دوره هجدهم، شماره تابستان ۱۴۰۴

مجله فناوری اطلاعات در طراحی مهندسی

Information Technology in Engineering Design

http://sanad.iau.ir/journal/ited

تشخیص حملات در اینترنت اشیا با استفاده از الگوریتم‌های مبتنی بر طبقه‌بند ترکیبی و بهینه‌سازی شاهین هریس

محمدحسین اختیاری*[(۱)] مهدی جعفری شهباززاده[(۲)] مهدیه اسلامی[(۳)]

(۱) گروه مهندسی برق، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران *

(۲) گروه مهندسی برق، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران

(۳) گروه مهندسی برق، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران

(تاریخ دریافت: ۱۴۰۳/۰۹/۰۹ تاریخ پذیرش: ۱۴۰۴/۰۳/۱۰)

چکیده

شبکه‌های کامپیوتری به دلیل آسیب‌پذیری‌های متعدد، همواره در معرض انواع حملات قرار دارند. با توجه به حجم بالا و پیچیدگی داده‌های ترافیک شبکه، تشخیص دقیق نفوذ با استفاده از مدل‌های یادگیری ماشین چالش‌برانگیز و زمان‌بر است. در این مقاله، یک چارچوب نوآورانه برای تشخیص نفوذ ارائه می‌شود که ترکیبی از الگوریتم بهینه‌سازی شاهین هریس (Harris Hawks Optimization – HHO) و مدل‌های یادگیری ماشین است. در گام نخست، الگوریتم HHO با شبیه‌سازی رفتار شکار شاهین‌ها، مجموعه‌ای از ویژگی‌های مؤثر را از داده‌ها انتخاب کرده و با کاهش ابعاد، موجب افزایش سرعت و دقت تشخیص می‌گردد. سپس طبقه‌بندی حملات با استفاده از ترکیبی از طبقه‌بندهای قدرتمند انجام می‌شود. عملکرد روش پیشنهادی بر روی مجموعه‌داده NSL-KDD مورد ارزیابی قرار گرفته است. نتایج تجربی نشان می‌دهند که مدل پیشنهادی موفق به دستیابی به دقت ۹۷٫۹۹٪ برای حملات DoS، ۹۸٫۹۳٪ برای Probe، ۹۶٫۹۷٪ برای R2L و ۹۷٫۹۹٪ برای U2R شده است. همچنین میزان فراخوانی برای این دسته‌ها به ترتیب ۹۶٫۹۸٪، ۹۷٫۸۳٪، ۹۸٫۰۱٪ و ۸۸٫۸۹٪ به دست آمده است. این نتایج مؤید برتری و کارایی روش پیشنهادی در مقایسه با روش‌های موجود در تشخیص دقیق انواع حملات است.

کلمات کلیدی: تشخیص ناهنجاری، تشخیص نفوذ، شاهین هریس، داده‌کاوی.

*عهده‌دار مکاتبات:

محمدحسین اختیاری

نشانی: گروه مهندسی برق، واحد کرمان، دانشگاه آزاد اسلامی، کرمان، ایران

پست الکترونیکی: imthemhe@gmail.com

۱- مقدمه

توسعه سریع اینترنت اشیا، بسیاری از صنایع مانند خانه‌های هوشمند، کشاورزی هوشمند، و مراقبت‌های بهداشتی را دگرگون کرده است [1]. پیش‌بینی می‌شود تعداد دستگاه‌های اینترنت اشیا تا سال ۲۰۲۵ به بیش از ۴,۱ میلیارد دستگاه برسد [2]. این دستگاه‌ها با وجود نقش حیاتی خود در زندگی روزمره، به دلیل اتصال گسترده به اینترنت در معرض تهدیدات امنیتی مختلفی قرار دارند. به‌عنوان نمونه، تبادل اطلاعات از طریق اینترنت دستگاه‌های IoT را به اهدافی آسیب‌پذیر برای حملات شبکه‌ای متعدد تبدیل می‌کند. طبق گزارش‌ها، حملات جدید بات‌نت در سال ۲۰۲۰ به‌طور چشمگیری افزایش یافته و ۵۷ درصد از دستگاه‌های IoT در برابر حملات آسیب‌پذیر بوده‌اند [3]. ازاین‌رو، بهبود امنیت دستگاه‌های IoT به یکی از موضوعات حیاتی در پژوهش‌های امنیتی تبدیل شده است [4-5].

