DOR: 20.1001.1.2538161.1404.10.2.6.4

DOI: 10.22034/csj.2025.227853
تابستان ۱۴۰۴، دوره ۱۰، شماره ۲، پیاپی ۳۷

دریافت مقاله: ۱۴۰۴/۰۲/۰۷
پذیرش مقاله: ۱۴۰۴/۰۳/۱۳

نوع مقاله: مروری

مروری بر سیستم‌های تشخیص نفوذ در شبکه با استفاده از مدل‌های مبتنی بر یادگیری ماشین

صدرا پورخسروانی
کارشناسی ارشد، گروه مهندسی کامپیوتر، دانشگاه بیرجند، ایران
پست الکترونیک: sadraporkhosravani@gmail.com

زهرا حیدران داروقه امنیه
گروه برق، واحد دولت آباد، دانشگاه آزاد اسلامی، اصفهان، ایران
پست الکترونیک: zahraheydaran@iau.ac.ir

مهدی حسین‌زاده هرویان
گروه برق و کامپیوتر، واحد تربت حیدریه، دانشگاه آزاد اسلامی، تربت حیدریه، ایران
پست الکترونیک: mhzh52@iau.ac.ir

ایمان ذباح*
گروه برق و کامپیوتر، واحد تربت حیدریه، دانشگاه آزاد اسلامی، تربت حیدریه، ایران
پست الکترونیک: im.zabbah@iau.ac.ir

چکیده

امروزه، شبکه‌های کامپیوتری نقش بسیار مهمی در ارتباطات و تبادل داده‌ها داشته و گسترش این شبکه‌ها شرایط مناسبی را برای حملات سایبری و نفوذ فراهم کرده است. نفوذ یکی از فعالیت‌های غیرقانونی است که امنیت اطلاعات و محرمانیت را به خطر می‌اندازد و یا دسترسی غیرمجاز به منابع سازمانی را فراهم می‌کند. سیستم‌های تشخیص نفوذ (IDS)، عامل مهمی هستند و حملاتی را که توسط باروهای[1] سنتی قابل رصد نیستند، شناسایی می‌کنند. با این حال، حملات مختلف رفتارهای خاص خود را دارند و بهبود تشخیص نوع حمله همچنان یکی از چالش‌های مدل‌های تشخیص نفوذ محسوب می‌شود. برای حل مشکلات مربوط به توسعه سیستم‌های تشخیص نفوذ (IDS)، محققان بسیاری به تمرکز بر روش‌های یادگیری ماشین و یادگیری عمیق پرداخته‌اند. این روش‌ها قادرند به طور خودکار تفاوت‌های اساسی بین داده‌های معمولی و داده‌های غیرطبیعی را کشف کنند و همچنین قابلیت تعمیم قوی دارند که این امر امکان شناسایی حملات ناشناخته را فراهم می‌کند. در این پژوهش به بررسی روش‌های مختلف تشخیص نفوذ با استفاده از یادگیری ماشین می‌پردازیم این چارچوب طبقه‌بندی برای امنیت سایبری مناسب است و محققان در این بررسی به مفهوم و طبقه‌بندی IDSها،

* نویسندهٔ مسئول

1-Firewall

الگوریتم‌های یادگیری ماشین و یادگیری عمیق، و چالش‌ها و تحولات آتی در این حوزه پرداخته‌اند.

کلید واژه‌ها: یادگیری ماشین، یادگیری عمیق، سیستم تشخیص نفوذ، امنیت سایبری

۱. مقدمه

رشد فنّاوری اینترنت حجم وسیعی از ارتباطات داده‌ها را گسترش داده است. تهدیدات و حملات شبکه و مسائل امنیتی و مهم، مشکل جامعیت و یکپارچگی داده‌ها و از دست رفتن و حذف اطلاعات را افزایش می‌دهد. سیستم‌های تشخیص نفوذ (IDS) فناوری مهمی است که در جهت تشخیص نفوذهایی که به شبکه وارد می‌شود کمک زیادی می‌نماید. به عنوان نمونه در حملات نوع DDOS، مهاجمان با ارسال حجم عظیمی از ترافیک جعلی از چندین منبع، شبکه را بار اضافی می‌کنند تا سرویس‌دهی آن مختل شود. [۱]. تشخیص نفوذ عملیاتی است که در آن فعالیت‌ها و رخدادهای یک سیستم یا شبکه پویش شده و بر اساس این جستجوها انجام نفوذ به آن شبکه یا سیستم، اطلاع رسانی می‌شود[۲]. در این حالت ممکن است محرمانگی، صحت و افشا و تمامیت و یا دسترسی‌پذیری به منابع سازمان و اطلاعات مهم دچار اختلال و حتی افشا شود. امروزه تشخیص نفوذ یکی از حیاتی‌ترین موارد امنیت اینترنت را تشکیل می‌دهد. به عبارت دقیق‌تر به‌منظور به وجود آوردن امنیت در سیستم‌های کامپیوتری ضمن استفاده از باروها و تجهیزات امنیتی دیگر، سیستم‌های تشخیص نفوذ نیز مورد نیاز می‌باشد [۳].

اولین سیستم تشخیص نفوذ در سال ۱۹۸۰ توسط جیمز اندرسون که یکی از پیشگامان این حوزه بود ارائه شد [٤]. از آن زمان، بسیاری از سیستم‌های IDS رشد و تکامل پیدا کردند. با این حال، بسیاری از IDSها هنوز از نرخ هشدار کاذب بالایی برخوردارند و تلاش برای بهبود تشخیص حملات و کاهش خطرات همچنان ادامه دارد. به طوری که بسیاری از محققان بر روی توسعه IDS با نرخ تشخیص بالاتر و کاهش نرخ هشدار کاذب در حال تلاش هستند. مشکل دیگر IDSهای موجود این است که آنها توانایی تشخیص حملات ناشناخته را ندارند. از آنجا که محیط شبکه به سرعت تغییر می‌کند، انواع حملات جدید به طور مداوم ظاهر می‌شوند. بنابراین، لازم است IDSهایی ایجاد شوند که توان شناخت حملات ناشناخته را داشته باشند.

برای حل مشکلات فوق، پژوهشگران بر روی ساخت سامانه‌های تشخیص نفوذ با استفاده از روش‌های یادگیری ماشین متمرکز شده‌اند. یادگیری ماشین یک نوع تکنیک هوش مصنوعی است که می‌تواند به صورت خودکار اطلاعات مفیدی را از مجموعه داده‌های حجیم کشف کند [٥]. سامانه‌های تشخیص نفوذ مبتنی بر یادگیری ماشین به شرط آن که داده‌های آموزش کافی در دسترس باشند و مدل‌های یادگیری ماشین دارای تعمیم‌پذیری کافی برای تشخیص نسخه‌های حمله و حملات نوآورانه باشند، عملکرد مطلوبی در تشخیص خواهند داشت. مانند حملات Zero-day که از آسیب‌پذیری‌های ناشناخته و وصله‌نشده در نرم‌افزارها، سیستم‌عامل‌ها یا سخت‌افزارها سوءاستفاده می‌کنند. از آنجا که این آسیب‌پذیری‌ها هنوز توسط توسعه‌دهندگان شناسایی و رفع نشده‌اند، هیچ راهکار امنیتی از پیش تعیین‌شده‌ای برای مقابله با آنها وجود ندارد مانند، حمله استاکس‌نت (Stuxnet) است که در سال ۲۰۱۰ با بهره‌گیری از چندین آسیب‌پذیری Zero-day در ویندوز، سیستم‌های صنعتی ایران را هدف قرار داد [٦]. علاوه بر این، سامانه‌های تشخیص نفوذ مبتنی بر یادگیری ماشین به طور زیادی به دانش حوزه وابسته نیستند؛ بنابراین، طراحی آنها چندان پیچیده نیست. یادگیری عمیق یک شاخه از یادگیری ماشین است که می‌تواند عملکردهای برجسته‌ای داشته باشد. در مقایسه با تکنیک‌های سنتی یادگیری ماشین، روش‌های یادگیری عمیق، در مقابله با داده‌های بزرگ بسیار توانمند هستند. علاوه بر این، روش‌های یادگیری عمیق می‌توانند به صورت خودکار

ویژگی را از داده‌های خام یاد بگیرند؛ یک ویژگی قابل توجه در یادگیری عمیق، ساختار عمیق است که شامل چندین لایه پنهان است. به عبارت دیگر، مدل‌های سنتی یادگیری ماشین مانند ماشین بردار پشتیبان (SVM) و نزدیکترین K همسایه (KNN) هیچ لایه پنهانی ندارند یا تنها یک لایه پنهان دارند. بنابراین، این مدل‌های سنتی

یادگیری ماشین به مدل‌های کم‌عمق نیز معروف هستند

هدف از این مطالعه، دسته‌بندی انواع سیستم‌های تشخیص نفوذ مبتنی بر یادگیری ماشین است که تا کنون پیشنهاد شده‌اند. بدین‌منظور برخی از مقالاتی که از سال ۲۰۱۹ تا ۲۰۲٤ منتشر شده‌اند مورد بررسی قرار گرفته اند. چند تحقیق قبلی نیز به عنوان بخشی از تلاش‌های تحقیقاتی بر اساس الگوریتم‌های یادگیری ماشینی که به کار گرفته شده‌اند، دسته‌بندی شده‌اند[۷]. در این پژوهش، منبع داده به عنوان مبنای اصلی دسته‌بندی سیستم‌های تشخیص نفوذ انتخاب شد، زیرا تحلیل‌ها نشان داد کیفیت و ماهیت داده‌های آموزشی مانند NSL-KDD، CIC-IDS2018 و UNSW-NB15 تأثیر مستقیمی بر عملکرد مدل‌ها دارد. مقایسه این چارچوب با دسته‌بندی سنتی (بر اساس روش تشخیص) در بخش ٤.۲ مقاله نشان داده شده است که رویکرد داده‌محور تفاوت‌های معناداری (۸٪ ± دقت) در عملکرد مدل‌ها را شناسایی می‌کند که در روش‌های کلاسیک نادیده گرفته می‌شود. برای مثال، مدل‌های ناهنجاری‌محور روی CIC-IDS2018 به دقت ۹٤ درصد رسیدند، در حالی که همین روش روی NSL-KDD تنها ۸٦ درصد دقت داشت. آزمون آماری ANOVA نیز تأثیر معنادار (p-value <۰/۰۵) منبع داده بر معیارهای ارزیابی را تأیید کرد. این چارچوب همچنین نیاز به به‌روزرسانی دوره‌ای داده‌ها و توسعه معیارهای ارزیابی خاص هر مجموعه‌داده را به عنوان راهکارهای عملیاتی پیشنهاد می‌دهد. بررسی الگوریتم‌های یادگیری ماشین مختلف که بر روی سیستم‌های تشخیص نفوذ استفاده شده‌اند، مناسب برای محققان یادگیری ماشین است. با این حال، این نوع سیستم طبقه‌بندی بیشتر بر فناوری‌های اجرای خاص تأکید می‌کند تا مشکلات حوزه امنیت سایبری. به عبارت دیگر، این تحقیق ها بعضا، به طور مستقیم به روش‌های حل مشکلات حوزه تشخیص نفوذ با استفاده از یادگیری ماشین نمی‌پردازند. لذا، در این نظرسنجی یک طبقه‌بندی جدید مرتبط با داده‌های سیستم‌های تشخیص نفوذ ارائه می‌شود و مطالعات مرتبط با این طبقه‌بندی را معرفی می‌کنیم.

