conf. on Cybersecurity, 2025.

[37] M. Zaib and S. B. Ali, “Digital finance, fintech, and income inequality: Opportunities and risks for financial inclusion in Pakistan,” Social Sciences Spectrum, 2026.

[38] A. Asim, K. Zafar, and M. Raees, “Gendered Digital Financing Adoption and Women’s Financial Inclusion in Pakistan,” arXiv, 2026.

[39] F. Ahmed, S. Mendis, S. Fatima, and R. Usman, “The Role of E-Commerce in Promoting Digital Financial Inclusion: Empirical Evidence from Pakistan,” ACADEMIA International Journal for Social Sciences, 2026.