برای مقابله با این تهدیدات، سیستم‌های تشخیص نفوذ (IDS) به‌منظور شناسایی رفتارهای مخرب و افزایش امنیت ارتباطات توسعه یافته‌اند. این سیستم‌ها با رصد بلادرنگ شبکه و شناسایی ناهنجاری‌ها، حملات را شناسایی و اخطار می‌دهند [6]. در سال‌های اخیر، یادگیری ماشین به‌عنوان یک ابزار کلیدی در سیستم‌های IDS معرفی شده است. الگوریتم‌های یادگیری ماشین نه تنها توانایی استخراج الگوهای پیچیده از داده‌های بزرگ را دارند، بلکه داده‌های غیرخطی و با ابعاد بالا را نیز مدیریت می‌کنند، که این ویژگی‌ها آن‌ها را برای تشخیص نفوذ در سیستم‌های IoT بسیار مناسب می‌سازد [7].

با وجود پیشرفت‌های حاصل‌شده در الگوریتم‌های یادگیری ماشین، چالش‌هایی مانند محدودیت منابع محاسباتی دستگاه‌های IoT، حجم بالای داده‌های ترافیکی، و وجود ویژگی‌های زائد و نامرتبط همچنان باقی است [8,9]. ویژگی‌های اضافی در داده‌ها می‌توانند باعث کاهش دقت مدل و افزایش هزینه‌های محاسباتی شوند. به همین دلیل، انتخاب ویژگی به‌عنوان یک راهکار مؤثر برای کاهش ابعاد داده و بهبود کارایی سیستم‌های تشخیص نفوذ مطرح شده است [10].

اخیراً الگوریتم‌های فراابتکاری به دلیل قابلیت‌های عالی جستجوی جهانی، توجه قابل توجهی را در زمینه انتخاب ویژگی به دست آورده‌اند [11]. الگوریتم‌های فراابتکاری رایج شامل الگوریتم ژنتیک ، بهینه‌سازی ازدحام ذرات (PSO) [12]، الگوریتم بهینه‌سازی نهنگ [13]، بهینه‌سازی گرگ‌خاکستری [14] و بازپخت شبیه‌سازی شده، از جمله موارد دیگر است. در میان این الگوریتم‌ها، GWO به دلیل سهولت پیاده‌سازی، سرعت همگرایی سریع و قابلیت‌های بهینه‌سازی قوی، توجه قابل توجهی را به خود جلب کرده است. برای استفاده بهتر از GWO برای مشکلات انتخاب ویژگی، یک بهینه‌سازی جدید گرگ‌خاکستری باینری [15] پیشنهاد گردید. هسو و همکاران [16] اشاره کرد که

روش‌های انتخاب ویژگی ترکیبی از روش‌های انتخاب ویژگی فردی بهتر عمل می‌کنند. علاوه بر این، حذف ویژگی بازگشتی، یک روش انتخاب ویژگی پوشش، تعادل خوبی بین دقت و زمان اجرا ایجاد می‌کند و سطح مشخصی از دقت را حفظ می‌کند و در عین حال زمان اجرا را کاهش می‌دهد .

در این میان، الگوریتم‌های فراابتکاری به دلیل قابلیت جستجوی جهانی، توجه زیادی در زمینه انتخاب ویژگی جلب کرده‌اند. الگوریتم بهینه‌سازی شاهین هریس (Harris Hawk Optimization) با الهام از رفتار شکار شاهین‌ها، از جمله الگوریتم‌های نوظهوری است که عملکرد برجسته‌ای در انتخاب ویژگی‌ها دارد. این الگوریتم با شناسایی ویژگی‌های کلیدی، ابعاد داده‌ها را کاهش داده و زمان آموزش مدل را بهبود می‌بخشد. شاهین هریس در مقایسه با الگوریتم‌های دیگر مانند PSO و GWO، علاوه بر دقت بالا، سرعت همگرایی مناسبی نیز ارائه می‌دهد. این ویژگی‌ها، شاهین هریس را به انتخابی ایده‌آل برای حل مشکلات انتخاب ویژگی در مجموعه داده‌های حجیم و پیچیده تبدیل کرده است.