بررسی مطالعات متعدد نشان می‌دهد [۸] که انواع ویژگی‌ها و روش‌های استخراج آنها در مدل‌های یادگیری ماشین متفاوت می‌باشند. بنابراین، سعی کرده‌ایم در این مطالعه داده‌های پردازش شده در امنیت سایبری و سیستم‌های تشخیص نفوذ را به صورت اجمالی دسته‌بندی کنیم. به عنوان مثال، این سیستم طبقه‌بندی می‌تواند به مشکلات زیر پاسخ دهد: (۱) چه ویژگی‌هایی بهترین نمایندگان برای حملات مختلف (مانند حملات نوع DDOS، R2L و PROBE) هستند؟ (۲) چه نوع داده‌ای برای تشخیص حملات خاص مناسب‌تر است؟ (۳) چه انواع الگوریتم‌های یادگیری ماشین برای یک نوع خاص داده مناسب‌تر هستند؟ (٤) چگونه روش‌های یادگیری ماشین بهبود سیستم‌های تشخیص نفوذ را در جنبه‌های مختلف ارتقا می‌دهند؟ در نهایت، چالش‌ها و توسعه‌های آینده روش‌های یادگیری ماشین برای سیستم‌های تشخیص نفوذ با خلاصه کردن مطالعات نماینده اخیر بحث می‌شوند. بخش بقیه این مقاله به شرح زیر است: بخش ۲ مفاهیم کلیدی و طبقه‌بندی سیستم‌های تشخیص نفوذ را معرفی می‌کند. بخش ۳ الگوریتم‌های یادگیری ماشینی متداول در سیستم‌های تشخیص نفوذ، معیارهای آنها و مجموعه داده‌های مرجع را معرفی می‌کند. بخش ٤ سیستم‌های تشخیص نفوذ را بر اساس منابع داده‌ای دسته‌بندی می‌کند و فرآیند اعمال یادگیری ماشین بر روی سیستم‌های تشخیص نفوذ را خلاصه می‌کند. بخش ٥ چالش‌ها و جهت‌های آینده سیستم‌های تشخیص نفوذ مبتنی بر

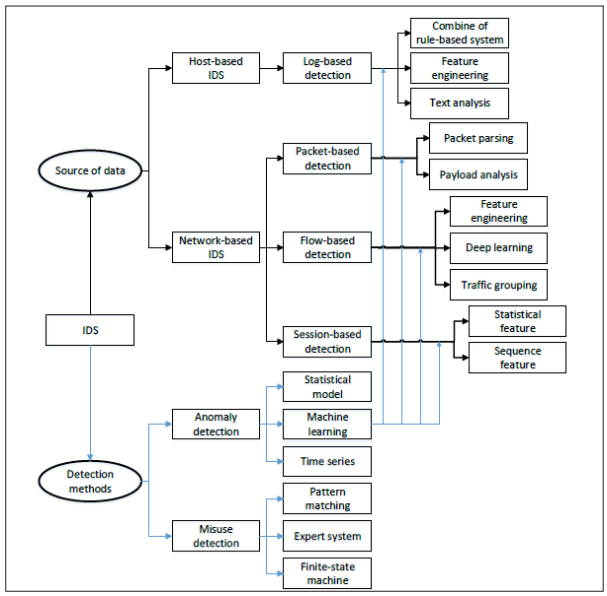
یادگیـــری ماشین را بحث می‌کند و بخش ٦ به نتیجه مقاله می‌پردازد.

۲. مفهوم و طبقه‌بندی سیســـتم‌های تشخیص نفوذ(IDS)

نفوذ بـــه معنای تلاش برای دســـترسی بـــه اطلاعات درباره سیستم‌های کامپیوتری یا تخریب عملکرد سیستم به صورت غیرقانونی یا غیرمجاز است. یک IDS یک برنامه امنیتی کامپیوتری اســـت که هدف آن تشـــخیص یک طیف گســـترده از نقض‌های امنیـــتی اســـت، از تلاش‌های نفوذ توســـط افراد خارجی تا نفوذ و سوءاستفاده‌های داخلی‌تر [۹]. عملکردهای اصلی سیســـتم‌های تشخیص نفوذ شامل نظارت بر میزبان‌ها و شـــبکه‌ها، تجزیه‌وتحلیل رفتارهای سیســـتم‌های کامپیوتـــری، تولید هشـــدارها و پاســـخ به رفتارهای مشـــکوک اســـت. به دلیل نظارت بر میزبان‌ها و شبکه‌های مرتبط، سیســـتم‌های تشخیص نفوذ معمولاً در نزدیکی گره‌های شـــبکه محافظت شده (مانند سوئیچ‌ها در بخش‌های اصلی شبکه) نصب می‌شوند.

دو نـــوع روش دســـته‌بندی بـــرای ســـامانه‌های تشـــخیص نفـــوذ (IDS) وجـــود دارد: اول روش مبتنی بر تشـــخیص(detection-based method) و دوم روش‌های مبتنی بر منبع داده(data source-based methods) [۱۱]..

در بیـــن روش‌های مبتنی بر تشـــخیص، IDSهـــا می‌توانند به تشـــخیص سوءاســـتفاده(misuse detection) [۱۲]. و تشـــخیص ناهنجاری (anomaly detection) [۱۷]. تقسیم شـــوند. در بیـــن روش‌های مبتـــنی بر منبـــع داده، IDSها می‌توانند به روش‌های مبتـــنی بر میزبان (host-based) و مبتنی بر شـــبکه (network-based)تقسیم شوند [۱۰]. در این مطالعه ســـعی می‌کنیم این دو نوع روش دســـته‌بندی IDS را ترکیـــب کرده و منبـــع داده را به عنوان اصلی‌ترین ملاک دســـته‌بندی در نظر بگیریم و روش تشخیص را به عنوان یک عنصر دسته‌بندی ثانویه معرفی می‌کنیم. شکل ۱ دسته‌بندی انواع حملات را نشان می‌دهد. درباره روش‌های تشـــخیص، مطالعه بـــر روی روش‌هـــای یادگیری ماشین متمرکز می‌شود. نحوه اســـتفاده از یادگیری ماشین برای IDS با استفاده از انواع مختلف داده به تفصیل در بخش‌های بعدی معرفی می‌شود.



شکل (۱): دسته بندی سیستم‌های تشخیص نفوذ

همان‌طور که در شـــکل ۱ نشـــان داده شـــده است، در طبقه‌بندی بر اساس روش تشخیص، تشخیص سوءاستفاده شـــامل روش‌های مبتنی بر تطبیق الگو، سیستم‌های خبره و مبتنی بر ماشین‌های حالت متناهی می‌شـــود. تشخیص ناهنجاری شامل روش‌های مبتنی بر مدل آماری، مبتنی بر یادگیری ماشین و مبتنی بر سری زمانی است.

۳. طبقه‌بندی بر اســـاس روش‌های تشـــخیص (detection-based method)

در تشخیص سوءاستفاده که به آن تشخیص مبتنی بر امضا (signature-based) نیز گفته می‌شود، ایده اصلی این است که رفتارهای حملات به عنوان امضاها نمایان شوند. در این روش، فرآیند تشخیص با استفاده از یک پایگاه داده امضا، امضاهای نمونه‌ها را با هم مطابقت می‌دهد. مشـــکل اصلی در ســـاخت ســـامانه‌های تشـــخیص سوءاستفاده،

طراحی امضاهای کارآمد است [۱۱]. مزایای تشخیص سوءاستفاده در این است که نرخ اشتباه پایینی دارد و انواع حملات را به همراه دلایل ممکن به طور دقیق گزارش می‌دهد؛ اما معایب آن شامل نرخ اشتباه از دست رفته بالا، ناتوانی در تشخیص حملات ناشناخته و نیاز به نگهداری یک پایگاه داده امضا بزرگ است. ایده طراحی پشت تشخیص ناهنجار این است که یک نمایه رفتار عادی ایجاد کرده و سپس رفتارهای غیرعادی را بر اساس انحراف آنها از نمایه عادی تعریف کند. بنابراین، کلید برنامه‌ریزی یک سامانه تشخیص ناهنجار، تعریف دقیق یک نمایه عادی است. مزایای تشخیص ناهنجار شامل عمومپذیری قوی و توانایی تشخیص حملات ناشناخته است، در حالی که مشکلات آن شامل نرخ اشتباه بالا و ناتوانی در ارائه دلایل ممکن برای یک ناهنجاری هستند. تفاوت‌های اصلی میان تشخیص سوءاستفاده و تشخیص ناهنجار در جدول ۱ آورده شده‌اند.

۴. انواع داده‌های مورد استفاده در تشخیص نفوذ

آموزش مدل‌های یادگیر، نیاز به مقدار قابل توجهی داده دارد تا تقریباً همبستگی موثر که به معنای وجود رابطه معنادار و قابل اتکا بین داده‌های ورودی (Features) و خروجی‌های مورد انتظار (Labels) در فرآیند یادگیری ماشین است، به دست آید. این مسئله بخصوص در مورد یادگیرنده‌های نظارت شده (Supervised Learning) حائز اهمیت است. متأسفانه، بخش عمده‌ای از مجموعه داده‌های عمومی برای سیستم‌های تشخیص نفوذ معمولاً بسیار قدیمی است و نمایندگی ایده‌آل از شبکه را ارائه نمی‌دهد [۱۲]. رفع این مشکل ممکن است شامل جمع‌آوری دستی داده‌ها یا ایجاد نسخه‌های سفارشی از مجموعه داده‌های موجود شود. با این حال، عدم وجود یک محک مشترک برای پیاده‌سازی‌های جدید سیستم‌های تشخیص نفوذ، مقایسه روش‌ها از نظر دقت و هشدارهای مثبت زیاد را دشوار می‌کند. این نکته نشان می‌دهد که نیاز به مجموعه داده‌های استاندارد و به‌روز است تا امکان مقایسه و ارزیابی منصفانه روش‌های مختلف تشخیص نفوذ را فراهم کند. در این بخش تلاش شده است، به بررسی اجمالی از مجموعه داده‌های موجود برای آموزش شبکه‌های عصبی در ارتباط با پیاده‌سازی سیستم‌های تشخیص نفوذ ارائه شود که در مقالات مورد بحث و بررسی قرار گرفته‌اند. برخی از روش‌های سفارشی برای تولید داده‌های آموزشی نیز ذکر شده است [۱۳] تا به محققان کمک کند راه‌های جدیدی برای تأیید سیستم‌های تشخیص نفوذ خود پیدا کنند.

۴-۱. داده‌های DARPA ۱۹۹۹

DARPA ۱۹۹۹ جانشین DARPA ۱۹۹۸ است. در این مورد، داده‌ها برای مدت ۵ هفته جمع‌آوری شده است، ۳ هفته برای مجموعه آموزش و ۲ هفته برای مجموعه آزمون. تفاوت اصلی بین آن‌ها دامنه گسترده‌تری از سناریوهای حمله موجود است (Lippmann et al. 2000a). با این حال، علیرغم این که DARPA ۱۹۹۸ و DARPA ۱۹۹۹ به عنوان مجموعه‌داده‌های معمولاً استفاده شده برای آزمایش‌ها ارائه می‌شوند، در این پژوهش، ما روش‌های جدیدی که از آن‌ها استفاده کرده باشند، مشاهده نکردیم. دلیل احتمالی این موضوع این است که این مجموعه‌داده‌ها به طور کامل، قادر به شبیه‌سازی سیستم‌های فیزیکی شبکه نشده‌اند و در حال حاضر توسط پیشنهادهای جدید جایگزین می‌شوند. جدول ۱ انواع حملات [۱۴] و نوع هر حمله را در این مجموعه نشان می‌دهد. قابل ذکر است که در این پژوهش صرفا به عنوان مروری و تحلیل، بر ضرورت توسعه مجموعه‌داده‌های به‌روز تأکید شده است و ارزیابی مقایسه‌ای میان مجموعه‌های داده قدیمی و جدید در قالب آزمایش‌های تجربی، انجام نگرفته است.

۴-۲. مجموعه‌داده KDD Cup ۱۹۹۹

مجموعه داده KDD Cup ۱۹۹۹ یکی از مجموعه‌داده‌هایی است که بسیاری از محققین برای ارزیابی سیستم‌های

جدول(۱): انواع حملات و رده‌بندی آنها در مجموعه داده‌های DARPA1999

نام حمله	انواع حمله	تعداد حمله
Probe	اسکن پورت (Port Scanning)، اسکن آسیب‌پذیری (Vulnerability Scanning)، پینگ سوئیپ (Ping Sweep)، حملات Sweep (Network Sweep)، حملات OS Fingerprinting	۱۴۰۷۷
Dos	سیل SYN (SYN Flood)، سیل UDP (UDP Flood)، سیل ICMP (ICMP Flood)، حمله اسمورف (Smurf)، حمله پینگ مرگ (Ping of Death)، سیل HTTP (HTTP Flood)، حمله اسلولوریس (Slowloris)، تقویت NTP (NTP Amplification)، تقویت DNS(Amplification، DNS) حمله تیردراپ (Teardrop)	۵۳۳۸۳
R2l	حدس رمز عبور (Password Guessing)، تزریق SQL (SQL Injection)، اسکریپت‌نویسی بین‌سایتی (Cross-Site Scripting)، تزریق کد (Code Injection)، اجرای دستور از راه دور (Remote Command Execution)، سوءاستفاده از میانگیر (Buffer Overflow)، تزریق پوسته (Shell Injection)، سوءاستفاده از اعتبار (Credential Exploitation)، تزریق فایل (File Injection)، سوءاستفاده از سرویس‌های شبکه (Network Service Exploitation)	۳۷۵۶
U2r	سوءاستفاده از آسیب‌پذیری محلی (Local Privilege Escalation)، حمله تزریق کد محلی (Local Code Injection)، حمله دستکاری فایل‌های سیستمی (System File Manipulation)، حمله سوءاستفاده از سرویس‌های محلی (Local Service Exploitation)، حمله دسترسی به فایل‌های حساس (Sensitive File Access)، حمله جعل هویت سیستمی (System Identity Spoofing)، حمله تغییر تنظیمات سیستمی (System Configuration Tampering)، حمله سوءاستفاده از دسترسی‌های(Root Privilege Abuse) root، حمله دستکاری لاگ‌های سیستم (System Log Manipulation)، حمله نصب backdoor سیستمی (System Backdoor Installation)	۲۵۲

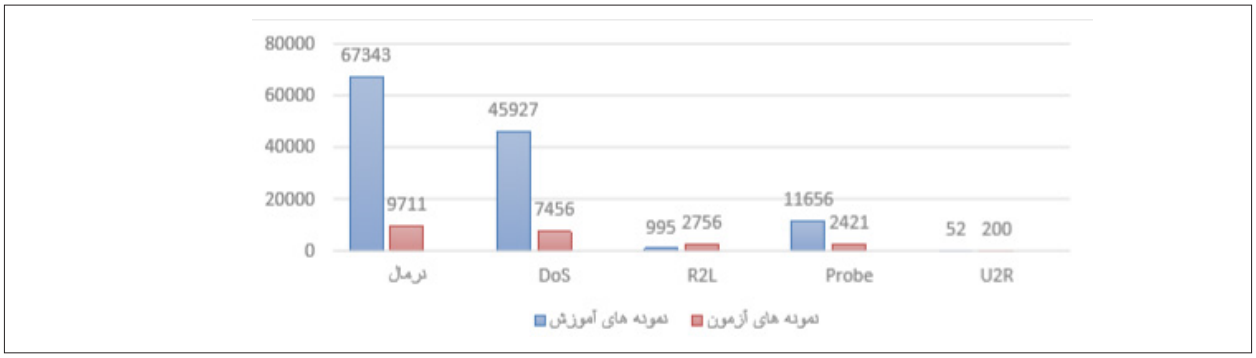
تشـــخیص نفوذ (IDSs) مورد اســـتفاده قـــرار داده‌اند. در این مطالعـــه به محدودیت‌ها و نقـــاط ضعف مجموعه‌داده KDD Cup 1999 در شبیه‌ســـازی حـــملات واقعی و نحوه بازنمایی تغییرات و حملات نوین شـــبکه‌های مدرن اشاره شده است. با این حال این مجموعه، با وجود فراگیری و رایج بودن در پژوهش‌هـــای پیشین، فاقد تطابق کامل با حملات پیشرفته و تغییرات ساختاری شبکه‌های امروزی است، که این موضوع تاثیر قابل توجهی بر کارایی و صحت مدل‌های تشـــخیص نفوذ دارد. در این مطالعه، ارزیابی یا مقایســـه عملکـــرد مدل‌های بـــر روی مجموعه‌داده‌های قدیمی مانند KDD Cup 1999 و مجموعه‌داده‌های جدیدتر انجام نشـــده است؛ اما، بر اساس پژوهش‌های مرجع، نتایج نشان می‌دهد اســـتفاده از داده‌های جدید و به‌روز، به دلیل انعکاس بهتر تغییرات شبکه و حملات نوین، می‌تواند منجر به بهبود دقت و تطابق مدل‌ها در شرایط عملی شود. با این حال با توجه به اهمیـــت این مجموعه‌داده به دلیل اســـتفاده در بسیاری از پژوهش‌هـــا در این بخـــش مورد بررسی قـــرار گرفته اســـت. این مجموعه داده از داده‌های TCP/IP مجموعه‌داده DARPA 1998 اســـتفاده می‌کند. در حالی که مجموعه داده DARPA 1998 شـــامل حدود ۵ میلیون رکورد در داده‌های آمـــوزشی و حدود ۲ میلیون رکـــورد در داده‌های آزمون اســـت، قســـمت آموزشی مجموعـــه داده KDD Cup 1999 حاوی حـــدود ۴،۹۰۰،۰۰۰ بردار اتصال اســـت. هر بردار دارای ۴۱ ویـــژگی اســـت و به عنوان اتصـــال عادی یا یک حمله دســـته‌بندی می‌شود. حملات در این مجموعه داده به یکی از چهار نوع حمله زیر تقسیم‌بندی می‌شوند:

Denial of service (DOS)، User to root (U2R)، Probe (probing)، و Root to local (R2L).

یکی از مشکلات این مجموعه‌داده این است که مجموعه داده KDD Cup 1999 از توزیـــع نامســـاوی حـــملات و رکوردهای تکراری برخوردار است[۱۵]. جدول ۲ انواع و کد حمله را در مجموعه‌داده KDD_cup نشان می‌دهد.

۴-۳. مجموعه‌داده NSL-KDD

NSL-KDD یک مجموعه‌داده است که برای رفع مشکلات مجموعه داده‌های DARPA و KDD Cup 1999 ایجاد شـــده



شکل(۲): تقسیم بندی نمونه‌های آموزش وآزمون در حملات پایگاه داده NSL-KDD

جدول(۲): انواع حملات و زیر حملات در پایگاه داده kdd_CUP

کد حمله	نوع حمله	تعداد حمله	کلاس حمله
۱	DOS	۲۸۰۷۹	Smurf
۲		۱۰۷۲۰۱	Neptune
۳		۲۲۰۳	Back
۴		۷۹۷	Teardrop
۵		۲۶۴	Pod
۶		۲۱	Land
۷	NORMAL	۹۷۲۷۸	Normal
۸	PROBE	۱۵۸۹	Satan
۹		۱۲۴۷	Ipsweep
۱۰		۱۰۴۰	Portsweep
۱۱		۲۳۱	Nmap
۱۲	R2L	۱۰۲۰	Warezclient
۱۳		۵۳	Guess_Pass
۱۴		۲۰	Warezmaster
۱۵		۱۲	Imap
۱۶		۸	FTP_Write
۱۷		۷	Multihop
۱۸		۴	Phf
۱۹		۲	Spy
۲۰	U2R	۳۰	Buffer_Overflow
۲۱		۱۰	Rootkit
۲۲		۹	Loadmodule
۲۳		۳	Perl

است. این پیشنهادی از طرف Tavallaee و همکاران (۲۰۰۹) بود. مزایای اصلی نسبت به KDD Cup 1999 عبارتند از :

عدم وجود رکوردهای تکراری در مجموعه‌داده آموزش و عدم وجود رکوردهای تکراری در مجموعه‌داده آزمون.

تعداد رکوردها قابل قبول است، بنابراین نیازی به ایجاد زیرمجموعه‌هایی از مجموعه‌داده برای آزمایش‌ها نیست.

تعداد معکوس رکوردهای خاص از هر گروه سطح دشواری به درصد رکوردها در مجموعه‌داده اصلی KDD Cup 1999 متناسب است.

از پایگاه داده NSL-KDD می‌توان برای ارزیابی موثر سیستم‌های تشخیص نفوذ استفاده کرد [۱۶] شکل ۲ تقسیم بندی داده ها در این مجموعه‌داده از نظر تعداد نمونه‌های آموزش و آزمون را نشان می‌دهد.

اگرچه داده‌های جدیدتر مانند CICIDS2017 یا UNSW–NB15 به‌روزتر و واقعی‌تر هستند، اما NSL-KDD هنوز به‌دلیل سادگی، مقایسه‌پذیری با کارهای گذشته و مناسب بودن برای آزمایش‌های مقدماتی، در بسیاری از تحقیقات ترجیح داده می‌شود.

۴-۴. پایگاه داده UNSW-NB15

استفاده گسترده از مجموعه داده‌های KDD Cup 1999 و NSL-KDD منجر به کشف چالش‌های زیر شد:

- عدم وجود برخی ویژگی‌های حمله با تأثیر کم،
- عدم وجود برخی از الگوهای ترافیک (مانند عادی و مدرن)،
- عدم تطابق در توزیع مجموعه داده‌های خاص (آموزشی در مقایسه با آزمونی)[۱۷]

UNSW-NB15 به عنوان پاسخ به مشکلات فوق ایجاد شد. این مجموعه داده با استفاده از ابزار IXIA Perfect-Storm در آزمایشگاه محدوده سایبری مرکز امنیت سایبری استرالیا (ACCS) ایجاد شد UNSW-NB15 شامل ۴۹ ویژگی است. این پایگاه داده شامل ۱۰

جدول (۳): ویژگی‌های پایگاه داده UNSW-NB15

ردیف	نام ویژگی	توضیح ویژگی	۲۱	swin	اندازه پنجره ارسالی
۱	attack_cat	دسته‌بندی نوع حمله	۲۲	stcpb	اندازه بافر TCP ارسالی
۲	dur	مدت زمان اتصال	۲۳	dtcpb	اندازه بافر TCP دریافتی
۳	proto	پروتکل شبکه (مانند TCP، UDP)	۲۴	dwin	اندازه پنجره دریافتی
۴	service	سرویس شبکه مورد استفاده	۲۵	tcprtt	زمان رفت و برگشت TCP
۵	state	وضعیت اتصال	۲۶	synack	زمان بین SYN و ACK
۶	spkts	تعداد بسته‌های ارسالی	۲۷	ackdat	زمان بین ACK و DATA
۷	dpkts	تعداد بسته‌های دریافتی	۲۸	smean	میانگین اندازه بسته‌های ارسالی
۸	sbytes	حجم داده ارسالی (بر حسب بایت)	۲۹	dmean	میانگین اندازه بسته‌های دریافتی
۹	dbytes	حجم داده دریافتی (بر حسب بایت)	۳۰	trans_depth	عمق تراکنش
۱۰	rate	نرخ انتقال داده	۳۱	response_body_len	طول بدنه پاسخ
۱۱	sttl	زمان حیات (TTL) بسته‌های ارسالی	۳۲	ct_srv_src	تعداد اتصالات از سرویس به منبع
۱۲	dttl	زمان حیات (TTL) بسته‌های دریافتی	۳۳	ct_state_ttl	تعداد اتصالات بر اساس حالت و TTL
۱۳	sload	نرخ بارگذاری ارسالی	۳۴	ct_dst_ltm	تعداد اتصالات به مقصد در زمان اخیر
۱۴	dload	نرخ بارگذاری دریافتی	۳۵	ct_src_dport_ltm	تعداد اتصالات از مبدا به پورت مقصد
۱۵	sloss	تعداد بسته‌های ارسالی از دست رفته	۳۶	ct_dst_sport_ltm	تعداد اتصالات از مقصد به پورت مبدا
۱۶	dloss	تعداد بسته‌های دریافتی از دست رفته	۳۷	ct_dst_src_ltm	تعداد اتصالات بین مقصد و مبدا در
۱۷	simpkt	میانگین اندازه بسته‌های ارسالی	۳۸	is_ftp_login	آیا اتصال FTP لاگین دارد؟
۱۸	dimpkt	میانگین اندازه بسته‌های دریافتی	۳۹	ct_ftp_cmd	تعداد دستورات FTP در اتصال
۱۹	sjit	تغییرات زمانی در بسته‌های ارسالی	۴۰	ct_flw_http_mthd	تعداد روش‌های HTTP در جریان
۲۰	djit	تغییرات زمانی در بسته‌های دریافتی	۴۱	ct_src_ltm	تعداد اتصالات از مبدا در زمان اخیر

ردهٔ مختلف است که می‌توان برچسب چند رده‌ای برای آن در نظر گرفت. انواع حملات این پایگاه داده عبارتند از: Fuzzers، Analysis، Backdoors، DoS، Exploits، Generic، Reconnaissance، Shellcode، Worms

ویژگی‌های حملات در این پایگاه داده در جدول ۳ نشان داده شده است.