در این پژوهش، یک رویکرد جدید مبتنی بر ترکیب الگوریتم شاهین هریس برای انتخاب ویژگی و مدل‌های یادگیری ماشین برای تشخیص نفوذ در شبکه‌های IoT ارائه شده است. ابتدا داده‌ها پیش‌پردازش شده و ویژگی‌های مؤثر با استفاده از شاهین هریس انتخاب می‌شوند. سپس با بهره‌گیری از ترکیب چندین طبقه‌بند، حملات شناسایی می‌شوند. این روش بر روی مجموعه‌داده NSL-KDD ارزیابی شده و نتایج نشان می‌دهند که دقت بالاتر از ۹۸ درصد حاصل شده است، که کارایی و برتری روش پیشنهادی را تأیید می‌کند.

۲- کارهای مرتبط

اخیراً تکنیک‌های یادگیری ماشین به‌طور گسترده در زمینه تشخیص نفوذ در اینترنت اشیا به کار گرفته شده و نتایج بسیار خوبی به دست آورده است. سیستم‌های تشخیص نفوذ مبتنی بر یادگیری ماشینی معمولاً به دو بخش تقسیم می‌شوند. بخش اول پیش پردازش داده است که شامل پیش پردازش داده‌ها قبل از وارد کردن آن به مدل است. این شامل انتخاب ویژگی و مدیریت مجموعه داده‌های نامتعادل برای ارائه ورودی‌های بهتر به مدل است. بخش دوم طبقه‌بندی‌کننده است، جایی که انتخاب یک مدل مناسب می‌تواند نرخ تشخیص نفوذ را به حداکثر برساند. بنابراین، بسیاری از محققان تلاش خود را بر روی این دو جنبه برای ایجاد سیستم‌های تشخیص نفوذ قدرتمند متمرکز کرده‌اند. در این بخش به بررسی کارهای اخیر می‌پردازیم.

در [16]، بوکه و همکاران. (۲۰۲۳) مدلی را برای تشخیص هوشمند حملات DDoS بر اساس درخت تصمیم و روش انتخاب ویژگی بهینه شده با شاخص جینی ارائه کرد. این مدل بر روی مجموعه داده UNSW-NB15 با بیش از ۱۱۴۰۰۰۰ نمونه آزمایش شد و به دقت ۹۸ درصد دست یافت که از روش‌های پیشرفته مانند Random Forest و XGBoost بهتر عمل کرد. روش انتخاب ویژگی با استفاده از شاخص جینی، ابعاد داده‌ها را کاهش می‌دهد و از برازش بیش از حد جلوگیری می‌کند.

در [17] مصطفی و همکاران. (۲۰۲۳) مدلی مبتنی بر شبکه‌های عصبی حافظه کوتاه مدت بلند مدت (LSTM) برای شناسایی حملات DDoS پیشنهاد کرد. این مدل که توانایی یادگیری وابستگی‌های طولانی مدت را دارد، در تشخیص حملات DDoS به دقت بالایی دست یافت و همچنین با انواع مختلف حملات تولید شده توسط GAN آزمایش شد. نتایج نشان داد که این مدل برای تشخیص ترافیک GAN متخاصم با دقت بین ۹۱,۷۵ تا ۱۰۰ درصد مؤثر است.

در [18]، Khanday و همکاران. (۲۰۲۳) یک سیستم سبک وزن برای تشخیص نفوذ با استفاده از تکنیک‌های پیش پردازش داده‌ها و ترکیبی از یادگیری ماشین و طبقه‌بندی کننده‌های یادگیری عمیق معرفی کرد. این سیستم از مجموعه داده‌های BOT-IoT و TON-IoT استفاده کرد و آزمایش‌هایی برای طبقه‌بندی حملات DDoS انجام شد. عدم تعادل داده‌ها در BOT-IoT در مقایسه با TON-IoT بیشتر بود.

در [19] رانی و همکاران. (۲۰۲۳) بر شناسایی و جلوگیری از حملات DDoS مانند SYN و Slowloris در ارتباطات D2D متمرکز شد. با شبیه‌سازی حملات در یک شبکه D2D و ایجاد یک مجموعه داده ویژه، مدل یادگیری ماشینی آن‌ها قادر به شناسایی و جلوگیری از حملات بود.