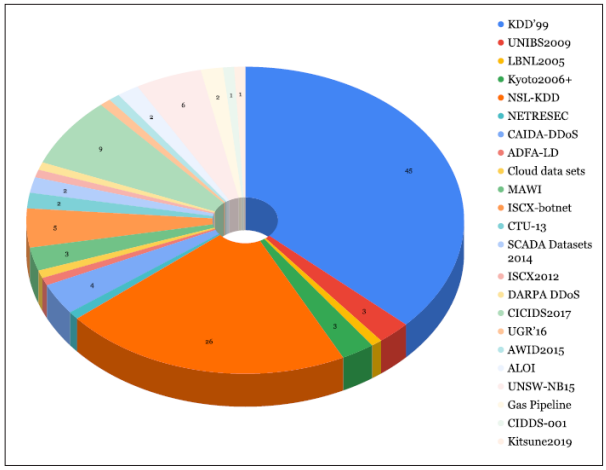
۴-۵. Kyoto2006+

یکی دیگر از داده‌های مرجع عمومی برای آموزش و آزمون سیستم‌های تشخیص نفوذ (IDS) مجموعه‌داده Kyoto2006+ است. این مجموعه‌داده شامل ۲۴ ویژگی مختلف مرتبط با شبکه است که از کارسازهای نصب شده در دانشگاه کیوتو استخراج شده است [۱۸]. ۱۴ ویژگی از KDD Cup 99 به دست آمده‌اند، در حالی که ۱۰ ویژگی دیگر به‌تازگی اضافه شده‌اند. داده‌ها برای سه سال از سال ۲۰۰۶ تا ۲۰۰۹ جمع‌آوری شده است. که به عنوان یک جایگزین برای KDD Cup 1999 ایجاد شده است. همچنین نسخه مرجع دارای ۱۷ ویژگی است که شامل ۱۴ ویژگی مشتق شده از KDD Cup 1999 و ۳ ویژگی اضافی است (مجموعه داده کیوتو ۲۰۱۵).

۴-۶. سایر پایگاه داده‌ها

بعضی از محققان تصمیم گرفتند که با مجموعه‌داده‌های عمومی دیگری به جز مجموعه‌داده‌های عمومی فوق‌الذکر آزمایش کنند. ارفانی و همکاران (۲۰۱۶) از شش مجموعه‌داده واقعی و دو مجموعه‌داده مصنوعی استفاده

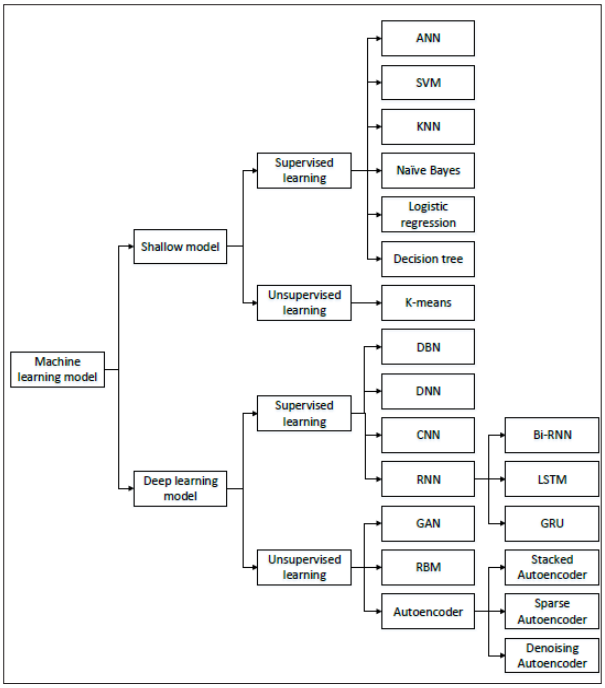
کردند. شش مجموعه‌داده واقعی از مخزن یادگیری ماشین UCI دریافت شده بودند. در هر صورت پایگاه داده‌های دیگری نیز در حوزه امنیت سایبری و سایر حوزه‌های علوم کامپیوتر وجود دارد. اما به دلیل کمبود این مجموعه‌داده‌ها،



شکل (۳): انواع پایگاه داده‌های مورد استفاده در سیستم‌های تشخیص نفوذ[۱۹].

۵. انواع روش‌های تشخیص نفوذ

دو روش اصلی از یادگیری ماشین وجود دارد: یادگیری نظارت‌شده[۲] و یادگیری بدون نظارت[۳]. یادگیری نظارت شده بر روی اطلاعات مفید موجود در داده‌های برچسب‌خورده تکیه دارد. مدل‌های طبقه‌بندی معمولا رایج‌ترین مدل‌ها در یادگیری‌های نظارت شده است که در سیستم‌های IDS بیشتر استفاده قرار می‌گیرد [۲۰]. اما برچسب‌گذاری داده به صورت دستی گران‌قیمت و زمان‌بر است. بنابراین، کمبود داده‌های برچسب‌خورده کافی، اصلی‌ترین مانع یادگیری نظارت‌شده است. برعکس، یادگیری بدون نظارت می‌تواند اطلاعات ارزشمندی را از داده‌های بدون برچسب استخراج کند. با این حال، عملکرد (ماتریس آشفتگی[۴]) تشخیصی روش‌های یادگیری بدون نظارت معمولاً پایین‌تر از روش‌های یادگیری نظارت شده است. الگوریتم‌های معمول یادگیری ماشینی استفاده شده در سیستم‌های تشخیص نفوذ به دو دسته کلی تقسیم می‌شوند. یکی روش‌های سنتی و دیگری روش‌های مبتنی بر یادگیری عمیق. شکل ٤ این دسته‌بندی‌ها را نشان داده



شکل(۴): دسته بندی کلی روش‌های تشخیص نفوذ در شبکه (روش‌های سنتی- روش‌های عمیق)

۵-۱. روش‌های سنتی مبتنی بر یادگیری ماشین مورد استفاده در تشخیص نفوذ

مدل‌های سنتی یادگیری ماشین (مدل‌های کم عمق) برای سیستم‌های تشخیص نفوذ اصلی شامل شبکه‌های عصبی مصنوعی[۵] (ANN)، ماشین بردار پشتیبان (SVM)[۶]، نزدیک‌ترین K همسایه[۷] (KNN)، بیز ساده، رگرسیون لجستیک[۸] (LR)، درخت تصمیم، خوشه‌بندی و روش‌های ترکیبی هستند. برخی از این روش‌ها برای چند دهه مورد مطالعه قرار گرفته‌اند و روش‌شان بالغ است. آن‌ها نه تنها بر روی اثر تشخیص تمرکز دارند بلکه بر مسائل عملی نیز مانند کارآیی تشخیص و مدیریت داده. مزایا و معایب مدل‌های کم عمق مختلف در جدول ٤ نشان داده شده است

مفهوم بیش‌برازش جدول (٤): روش‌های سنتی رایج

2-Supervised Learning
3-Unsupervised Learning
4-Confusion Matrix

5- Artificial Neural network
6- Support Vector Machine
7- K nearest neighbours
8- Regression Logistic

جدول (۴): روش‌های سنتی رایج در مواجه با نفوذ، ویژگی ها و معایب هر روش

نمونه مقاله تشخیص نفوذ	اقدامات بهبودی	معایب	مزایا	الگوریتم
۲۲،۲۱	بهبود بهینه‌سازها، توابع فعال‌سازی و توابع خطا	تمایل به بیش‌برازش؛ حساسیت به گیر افتادن در یک بیشینه محلی؛ زمان‌بر بودن آموزش مدل؛	توانایی مقابله با داده‌های غیرخطی؛ توانایی مناسب برای هماهنگ‌سازی قوی؛	شبکه عصبی
۲۳ ۷	پارامترهای بهینه‌سازی با بهینه‌سازی از طریق الگوریتم حشره‌شناسی (PSO) بهینه‌سازی شده است	عملکرد مناسبی در داده‌های بزرگ یا وظایف دسته‌بندی چندگانه ندارد؛ حساسیت به پارامترهای تابع هسته	یادگیری اطلاعات مفید از مجموعه داده‌های کوچک؛	بردار پشتیبان
۲۴	تعداد مقایسات را با عدم مساوی‌گری مثلثی کاهش داد؛ پارامترها با استفاده از بهینه‌سازی	دقت پایین در رده‌های کمیته؛ زمان آزمون طولانی؛ حساسیت به پارامتر K	قابلیت استفاده از داده‌های حجیم؛ مناسب برای داده‌های غیرخطی؛ آموزش سریع؛ مقاومت در برابر نوفه	نزدیک‌ترین K همسایه
۲۵ ۲۶	متغیرهای پنهان برای واگرایی از فرضیه مستقل وارد شده است	عملکرد مناسبی در داده‌های مربوط به ویژگی ندارد.	مقاومت در برابر نوفه؛ قابلیت یادگیری به صورت تدریجی	بیز ساده
۲۷	منظّم کردن برای جلوگیری از بیش‌برازش وارد شده است	عملکرد مناسبی در داده‌های غیرخطی ندارد؛ تمایل به بیش‌برازش	ساده، قابل آموزش سریع؛ مقیاس‌پذیری خودکار ویژگی‌ها	لجستیک رگرسیون
۲۸	مجموعه داده‌ها با استفاده از SMOTE متعادل شده؛ متغیرهای پنهان وارد شده‌اند	نتایج دسته‌بندی به جمعیت عمده متمایل است؛ همبستگی داده‌ها را نادیده می‌گیرد	انتخاب خودکار ویژگی‌ها؛ تفسیر قوی	درخت تصمیم
۲۹	مقداردهی روش اولیه بهبود یافته است	عملکرد مناسبی در داده‌های غیرمحدب ندارد؛ حساسیت به مقداردهی اولیه؛ حساسیت به پارامتر K	ساده، قابل آموزش سریع؛ مقیاس‌پذیری قوی؛ قابلیت تطبیق با داده‌های بزرگ	

در مواجه با نفوذ، ویژگی ها و معایب هر روش ۹ که در جدول فوق بدان اشاره شده است و امری محتمل در بسیاری از روش‌های تشخیص نفوذ می‌باشد، این است که مدل به طور بیش از حد در داده‌های آموزشی متمرکز شده و جزئیات و نوسانات خاص آن داده‌ها را یاد می‌گیرد، به گونه‌ای که عملکرد آن بر روی داده‌های آزمون کاهش می‌یابد. در نتیجه، مدل در داده‌های آموزش عملکرد خوبی دارد ولی در داده‌های ناآشنا و خارج از نمونه، رفتار ضعیف‌تری نشان می‌دهد.

۲-۵. روش‌های عمیق مورد استفاده در تشخیص نفوذ

مدل‌های یادگیری عمیق شامل شبکه‌های عمیق متنوعی هستند. از میان آن‌ها، شبکه‌های عمیق برای آموزش سریع (DBNs)، شبکه‌های عصبی عمیق (DNNs)، شبکه‌های عصبی هم‌آمیختی (CNNs) و شبکه‌های عصبی بازگشتی (RNNs) مدل‌های یادگیری نظارت‌شده هستند، در حالی که خودرمزگذارها، ماشین‌های بولتزمن محدود (RBMs) و شبکه‌های مولّد هم‌آوردی (GANs) مدل‌های یادگیری بدون نظارت هستند. تعداد مطالعات مبتنی بر یادگیری عمیق در زمینه سیستم‌های تشخیص نفوذ (IDSs) از سال ۲۰۱۵ به حال به طور سریع افزایش یافته است. مدل‌های یادگیری عمیق به طور مستقیم نمایش ویژگی‌ها را از داده‌های اصلی مانند تصاویر و متون یاد می‌گیرند، بدون نیاز به مهندسی دستی ویژگی برای مجموعه داده‌های بزرگ، روش‌های یادگیری عمیق مزیت قابل‌توجهی نسبت به مدل‌های سنتی (کم عمق)

9-Overfitting

دارند. در مطالعه یادگیـــری عمیق، تاکید اصلی بر روی معماری شبکه، انتخاب اَبَرپارامتر و راهبرد بهینه‌سازی اســـت. مقایسه‌ای از مدل‌های یادگیری عمیق مختلف در جدول ۵ نشان داده شده است.

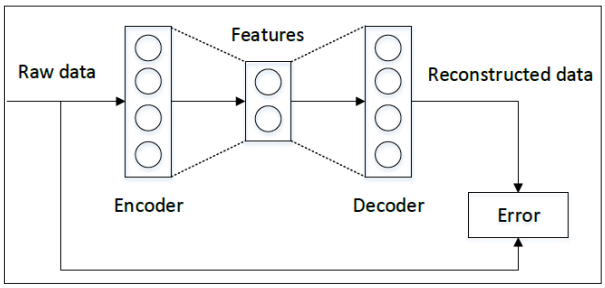
جدول (۵): روش‌های مبتنی بر یادگیری عمیق در سیستم‌های IDS

توابع	وضعیت نظارت	الگوریتم‌های مناسب برای انواع داده‌ها	الگوریتم
استخراج ویژگی؛ کاهش ویژگی؛ حذف نوفه	بدون نظارت	داده‌های خام؛ بردارهای ویژگی	Autoen-coder
استخراج ویژگی؛ کاهش ویژگی؛ حذف نوفه	بدون نظارت	بردارهای ویژگی	RBM
استخراج ویژگی؛ طبقه‌بندی	نظارت شده	بردارهای ویژگی	DBN
استخراج ویژگی؛ طبقه‌بندی	نظارت شده	بردارهای ویژگی	DNN
استخراج ویژگی؛ طبقه‌بندی	نظارت شده	داده‌های خام؛ بردارهای ویژگی	CNN
استخراج ویژگی؛ طبقه‌بندی	نظارت شده	داده‌های خام؛ بردارهای ویژگی؛ داده‌های دنباله‌ای	RNN
افزایش داده؛ آموزش هماورد مولد	بدون نظارت	نظارت‌شده، استخراج ویژگی	GAN

با توجه به اهمیـــت روش‌های عمیق در سیســـتم‌های تشخیص نفوذ در ادامه به بررسی آنها می‌پردازیم.

۲-۲-۵. یادگیرهای عمیق خودرمزگذار[10]

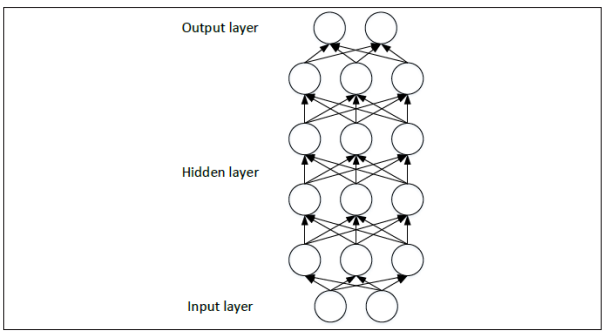
شـــامل دو بخش متقارن، یک رمزگذار و یک رمزگشا می‌باشـــد، همان‌طور که در شکل ۵ نشان داده شده است. رمزگـــذار ویژگی‌هـــا را از داده خـــام اســـتخراج می‌کند و رمزگشـــا داده را از ویژگی‌های اســـتخراج شده بازسازی می‌کنـــد. در طول آموزش، اختلاف بیـــن ورودی رمزگذار و خروجی رمزگشـــا به تدریج کاهش می‌یابد. هنگامی که رمزگشـــا موفق به بازســـازی داده از طریـــق ویژگی‌های استخراج شده می‌شـــود، به این معنی است که ویژگی‌های استخراج شـــده توسط رمزگذار بیانگر جوهر داده هستند. بسیاری از نســـخه‌های بهبود یافته مشـــهور GAN وجود دارد، کـــه در تشـــخیص نفوذ بـــه کار رفته‌انـــد [۳۰]. علت استفاده در تشخیص نفوذ در این شبکه ها وجود داده‌های نامتوازن در برخی از انواع حمله اســـت. در این خصوص می‌توان به مطالعات [۴۰،۲۱] اشاره کرد.



شکل (۵): ساختار شبکه‌های عمیق خودرمزگذار

۲-۲-۵. شبکه عصبی عمیق (DNN)[11]

یک اســـتراتژی آموزش قبلی لایه به لایه و تنظیم دقیق امکان ســـاخت شـــبکه‌های عصبی عمیق با چندین لایه را فراهم می‌کند، همانطور که در شـــکل ٦ نشـــان داده شـــده اســـت. هنگام آموزش یک DNN، پارامترها ابتدا با استفاده از داده‌هـــای بدون برچســـب یاد گرفته می‌شـــوند که یک مرحله یادگیری ویژگی بدون نظارت است. سپس، شبکه از طریق داده‌های دارای برچسب تنظیم می‌شود که یک مرحله یادگیری نظارت شـــده است. دســـتاوردهای شگفت‌انگیز شـــبکه‌های عصبی عمیق به طور اصلی به مرحله یادگیری ویژگی بدون نظارت برمی‌گـــردد. مراجع [۳۱،۳۲] از جمله تحقیقاتی هســـتند که به بررسی این شبکه و توانایی ان در تشخیص نفوذ پرداختند. شـــکل زیر ساختار این شبکه را نشان می‌دهد.



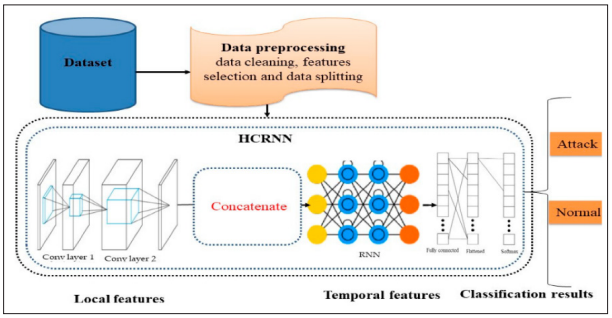
شکل (۶): ساختار شبکه‌های عمیق DNN

10-Autoencoder Deep

11-Deep Neural Network

۳-۲-۵. شبکه عصبی هم‌آمیختی (CNN) [۱۲]

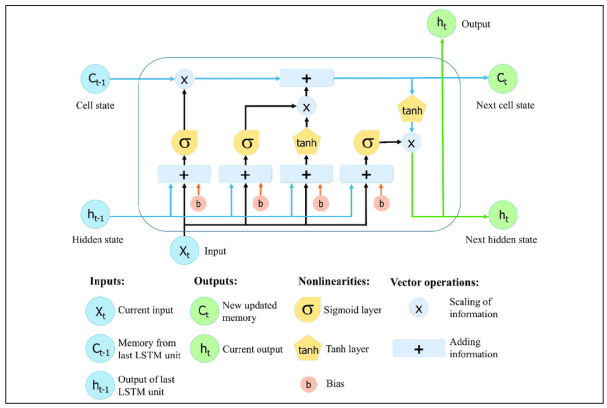
شبکه‌های عمیق هم‌آمیختی، طراحی شده‌اند تا سیستم بصری انسان (HVS) را تقلید کنند؛ بنابراین، CNNها در زمینه بینایی ماشین دستاوردهای بزرگی کسب کرده‌اند [۳۳]. یک CNN با لایه‌های هم‌آمیختی و لایه‌های یکپارچه‌ساز به صورت ترکیب شده استوانه‌ای است که در شکل ۷ نشان داده شده است. لایه‌های هم‌آمیختی برای استخراج ویژگی‌ها استفاده می‌شوند و لایه‌های یکپارچه‌ساز برای افزایش قابلیت تعمیم ویژگی‌ها استفاده می‌شوند. CNNها بر روی داده‌های دو بعدی (2D) کار می‌کنند، بنابراین داده ورودی باید به صورت ماتریسی برای شناسایی حملات ترجمه شود. شکل زیر برگرفته از مرجع [۳۴] است که به بررسی تشخیص نفوذ در دو کلاس نرمال و حمله با استفاده از یادگیرهای عمیق CNN می‌پردازد. در پژوهش مذکور، از یک شبکه عصبی هم‌آمیختی (CNN) برای ایجاد یک چارچوب شناسایی ترکیبی مبتنی بر DL استفاده می‌شود که حملات سایبری مخرب را در شبکه پیش‌بینی و طبقه‌بندی می‌کند.



شکل (۷): تشخیص نفوذ با استفاده از شبکه عمیق CNN [۳۴]

۴-۲-۵. شبکه عصبی بازگشتی (RNN) [۱۳]

RNNها شبکه‌هایی هستند که برای داده‌های متوالی طراحی شده‌اند و به طور گسترده در پردازش زبان طبیعی (NLP) استفاده می‌شوند [۳۵]. این شبکه‌ها عموما برای مسائلی استفاده می‌شوند که ویژگی‌های داده‌های متوالی در آنها وجود دارد؛ تجزیه‌وتحلیل داده‌های منزوی از دنباله معنی ندارد. برای به دست آوردن اطلاعات زمینه‌ای، هر واحد در یک RNN نه تنها وضعیت فعلی بلکه وضعیت‌های قبلی را نیز دریافت می‌کند. ساختار یک RNN در شکل ۸ نشان داده شده است. در واقع، RNNهای استاندارد تنها با دنباله‌های محدود کار می‌کنند. برای حل مشکل وابستگی بلندمدت، بسیاری از نوع‌های مختلف RNN پیشنهاد شده‌اند، مانند حافظه کوتاه‌مدت طولانی (LSTM). پژوهش‌های متعددی در خصوص تشخیص نفوذ با استفاده از شبکه‌های LSTM انجام شده است و این مقوله مورد توجه محققین بوده است [۳۶,۳۷]. شکل ۸ شبکه عمیق LSTM را نشان می‌دهد.



شکل (۸): تشخیص نفوذ با استفاده از شبکه‌های بازگشتی [۳۶] LSTM

در مسئلهٔ تشخیص نفوذ، الگوهای حمله و رفتارهای غیر عادی معمولاً در طول زمان و بر اساس جریان‌های پیوسته داده‌ها تداوم و رابطه‌های بلندمدت دارند. به عنوان مثال، حملات پیچیده و مداومی مانند فعالیت‌های مخرب که در طول زمان تکرار یا تغییر می‌یابند، نیازمند درک وابستگی‌های زمانی بلندمدت هستند. شبکه‌های LSTM به دلیل ساختار درونی خود، قادرند این رابطه‌های طولانی مدت در داده‌های جریان پیوسته را، که ممکن است چندین ساعت یا روز ادامه داشته باشند، مدل‌سازی و درک کنند[۴۱].

۶. ارزیابی عملکرد یادگیرها

بسیاری از معیارها برای ارزیابی روش‌های یادگیری

12-Convolutional Neural Network
13-Recurrent Neural Network

ماشین استفاده می‌شود. مدل‌های بهینه با استفاده از این معیارها انتخاب می‌شوند. برای اندازه‌گیری جامع اثر تشخیص، معیارهای چندگانه اغلب به طور همزمان در تحقیقات سیستم‌های مبتنی بر شبکه استفاده می‌شود. در تحقیقات مختلف در خصوص تشخیص نفوذ، به منظور تحلیل پاسخ‌های مدل‌ها و محاسبه دقت آنها از پارامترهای مختلفی استفاده شده است. شاخص‌های دقت، حساسیت، صحت از این جمله می‌باشند. معیار حساسیت یا صحت بیانگر درصد تشخیص دادن درست مدل مثبت است. سنجش G-mean نشان‌دهنده تعادل بین کارایی هر طبقه است. سنجش‌های دقت و صحت تنها روی دقت ردهٔ درست تمرکز دارند اما سنجش G-mean کارایی هر دو رده را در نظر می‌گیرند[۴۸]. شایان ذکر است که در این پژوهش، تحلیل آماری قوی یا آزمون‌های معنی‌دار برای مقایسه این معیارها در برابر یکدیگر صورت نگرفته است، اما نتایج تجربی و نمودارهای تحلیل، مؤثر بودن G-mean در این حوزه را نشان می‌دهد. نتایج مطالعات مختلف نشان می‌دهد که معیار G-mean به خوبی تعادل بین عملکرد رده‌های مختلف را برقرار می‌کند و نسبت به معیارهای سنتی، در مواجهه با داده‌های نامتوازن، حساسیت و قابلیت نمایان‌سازی بهتر عملکرد مدل را دارا است[۵۳]. در جدول زیر معیارهای ارزیابی که در این مطالعه مورد استفاده قرار گرفته است، نشان داده شده است.