در [20] حسن و همکاران. (۲۰۲۳) همچنین آسیب‌پذیری‌های پروتکل‌های ارتباطی شبکه هوشمند را مطالعه کرد، تکنیک‌های حمله DDoS و روش‌های تشخیص را بررسی کرد و یک روش ترکیبی مبتنی بر یادگیری ماشین برای یک سیستم شبکه هوشمند پایدار پیشنهاد کرد.

لازارینی و همکاران [26] یک سیستم تشخیص نفوذ اینترنت اشیاء را با استفاده از رویکرد انباشته کردن مجموعه ساخت. آنها چهار مدل مختلف یادگیری برای شناسایی و طبقه‌بندی حملات در محیط‌های اینترنت اشیا ترکیب کردند. آزمایش‌های باینری و چند کلاسه بر روی مجموعه داده‌های Ton-IoT و CIC-IDS2017 انجام شد. نتایج نشان داد که روش پیشنهادی قادر به تشخیص اکثر حملات با نرخ مثبت کاذب به خصوص پایین و منفی کاذب است. با این حال، این رویکرد چهار مدل مختلف را ادغام می‌کند که به مقدار قابل توجهی از منابع و ارزیابی بیشتر عملکرد آن در دستگاه‌های واقعی اینترنت اشیا نیاز دارد. آلانی [27] از حذف ویژگی بازگشتی مبتنی بر اهمیت ویژگی برای انتخاب

ویژگی در مجموعه داده استفاده کرد و ۱۱ ویژگی مهم را انتخاب کرد. آنها از طبقه‌بندی درخت تصمیم برای طبقه‌بندی و توضیح افزودنی Shapley (SHAP) برای توضیح ویژگی‌ها و طبقه‌بندی‌کننده انتخاب‌شده استفاده کردند. روش پیشنهادی به دقت ۰,۹۹۹۷ در مجموعه داده WUSTL-IIOT-2021 دست یافت. نظام الدین [25] از نرمال‌سازی درجه‌بندی عدد صحیح برای پیش‌پردازش داده‌ها استفاده کرد و از بهینه‌ساز الهام گرفته از موش‌های مبتنی بر یادگیری برای انتخاب ویژگی برای حفظ ویژگی‌های مهم استفاده کرد. آزمایشات روی یک مجموعه داده ترکیبی دقت تشخیص بهبود یافته را در مقایسه با سایر روش‌های پیشرفته نشان داد. شارما و همکاران [28] یک سیستم تشخیص نفوذ اینترنت اشیا را بر اساس مدل شبکه عصبی عمیق برای محافظت بهتر از امنیت دستگاه‌های اینترنتی پیشنهاد کرد. آن‌ها از یک شبکه متخاصم مولد برای ترکیب داده‌های کلاس حمله اقلیت استفاده کردند و از روش فیلتر ضریب همبستگی پیرسون برای انتخاب ویژگی استفاده کردند. نتایج تجربی روی مجموعه داده UNSW-NB15 با داده‌های متعادل به دقت ۹۱ درصد دست یافت.