جدول (۶): سنجش‌های ارزیابی در مطالعات مختلف به منظور تشخیص نفوذ

نام سنجش	رابطه سنجش	نام سنجش	رابطه سنجش
نرخ تشخیص درست ردهٔ مثبت	$TPR = \frac{TP}{TP + FN}$	صحت	$Accuracy = \frac{TP+TN}{TN+TP+FN+FP}$
نرخ تشخیص درست ردهٔ منفی	$TNR = \frac{TN}{TN + FP}$	دقت	$precision = \frac{TP}{TP+FP}$
نرخ تشخیص غلط ردهٔ مثبت	$FPR = \frac{FP}{TN + FP}$	حساسیت	$recall = \frac{TP}{TP+FN}$
نرخ تشخیص غلط ردهٔ منفی	$FNR = \frac{FN}{TP + FN}$	G-mean	$G\ mean = \sqrt{TPR.TNR}$

در اینجا TP به مثبت‌های واقعی، FP به مثبت‌های کاذب، TN به منفی‌های واقعی و FN به منفی‌های کاذب اشاره دارد. هدف یک سیستم تشخیص نفوذ (IDS) شناختن حملات است؛ بنابراین، نمونه‌های حمله معمولاً به عنوان مثبت‌ها در نظر گرفته می‌شوند و نمونه‌های عادی معمولاً به عنوان منفی‌ها در نظر گرفته می‌شوند. در تشخیص حمله، معیارهای مورد استفاده اغلب شامل دقت، بازخوانی (یا نرخ تشخیص)، FNR (یا نرخ اشتباه گذاری) و FPR (یا نرخ اشتباه مثبت) می‌باشد.

۷. بررسی برخی از پژوهش‌های انجام شده در خصوص تشخیص نفوذ

به طور کلی رویکردهای عمیق در سال‌های اخیر مورد توجه پژوهشگران قرار گرفته است. همچنین ترکیب مدل‌های عمیق نیز به عنوان یک راه حل جهت بهبود تشخیص حملات در مقالات متعدد دیده می‌شود [۳۸]. لذا در این بخش به اختصار به چند نمونه از مدل‌های عمیق و مدل‌های ترکیبی عمیق اشاره می‌شود.

درپژوهش چوان و همکاران، یک سیستم تشخیص نفوذ براساس یادگیری عمیق با استفاده از شبکه‌های عصبی بازگشتی IDS - RNN مورد بررسی قرار گرفته است[۳۹]. در پژوهشی دیگر با استفاده از شبکه‌های عمیق هم‌آمیختی CNN یک مدل عمیق بر روی پایگاه داده KDD-NSL پیاده‌سازی شده است. این مطالعه نشان می‌دهد که عملکرد مدل IDS نسبت به عملکرد مدل‌های مبتنی بر روش‌های یادگیری ماشین سنتی و روش‌های یادگیری عمیق نوین در طبقه‌بندی چند رده‌ای برتری چشمگیری دارد[۴۰]. در مطالعه‌ای دیگر نویسندگان با ترکیب دو شبکه عمیق LSTM و شبکه CNN روش موثری را برای مسئلهٔ تشخیص نفوذ ارائه کرده‌اند. حصول دقت بالا در مجموعه داده KDD NSL بدون اعمال هرگونه تنظیم اَبَرپارامتر در مجموعه‌داده KDD NSL از دستاوردهای این پژوهش است[۴۱].

در پژوهش محمد و همکاران یک مدل یادگیری عمیق

ترکیبی برای تشخیص موثر نفوذهای شبکه بر اساس یک شبکه عصبی هم‌آمیختی (CNN) و نیز یک شبکه حافظه کوتاه مدت با وزن/کاهش یافته (WDLSTM) پیشنهاد دادند [۴۲]. ریاض و همکاران، از یک الگوریتم جدید برای انتخاب ویژگی استفاده در سیستم‌های تشخیص استفاده نمودندکه شامل میدان تصادفی شرطی و انتخاب ویژگی مبتنی بر ضریب همبستگی خطی (LCFS-CRF) و یک شبکه عصبی هم‌آمیختی (CNN) می‌باشد. مدل پیشنهادی نسبت به CNN عملکرد بهتری از نظر دقت تشخیص (۹۸/۸ درصد)، زمان آموزش کمتر (۰/۵۷ ثانیه) و زمان آزمایش (۰/۲۶ ثانیه) و نرخ هشدار نادرست داشته است[۴۳]. ترکیب شبکه‌های عمیق LSTM-CNN نیز در برخی از مطالعات مشاهده می‌شود که نتیجه آن حصول دقت بالاتر در سیستم تشخیص نفوذ می‌باشد[۴۴].

به طور مشابه در تحقیق [۴۵]، یک سیستم IDS-DL (سیستم تشخیص نفوذ بر مبنای یادگیری عمیق) ارائه شده است که از یک شبکه عصبی هم‌آمیختی (CNN) و شبکه حافظه کوتاه‌مدت بلند (LSTM) در یک ساختار پیوندی استفاده می‌نماید. این سیستم برای استخراج ویژگی‌های فضایی و زمانی از داده‌های ترافیک شبکه به منظور تشخیص نفوذ استفاده می‌کند. نتیجه این مطالعه در آزمایش‌های چندطبقه‌بندی، IDS-DL حصول دقت ۹۸/۶۷ درصد و دقت تشخیص هر نوع حمله بیش از ۹۹/۵۰ درصد بود. به طورکلی مطالعاتی که با استفاده از ساختارهای ترکیبی به بررسی نوع حمله پرداخته‌اند یا از ترکیب نظرات خبرگان استفاده نموده‌اند ویا با استفاده از معماری‌های عمیق اقدام به استخراج ویژگی‌های برتر پرداخته‌اند [۴۹].

محققان در [۴۶]، یک روش برای ارتقای کارایی سیستم تشخیص نفوذ مورد استفاده در کارساز اینترنت اشیاء پیشنهاد داده‌اند. آنها روش‌های یادگیری عمیق را بررسی و این روش‌ها را با روش‌های سنتی مقایسه کرده‌اند. معیار سنجش آنها پارامتر F1_score بوده است که نتیجه نشان می‌دهد روش‌های یادگیری عمیق امتیاز 1F بالاتری را نسبت به مدل‌های سنتی دارند و استفاده از سیستم تشخیص نفوذ مبتنی بر مدل ترتیبی با استفاده از این روش‌های یادگیری عمیق می‌تواند به افزایش امنیت کارسازهای اینترنت اشیاء کمک کند. در [۴۷] یک چارچوب انعطاف‌پذیر بر مبنای شبکه‌های عصبی عمیق (DNN) ارائه شده است به منظور تشخیص حملات جدید در شبکه و بهبود کارایی سیستم تشخیص. این چارچوب دارای نرخ دقت بسیار بالا و نرخ اشتباهات نادرست (FAR) پایین است. در محیط اینترنت اشیاء، اندازه‌گیری تعادل بین دقت و حساسیت در تشخیص حملات یا حالات نادر اهمیت زیادی دارد. معیار F1-score، که تعادل میان Precision و Recall را نشان می‌دهد، برای ارزیابی مدل‌هایی مفید است که هم نیازمند دقت بالا هستند و هم به Sensitivity، یعنی توانایی شناسایی موارد مثبت، اهمیت می‌دهند. همچنین، نرخ خطای نادرست (FAR) یا نرخ هشدار نادرست، نشان می‌دهد چه مقدار سیستم، هشداری نادرست صادر می‌کند، که در محیط‌های حساس مانند اینترنت اشیاء، کاهش این نرخ برای جلوگیری از هشدارهای نادرست و صرف منابع اهمیت دارد. در مقایسه با معیارهایی مانند AUC یا Precision-Recall، که ویژه ارزیابی بر اساس تعدادی از مقادیر آستانه کار می‌کنند، این معیارها در تصمیم‌گیری‌های عملی راحت‌تر و مستقیم‌تر هستند و به طور خاص در سناریوهای حساس در مورد توازن میان Precision و Recall، اهمیت بیشتری دارند. به طور کلی، این چارچوب پیشنهادی برای شناسایی حملاتی که هنوز به صورت جهانی شناخته نشده‌اند، کارآمد و موثر است. جدول ذیل برخی دیگر از مطالعات این حوزه را نشان می‌دهد.

در برخی از مطالعات نیز روش‌های متعدد یادگیری ماشین مورد بررسی قرار گرفته و سپس مقایسه ای بین دقت آنها انجام شده و در نهایت بهترین روش در آن مطالعه معرفی شده است. جدول زیر نمونه ای از مطالعات را نشان می‌دهد.

جدول (۶) مقایسه برخی از مطالعات انجام شده و متد طبقه بندی

شماره مرجع	طبقه‌بند	دقت	تعداد کلاس	سال پژوهش	مجموعه داده	شماره
[۴۹]	ترکیب شبکه‌های عصبی MLP	٪۹۶/۷۱	۳	۲۰۲۰	KDD-99	۱
[۵۰]	بردار یادگیر ماشین (SVM)	٪۹۹/۹۶	۲	۲۰۱۹	KDD-99	۲
[۵۱]	درخت تصمیم (DT)و ترکیب آن با (SVM)	٪۹۹/۱۰	۳	۲۰۲۳	KDD-99	۳
[۵۲]	کلنی مورچه‌ها (Ant colony)	٪۹۴/۸۶	۳	۲۰۲۱	KDD-99	۴
[۵۳]	کاهش ابعاد با PCA و الگوریتم ژنتیک (GA)	٪۹۶/۳۷	۳	۲۰۲۱	KDD-99	۵
[۵۴]	نزدیک‌ترین K همسایه (KNN) و الگوریتم جمعی ذرات (PSO)	٪۹۲/۹۰	۲	۲۰۲۱	KDD-99	۶
[۵۵]	شبکه عصبی احتمالی (RBF)	---	۲	۲۰۲۳	Binary dataset	۷
[۵۶]	کاهش ابعاد به ۱۰ مولفه و درخت تصمیم C4.5-	٪۹۷/۸	۳	۲۰۲۰	KDD CUP and UNB ISCX	۱۱
[۵۷]	پیش پردازش و نرمال‌سازی دادها، انتخاب ویژگی و استفاده از روش SVM		Multi class	۲۰۱۶	NSL-KDD	۱۲
[۵۸]	استفاده از شبکه عصبی خود سازمانده som و مقایسه آن با درخت تصمیم J48	---	Multi class	۲۰۲۰	NSL-KDD , KDD'99	۱۳
[۵۹]	استفاده از روش k_mean ترکیب آن با سایر طبقه‌بندها	٪۹۳/۵۰	۳	۲۰۲۲	KDD-cup'99	۱۴
[۶۰]	طبقه‌بندی به روش kmean	---	Multi class	۲۰۲۰	KDD-cup'99	۱۵
[۶۱]	استفاده از الگوریتم CNN برای طبقه‌بندی.	---	۵	۲۰۲۰	KDD-cup'99	۱۶
[۶۲]	روش ترکیبی فازی برای استخراج خودکار قوانین	--	۵	۲۰۲۱	KDD-cup'99	۱۷
[۶۳]	ترکیب روش عصبی و فازی (FNN)	٪۹۳	۵	۲۰۲۱	KDD-cup'99	۱۸
[۶۴]	استفاده از الگوریتم‌های ترکیبی CNN و RNN برای طبقه‌بندی داده‌ها	٪۹۴	۲	۲۰۱۹	KDD-cup'99	۱۹

۸. جمع‌بندی

در این پژوهش، با هدف مرور مطالعات مختلف در حوزه تشـــخیص نفوذ با اســـتفاده از مدل‌های یادگیری ماشین، تلاش شـــد تا ابتدا یک طبقه‌بندی از سیستم‌های تشخیص نفوذ (IDS) معرفی و ســـپس الگوریتم‌های یادگیری ماشین متعددی که در این حوزه استفاده می‌شوند معرفی گردند. بر اساس این طبقه‌بندی، منابع داده متفاوتی که در مطالعات محققین مورد توجه بوده است دسته‌بندی و به طور اجمالی بررسی گردید. حملات متعددی که در هر یک از این پایگاه داده‌ها وجود داشـــت مانند حملات U2L و R2L و یا حملات DOS و Probe مـــورد بررسی قـــرار گرفت. در ادامه برای IDSهایی که از این انواع داده‌های مختلف استفاده می‌کنند، تکنیک‌های یادگیری ماشین (به ویژه الگوریتم‌های یادگیری عمیق) و ســـناریوهای کاربردی بررسی گردید. در مرور مطالعات انجام شده، بیشـــتر تحقیقات بر توسعه و بهبود دقت و کارایی مدل‌های عمیق تمرکز داشته‌اند، اما در برخی از مطالعات تمرکز بر روی زمان تشـــخیص وجود داشت و راهکارهای خاص برای کاهش زمان اجرا نظیر استفاده از مدل‌هایی که بر روی سخت‌افزارهای تسریع‌کننده مانند GPU یا TPU، ارائه شـــده بود، مانند مطالعه [۸۰] که در آن مطالعه چارچوبی برای افزایش سرعت پردازش بسته‌ها با استفاده از GPUها ارائه شـــده است. و با مطالعه موردی، مزایای این رویکرد را در اجرای سیســـتم تشـــخیص نفوذ (IDS) تحلیل می‌کند. نتایج نشـــان می‌دهد که با استفاده از این چارچوب، ســـرعت سیســـتم از ۵۰ مگابیت بر ثانیه به ۱ گیگابیت بـــر ثانیه افزایش یافته و مصـــرف منابع CPU

جدول (۷) مقایسه برخی از مطالعات ترکیبی و بهترین دقت به دست آمده

شماره مقاله	بهترین نتایج به دست آمده	بهترین الگوریتم	الگوریتم نظارت شده	پایگاه داده
۶۵	% Accuracy = 99.7 % FPR = 0.005	Random forest	J48, Random forest	NSL-KDD
۶۶	% Accuracy = 99.9 FPR = 0.001%	Random forest	Naive Bayes, Bayes-Net, Logistics, Random forest, J48, Bagging, OneR, PART, ZERO	NSL-KDD
۶۷	% Accuracy = 99 FPR= NA	Random forest	SVM, GMM, Random forest, logistics regression	NSL-KDD
۶۸	Accuracy = 94.0% FPR=NA	ANN	Artificial neural networks (ANN), SVM	NSL-KDD
۶۹	Accuracy = 99.0% FPR=NA	Random forest	k-means, random forest, naïve bayes, SVM	KDD'99
۷۰	Accuracy=98.7% FPR=NA	SVM	SVM	KDD'99
۷۱	Accuracy=91.0% FPR=NA	BBNN	BBNN	KDD'99
۷۲	Accuracy=81.83% FPR=NA	Adaboost	Adaboost	CICIDS2017
۷۳	Accuracy=98.8% FPR=.001%	QDA	LDA, QDA, BN, RF	CICIDS2017
۷۴	Accuracy=96.4	RF	ANN, RF	CICIDS2017
۷۵	Accuracy=96.7% FPR=.13%	J48	SVM, J48, RF, Zero	UNSW-NB15
۷۶	Accuracy=97.0% FPR=NA	DNN	DNN, RF	UNSW-NB15

تقریباً ٤٠ درصد کاهش یافته اســت. ویا در پژوهش [۸۱] که یک رویکرد ECU مبتنی بر FPGA ترکیبی پیشنهاد شده اســت که می‌تواند عملکرد IDS را از طریق یک شتاب‌دهنده سخت‌افزاری اختصاصی آماده که یک مدل تشخیص نفوذ CNN عمیق را پیاده‌سازی می‌کند، به طور شفاف ادغام کند. نتایج نشــان می‌دهد که رویکرد پیشــنهادی، دقت متوسط بیش از ۹۹ درصد را در بین مجموعه داده‌های حمله متعدد با نرخ تشخیص کاذب ۰/٦٤ درصد حاصل شده است، در حــالی که ۹٤ درصد انرژی کمتری مصــرف می‌کند و در مقایســه با پیاده‌ســازی‌های IDS روی GPUها، به کاهش ۵۱/۸ درصد در تأخیر پردازش هر پیام دست یافته است.

مدل‌های یادگیری عمیق نقشی روزافزونی در تشخیص نفوذ دارنــد و لذا در ایــن مطالعه به عنــوان یک رویکرد برجســته، بیشــتر مورد توجه قرار گرفتنــد. رویکردهای یادگیری عمیق شــامل شــبکه‌های عمیق چندگانه هستند کــه می‌توانند عملکرد IDSها را بهبود بخشــند. نســبت به مدل‌های یادگیری ماشین ســنتی، مدل‌های یادگیری عمیق توانایی تطبیق و عمومی‌ســازی قوی‌تری دارند. علاوه بر این، رویکردهای یادگیری عمیق مستقل از مهندسی ویژگی و دانش حوزه‌ای هســتند که مزیت برجســته‌ای نسبت به مدل‌های یادگیری ماشین ســنتی دارند. با این حال، زمان اجرای مدل‌های یادگیری عمیق اغلب برای برآورده کردن نیاز واقعی زمانی IDSها بسیار طولانی اســت. با خلاصه کردن مطالعات معمول اخیر، این مقاله چالش‌ها و روندهای آینده در ایــن حوزه را تجزیه‌وتحلیل کرده اســت. کمبود مجموعــه داده‌های موجود ممکن اســت بزرگترین چالش باشــد. بنابرایــن رویکردهــای یادگیری بــدون نظارت و یادگیری تدریجی چشم‌اندازهای توسعه گسترده‌ای دارند.

برای IDSاهای عملی، قابل فهم بودن امر ضروری است. زیرا مدل‌های قابل تفسیر قانع‌کننده هستند و می‌توانند کاربران را در تصمیم‌گیری هدایت کنند. قابلیت تفسیر مدل‌ها ممکن است به عنوان یک جهت تحقیقات مهم درباره IDSاها در آینده محسوب شود. در این مطالعه، ضمن بررسی مدل‌های یادگیری عمیق برای تشخیص نفوذ، اهمیت قابلیت تفسیر[۱۴] نیز مورد توجه قرار گرفته است. روش‌هایی مانند SHAP و LIME در برخی از پژوهش‌های مرورشده به کار رفته‌اند [۸۲] تا تصمیم‌های مدل‌های پیچیده مانند CNN-LSTM و خودرمزگذار را شفاف‌سازی کنند. با این حال، چالش اصلی، تضاد بین دقت و تفسیرپذیری در مدل‌های عمیق است. به عنوان راهکار، پیشنهاد می‌شود از رویکردهای ترکیبی استفاده شود؛ به این صورت که ابتدا یک مدل عمیق مبدّل برای تشخیص اولیه با دقت بالا به کار رود و سپس خروجی آن توسط یک مدل تفسیرپذیر مانند درخت تصمیم یا قوانین نمادین تحلیل شود. این روش می‌تواند هم زمان دقت بالا و شفافیت تصمیم‌گیری را برای IDSاهای عملیاتی فراهم کند. همچنین، توسعه داده‌های برچسب‌گذاری‌شده تفسیری و استفاده از یادگیری تقویتی قابل تفسیر می‌تواند به عنوان جهت‌گیری آینده این حوزه در نظر گرفته شود.

۹. مراجع

1. Ozkan-Okay, M., Samet, R., Aslan, Ö., & Gupta, D. (2021). A comprehensive systematic literature review on intrusion detection systems. IEEE Access, 9, 157727-157760.
2. Thakkar, A., & Lohiya, R. (2022). A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions. Artificial Intelligence Review, 55(1), 453-563.
3. Hande, Y., & Muddana, A. (2021). A survey on intrusion detection system for software defined networks (SDN). In Research Anthology on Artificial Intelligence Applications in Security (pp. 467-489). IGI Global.
4. Anderson, J.P. (1980). Computer Security Threat Monitoring and Surveillance; Technical Report; James P. Anderson Company: Philadelphia, PA , USA.
5. Michie, D.; Spiegelhalter, D.J.; Taylor, C. (1994). Machine Learning, Neurall and Statistical Classification; Ellis Horwood Series in Artificial Intelligence: New York, NY, USA, Volume 13.
6. Buczak, A.L., & Guven, E. A. (2015). survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun. Surv. Tutor. 18, 1153–1176. [CrossRef]
7. Mohammadi, M., Rashid, T. A., Karim, S. H. T., Aldalwie, A. H. M., Tho, Q. T., Bidaki, M., ... & Hosseinzadeh, M. (2021). A comprehensive survey and taxonomy of the SVM-based intrusion detection systems. Journal of Network and Computer Applications, 178, 102983.
8. Asharf, J., Moustafa, N., Khurshid, H., Debie, E., Haider, W., & Wahab, A. (2020). A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions. Electronics, 9(7), 1177.
9. Dina, A. S., & Manivannan, D. (2021). Intrusion detection based on machine learning techniques in computer networks. Internet of Things, 16, 100462.
10. Aldweesh, A., Derhab, A., & Emam, A. Z. (2020). Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. Knowledge-Based Systems, 189, 105124.
11. Díaz-Verdejo, J., Muñoz-Calle, J., Estepa Alonso, A., Estepa Alonso, R., & Madinabeitia, G. (2022). On the detection capabilities of signature-based intrusion detection systems in the context of web attacks. Applied Sciences, 12(2), 852.
12. Mebawondu, J. O., Alowolodu, O. D., Mebawondu, J. O., & Adetunmbi, A. O. (2020). Network intrusion detection system using supervised learning paradigm. Scientific African, 9, e00497.
13. Ashfaq RAR, Wang X-Z, Huang JZ et al. (2017). Fuzziness based semisupervised learning approach for intrusion detection system. Inf Sci 378:484–497. https ://doi.org/10.1016/j.ins.2016.04.019
14. Aljawarneh, S., Aldwairi, M., & Yassein, MB. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. J Comput Sci 25:152–160. https ://doi.org/10.1016/j.jocs.2017.03.006
15. Dhanabal, L., & Shantharajah, DSP. (2015). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. Int J Adv Res Comput Commun Eng 4:446–452
16. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, AA. (2009). A detailed analysis of the KDD CUP 99 data set. In: Symposium on Computational Intelligence for Security and Defense Applications. Pp 1–6
17. Moustafa, N., & Slay, J. (2016). The evaluation of network anomaly detection systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Inf Secur J Glob Perspect 25:18–31. https //doi.org/10.1080/19393 555.2015.11259 74
18. Ambusaidi, MA., He, X., Nanda, P., & Tan, Z. (2016).

14-Interpretability

James, N. (2021, October). Intrusion Detection System using Deep Neural Networks (DNN). In 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA) (pp. 1-6). IEEE.