کریم و همکاران [29] یک الگوریتم انتخاب ویژگی را با استفاده از الگوریتم برای ازدحام پرندگان (BSA) برای بهبود عملکرد بهینه ساز نیروهای گوریل پیشنهاد کرد. آزمایش‌ها بر روی مجموعه داده‌های NSL-KDD، CICIDS-2017، UNSW-NB15، و Bot-IoT نشان دادند که GTO-BSA پیشنهادی به سرعت و عملکرد همگرایی بهتری دست یافت. موهی الدین و همکاران [36] یک سیستم تشخیص نفوذ اینترنت اشیا را بر اساس الگوریتم نزدیک‌ترین همسایه با استفاده از تجزیه و تحلیل مؤلفه اصلی ، آزمون‌های آماری تک متغیره، و الگوریتم ژنتیک برای انتخاب ویژگی ارائه کرد. آزمایشات روی مجموعه داده Bot-IoT به دقت بالای ۹۹,۹۹٪ دست یافت و در عین حال زمان پیش‌بینی را به طور قابل توجهی کاهش داد. لیو و همکاران [30] با پیشنهاد یک روش انتخاب ویژگی بر اساس یک الگوریتم ژنتیک، به موضوع ویژگی‌های جریان بیش از حد مؤثر بر سرعت تشخیص در سیستم‌های تشخیص نفوذ IoT پرداخت. آزمایش‌های گسترده روی مجموعه داده‌های Bot-IoT شش ویژگی را از ۴۰ ویژگی انتخاب کرد که به دقت ۹۹,۹۸ درصد و امتیاز F1 به ۹۹,۶۳ درصد دست یافت. الوشاه و همکاران [31] یک الگوریتم انتخاب ویژگی بسته بندی جدید را پیشنهاد کرد که از مستعمره پنگوئن امپراتور برای کاوش در فضای جستجو استفاده می‌کرد و K-نزدیک‌ترین همسایه‌ها را به عنوان طبقه‌بندی کننده انتخاب می‌کرد. نتایج تجربی بر روی مجموعه داده‌های معروف اینترنت اشیا، دقت بهبود یافته و کاهش اندازه ویژگی را در مقایسه با روش‌هایی مانند بهینه‌سازی ازدحام ذرات چند هدفه نشان داد. حسن و همکاران [32] از یک الگوریتم جستجوگر باینری پرتوی باینری برای انتخاب ویژگی برای حذف ویژگی‌های اضافی و نامربوط از مجموعه داده استفاده کرد و از یک طبقه‌بندی جنگل تصادفی برای طبقه‌بندی استفاده کرد. روش پیشنهادی بر روی مجموعه داده‌های NSL-KDD و CIC-IDS2017 مورد ارزیابی قرار گرفت و به ترتیب ۲۲ و ۳۸ ویژگی را

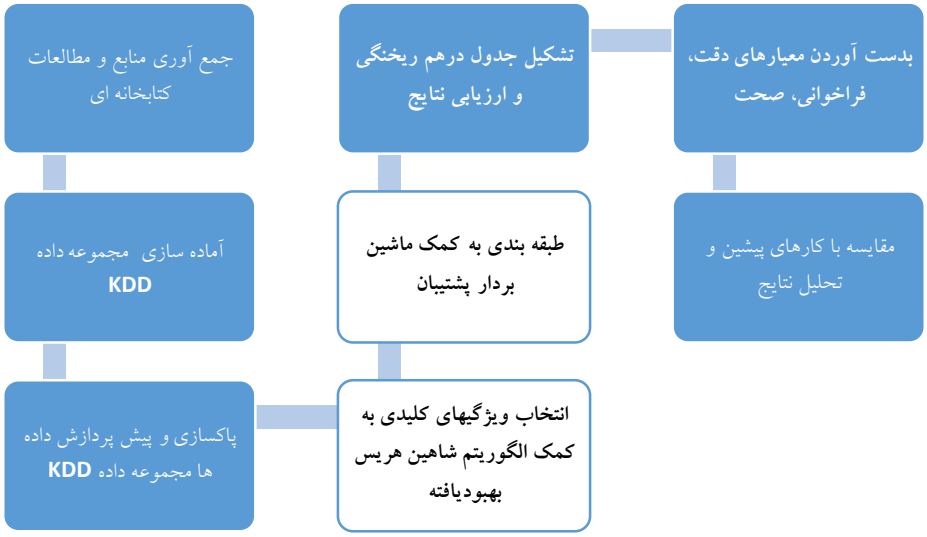
انتخاب کرد و به دقت ۹۸,۸٪ و ۹۹,۳٪ رسید. محی الدین و همکاران [33] برای کاهش پیچیدگی بهینه‌سازی انتخاب ویژگی، انتخاب ویژگی‌های مهم، و از XGBoost به عنوان طبقه‌بندی کننده استفاده کرد. نتایج تجربی روی مجموعه داده UNSW-NB15 به نرخ دقت ۹۹٪ و ۹۱٪ برای طبقه‌بندی باینری و چند کلاسه، و دقت ۹۸٪ برای طبقه‌بندی باینری در مجموعه داده CIC-IDS2017 دست یافت.