32. Li, L. H., Ahmad, R., Tsai, W. C., & Sharma, A. K. (2021, January). A feature selection based dnn for intrusion detection system. In 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-8). IEEE.
33. Kim, J., Kim, J., Kim, H., Shim, M., & Choi, E. (2020). CNN-based network intrusion detection against denial-of-service attacks. Electronics, 9(6), 916.
34. Khan MA. HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System. Processes. 2021; 9(5):834. Https://doi.org/10.3390/pr9050834
35. Chaibi, N., Atmani, B., & Mokaddem, M. (2020, October). Deep learning approaches to intrusion detection: A new performance of ANN and RNN on NSL-KDD. In Proceedings of the 1st International Conference on Intelligent Systems and Pattern Recognition (pp. 45-49).
36. Boukhalfa, A., Abdellaoui, A., Hmina, N., & Chaoui, H. (2020). LSTM deep learning method for network intrusion detection system. International Journal of Electrical and Computer Engineering, 10(3), 3315.
37. Hossain, M. D., Inoue, H., Ochiai, H., Fall, D., & Kadobayashi, Y. (2020). LSTM-based intrusion detection system for in-vehicle can bus communications. IEEE Access, 8, 185489-185502.
38. Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., & Chen, J. (2020). DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system. Security and communication networks, 2020, 1-11.
39. Yin, Chuanlong, Yuefei Zhu, Jinlong Fei, & Xinzheng He. (2017). “A deep learning approach for intrusion detection using recurrent neural networks.” Ieee Access 5, pp. 21954-21961.
40. Xu, W., Jang-Jaccard, J., Singh, A., Wei, Y., & Sabrina, F. (2021). Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. IEEE Access, 9, 140136-140146.
41. Ahsan, Mostofa, & Kendall E. Nygard. (2020).”Convolutional Neural Networks with LSTM for Intrusion Detection.” In CATA, pp. 69-79.
42. Hassan, Mohammad Mehedi, Abdu Gumaei, Ahmed Alsanad, Majed Alrubaian, and Giancarlo Fortino. (2020).”A hybrid deep learning model for efficient intrusion detection in big data environment.” Information Sciences 513, pp. 386-396.
43. Riyaz, & Sannasi Ganapathy. (2020). “A deep learning approach for effective intrusion detection in wireless networks using CNN.” Soft Computing 24, pp. 17265-17278.
44. Yao, Ruizhe, Ning Wang, Zhihui Liu, Peng Chen, and Xi-

Building an intrusion detection system using a filter-based feature selection algorithm. IEEE Trans Comput 65:2986–2998. https ://doi.org/10.1109/ TC.2016.2519914

19. Yao, H., Fu, D., Zhang, P., Li, M., & Liu, Y. (2019). Msml: a novel multilevel semi-supervised machine learning framework for intrusion detection system. IEEE Int Things J. 6(2):1949–59. https://doi.org/10.1109/JIOT.2018.2873125.
20. Talaei Khoei, T., & Kaabouch, N. (2023). A Comparative Analysis of Supervised and Unsupervised Models for Detecting Attacks on the Intrusion Detection Systems. Information, 14(2), 103.
21. Alrayes, F. S., Zakariah, M., Amin, S. U., Khan, Z. I., & Helal, M. (2024). Intrusion detection in IoT systems using denoising autoencoder. IEEE Access..
22. Maithem, M., & Al-Sultany, G. A. (2021, February). Network intrusion detection system using deep neural networks. In Journal of Physics: Conference Series (Vol. 1804, No. 1, p. 012138). IOP Publishing.
23. Abd, S. N., Alsajri, M., & Ibraheem, H. R. (2020). Rao-SVM machine learning algorithm for intrusion detection system. Iraqi Journal For Computer Science and Mathematics, 1(1), 23-27.
24. Wazirali, R. (2020). An improved intrusion detection system based on KNN hyperparameter tuning and cross-validation. Arabian Journal for Science and Engineering, 45(12), 10859-10873.
25. Kurniawan, Y. I., Razi, F., Nofiyati, N., Wijayanto, B., & Hidayat, M. L. (2021). I Bayes modification for intrusion detection system classification with zero probability. Bulletin of Electrical Engineering and Informatics, 10(5), 2751-2758.
26. Razdan, S., Gupta, H., & Seth, A. (2021, April). Performance analysis of network intrusion detection systems using j48Inaive bayes algorithms. In 2021 6th International Conference for Convergence in Technology (I2CT) (pp. 1-7). IEEE.
27. Besharati, E., Naderan, M., & Namjoo, E. (2019). LR-HIDS: logistic regression host-based intrusion detection system for cloud environments. Journal of Ambient Intelligence and Humanized Computing, 10, 3669-3692.
28. Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. IEEE Access.
29. Chapagain, P., Timalsina, A., Bhandari, M., & Chitrakar, R. (2022, January). Intrusion detection based on PCA with improved K-means. In International Conference on Electrical and Electronics Engineering (pp. 13-27). Singapore: Springer Singapore.
30. Lee, J., & Park, K. (2021). GAN-based imbalanced data intrusion detection system. Personal and Ubiquitous Computing, 25, 121-128.
31. Navya, V. K., Adithi, J., Rudrawal, D., Tailor, H., &

Engineering, Kuantan, Pahang, Malaysia, 29th July 2019 (pp. 591-603). Springer Singapore.
57. Thaseen, I. S. & Kumar, C. A. (2016). “Intrusion Detection Model using fusion of chi-square feature selection and multi class SVM,” Journal of King Saud University-Computer and Information Sciences.
58. Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., & Razaque, A. (2020). Cascaded hybrid intrusion detection model based on SOM and RBF neural networks. Concurrency and Computation: Practice and Experience, 32(21), e5233.
59. Chapagain, P., Timalsina, A., Bhandari, M., & Chitrakar, R. (2022, January). Intrusion detection based on PCA with improved K-means. In International Conference on Electrical and Electronics Engineering (pp. 13-27). Singapore: Springer Singapore.
60. Xu, J., Han, D., Li, K. C., & Jiang, H. (2020). A K-means algorithm based on characteristics of density applied to network intrusion detection. Computer Science and Information Systems, 17(2), 665-687.
61. Al-Haija, Q .A. & Zein-Sabatto, S. (2020). “An efficient deeplearning-based detection and classification system for cyberattacks in IoT communication networks,” Electronics, vol. 9, pp. 2152-2178.
62. Pradeep Mohan Kumar, K., Saravanan, M., Thenmozhi, M., & Vijayakumar, K. (2021). Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks. Concurrency and Computation: Practice and Experience, 33(3), e5242.
63. Hassan Nataj Solhdar M. (2021). Using Of Neuro-Fuzzy Classifier for Intrusion Detection Systems. C4I Journal; 5 (2) :47-61
64. Roopak, M., Tian, G.Y., & Chambers, J. (2019). “Deep learning models for cyber security in IoT networks,” in 2019 IEEE 9th annual computing and communication workshop and conference (CCWC), pp. 452-457.
65. Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE access, 6, 33789-33795.
66. Malhotra, H., & Sharma, P. (2019). Intrusion Detection using Machine Learning and Feature Selection. International Journal of Computer Network & Information Security, 11(4).
67. Belavagi, M. C., & Muniyal, B. (2016). Performance evaluation of supervised machine learning algorithms for intrusion detection. Procedia Computer Science, 89, 117-123.
68. Taher, K. A., Jisan, B. M. Y., & Rahman, M. M. (2019, January). Network intrusion detection using supervised machine learning technique with feature selection. In 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (pp. 643-646). IEEE.
69. El Mourabit, Y., Bouirden, A., Toumanari, A., & Mousanjun Sheng. (2021).”Intrusion Detection System in the Advanced Metering Infrastructure: A Cross-Layer FeatureFusion CNN-LSTM-Based Approach.” Sensors 21, no. 2.
45. Sun, Pengfei, Pengju Liu, Qi Li, Chenxi Liu, Xiangling Lu, Ruochen Hao, & Jinpeng Chen. (2020).”DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system.” Security and Communication Networks.
46. Zhong, Ming, Yajin Zhou, & Gang Chen. (2021). “Sequential Model Based Intrusion Detection System for IoT Servers Using Deep Learning Methods.” Sensors 21, no. 4: 1113.
47. Zhiqiang, Liu, Lin Zhijun, Gong Ting, & Shi Yucheng. (2021).”A Three-Layer Architecture for Intelligent Intrusion Detection Using Deep Learning.” In Proceedings of Fifth International Congress on Information and Communication Technology, pp. 245- 255. Springer, Singapore.
48. Heidari, A., & Jabraeil Jamali, M. A. (2023). Internet of Things intrusion detection systems: A comprehensive review and future directions. Cluster Computing, 26(6), 3753-3780.
49. Maroosi, A., Zabbah, E., & Ataei Khabbaz, H. (2020). Network Intrusion Detection using a combination of artificial neural networks in a hierarchical manner. Electronic and Cyber Defense, 8(1), 89-99.
50. Meftah, S., Rachidi, T. & Assem, N. (2019). “Network Based Intrusion Detection Using the UNSW-NB15 Dataset,” International Journal of Computing and Digital Systems, vol. 8, pp. 477-487.
51. Zou, L., Luo, X., Zhang, Y., Yang, X., & Wang, X. (2023). HC-DTTSVM: A Network Intrusion Detection Method Based on Decision Tree Twin Support Vector Machine and Hierarchical Clustering. IEEE Access, 11, 21404-21416.
52. Nadoomi, M. A., & Sina, M. (2020). Using Ant Colony Algorithm and Pairwise Learning to Classify Attack in Intrusion Detection Systems. Journal of Communication Engineering, 9(36), 39-53.
53. Halim, Z., Yousaf, M. N., Waqas, M., Sulaiman, M., Abbas, G., HusIn, M., ... & Hanif, M. (2021). An effective genetic algorithm-based feature selection method for intrusion detection systems. Computers & Security, 110, 102448.
54. Wazirali, R. (2021). Intrusion detection system using fknn and improved PSO. CMC-Comput Mater Contin, 67(2), 1429-1445.
55. Heidari, A., Navimipour, N. J., & Unal, M. (2023). A Secure Intrusion Detection Platform Using Blockchain and Radial Basis Function Neural Networks for Internet of Drones. IEEE Internet of Things Journal.
56. Slamet, Mohamed, I. I., & Samsuri, F. (2020). Campus hybrid intrusion detection system using snort and c4. 5 algorithm. In InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer

said, N. E. (2015). Intrusion detection techniques in wireless sensor network using data mining algorithms: comparative evaluation based on attacks detection. International Journal of Advanced Computer Science and Applications, 6(9), 164-172.

70. Li, Y., Xia, J., Zhang, S., Yan, J., Ai, X., & Dai, K. (2012). An efficient intrusion detection system based on support vector machines and gradually feature removal method. Expert systems with applications, 39(1), 424-430.
71. Shah, B., & Trivedi, B. H. (2015, February). Reducing features of KDD CUP 1999 dataset for anomaly detection using back propagation neural network. In 2015 Fifth International Conference on Advanced Computing & Communication Technologies (pp. 247-251). IEEE.
72. Yulianto, A., Sukarno, P., & Suwastika, N. A. (2019, March). Improving Adaboost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset. In Journal of Physics: Conference Series (Vol. 1192, No. 1, p. 012018). IOP Publishing.
73. Abdulhammed, R., Faezipour, M., Musafer, H., & Abuzneid, A. (2019, June). Efficient network intrusion detection using pca-based dimensionality reduction of features. In 2019 International Symposium on Networks, Computers and Communications (ISNCC) (pp. 1-6). IEEE.
74. Pelletier, Z., & Abualkibash, M. (2020). Evaluating the CIC IDS-2017 Dataset Using Machine Learning Methods and Creating Multiple Predictive Models in the Statistical Computing Language R. Science, 5(2), 187-191.
75. Hammad, M., El-medany, W., & Ismail, Y. (2020, December). Intrusion Detection System using Feature Selection With Clustering and Classification Machine Learning Algorithms on the UNSW-NB15 dataset. In 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT) (pp. 1-6). IEEE.
76. Faker, O., & Dogdu, E. (2019, April). Intrusion detection using big data and deep learning techniques. In Proceedings of the 2019 ACM Southeast Conference (pp. 86-93).
77. Ghajari, Ghazal, Ashutosh Ghimire, Elaheh Ghajari, & Fathi Amsaad. (2025). “Network Anomaly Detection for IoT Using Hyperdimensional Computing on NSL-KDD.” arXiv preprint arXiv: 2503.03031.
78. Ghajari, G., Ghajari, E., Mohammadi, H., & Amsaad, F. (2025). Intrusion Detection in IoT Networks Using Hyperdimensional Computing: A Case Study on the NSL-KDD Dataset. arXiv preprint arXiv:2503.03037.
79. Hegde, R., Likitha, S., & Natesh, M. (2024, November). Intrusion Detection System Using Machine Learning Based on NSL KDD Dataset. In 2024 International Conference on Recent Advances in Science and Engineering Technology (ICRASET) (pp. 1-5). IEEE.
80. Araujo, I., Natalino, C., & Cardoso, D. (2021). A GPU-assisted NFV framework for intrusion detection system. Computer Communications, 169, 92-98..
81. Khandelwal, S., Wadhwa, E., & Shreejith, S. (2022, July). Deep learning-based embedded intrusion detection system for automotive can. In 2022 IEEE 33rd international conference on application-specific systems, architectures and processors (ASAP) (pp. 88-92). IEEE.
82. Le, T. T. H., Kim, H., Kang, H., & Kim, H. (2022). Classification and explanation for intrusion detection system based on ensemble trees and SHAP method. Sensors, 22(3), 1154.