بررسی ادبیات نشان می‌دهد که اگرچه روش‌های فوق‌الذکر به عملکرد تشخیص نسبتاً بالایی دست یافته‌اند، هنوز فضا برای بهبود در نرخ تشخیص و بهینه‌سازی روش‌های انتخاب ویژگی وجود دارد. علاوه بر این، تنها تکیه بر یک روش انتخاب ویژگی منفرد ممکن است به زیرمجموعه ویژگی بهینه منجر نشود، که می‌تواند بر عملکرد تشخیص مدل‌ها تأثیر بگذارد.

۳- روش پیشنهادی

با گسترش تکنولوژی، تشخیص نفوذ به عنوان یک مقوله ضروری در مورد بررسی قرار گرفته است. سیستم‌های تشخیص نفوذ تلاش می‌کنند تا رفتارهای نرمال یا غیرنرمال کاربران را شناسایی و اطلاع رسانی کنند. سیستم‌های تشخیص نفوذ چالشی پیچیده با داده‌های ترافیک شبکه دارند. روش‌های IDS در سطوح مختلفی از دقت و صحت پیشنهاد و تولید می‌شوند . به این دلیل توسعه سیستم‌های تشخیص نفوذ موثر و قوی لازم است. در این مقاله قصد داریم به ارائه مدلی جهت تولید سیستم تشخیص نفوذ بپردازیم. در این راستا هدف اصلی این پروژه افزایش نرخ تشخیص درست و کاهش نرخ اشتباه می‌باشد. موضوع اصلی این تحقیق بر پایه تکنیک‌های داده‌کاوی بنا شده است و در این راستا از تکنیک‌های طبقه بندی برای تشخیص نفوذ بهره می‌گیریم. همچنین دیتاست مورد استفاده در این پژوهش مجموعه داده‌های NSL-KDD با چهار نوع حمله DOS، Probe، U2R و R2l ثبت شده در آن می‌باشد.

در این بخش به روش پیشنهادی تحقیق پرداخته می‌شود. شکل ۱ روش پیشنهادی تحقیق را نشان می‌دهد.



شکل ۱: روش پیشنهادی

همان‌طور که در شکل ۱ نشان داده شده است، مراحل روش پیشنهادی به صورت خلاصه به صورت زیر است:

الگوریتم اول: مراحل کلی روش پیشنهادی	
ورودی	مجموعه داده NSL-KDD از ورودی خوانده می‌شود.
۱	پیش پردازش روی داده‌ها انجام می‌شود:
۲	ویژگی‌های کلاس و غیرکلاس (۴۱) ویژگی از هم تفکیک می‌گردند.
۳	کاهش تعداد ابعاد به کمک شاهین هریس بهبود یافته انجام می‌شود.(۷ تا از ۴۱ ویژگی طبق این الگوریتم انتخاب می‌گردد.)
۴	فاز تشخیص نفوذ(مرحله یادگیری توسط الگوریتم شهین هریس بهبود یافته) به صورت زیر انجام می‌شود:
۵	ساخت مدل اولیه:
۶	داده‌های پیش پردازش شده به طبقه‌بند ماشین بردار پشتیبان داده می‌شود و مدل مبتنی بر ماشین بردار پشتیبان ساخته می‌شود.(SVM)
۷	الگوریتم شاهین هریس برای اعمال ویژگی‌های بهینه اعمال می‌گردد
۸	مرحله ارزیابی: ماتریس درهم ریختگی(Confusion matrix) برای داده‌های آزمایشی ساخته می‌شود.
خروجی	معیارهای دقت، فراخوانی، صحت

شکل ۲: الگوریتم یادگیری روش پیشنهادی

۳-۱- فاز پیش‌پردازش داده‌ها

فاز پیش پردازش داده ها شامل جایگزینی داده‌های سمبولیک مانند پروتکل و سرویس با مقدار گسسته و جدا ازهم وگسسته‌سازی مقادیر مشابه و هم جنس در یک رنج محدود و جایگزینی مقادیر از دست رفته با با میانه و میانگین داده‌های آموزشی است. مطابق با این الگوریتم مقادیر اصلی با مقادیر متناظر در محدوده دیگر جایگزین شدند. از آنجا که در مجموعه داده‌های NSL-KDD چهار نوع حمله اصلی به انوع مختلفی از حملات دسته‌بندی می‌شوند ما هریک از این حملات را با یکی از چهار نوع حمله اصلی برچسب گذاری می‌کنیم. این عملیات داده‌های ورودی را به داده‌های مورد پذیرش در سیستم تشخیص نفوذ پیشنهادی تبدیل می‌کند.

۳-۲- انتخاب ویژگی با استفاده از الگوریتم شاهین هریس

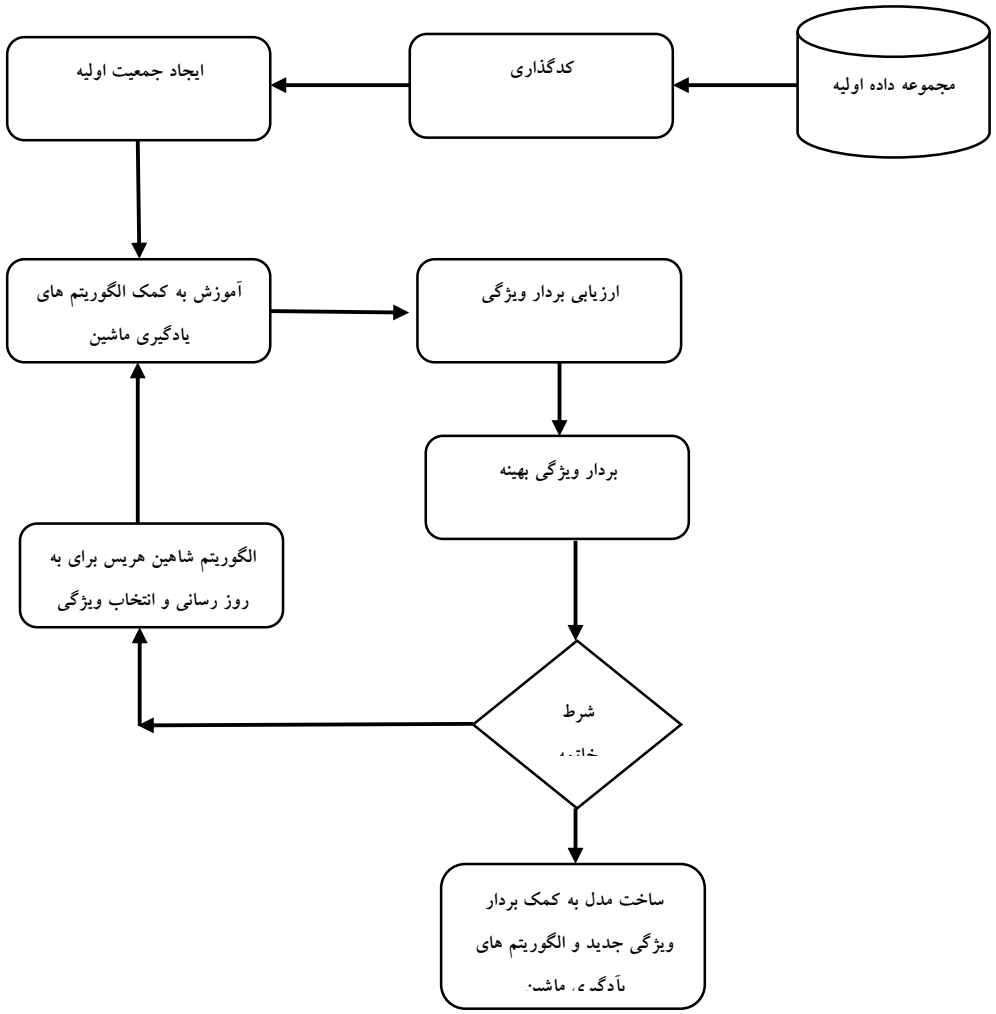
از الگوریتم شاهین هریس جهت یافتن مؤثرترین ویژگی‌ها به منظور مدل‌سازی بهینه‌ی تشخیص نفوذ شبکه استفاده می‌گردد. برای انتخاب ویژگی می‌توان از الگوریتم‌های فراابتکاری بهره گرفت. یکی از محبوب‌ترین الگوریتم‌های فراابتکاری مبتنی بر جمعیت، بهینه‌سازی شاهین هریس (HHO) است [2] که مکانیسم‌های شکار شاهین هریس در طبیعت را تقلید می‌کند. اگرچه شاهین هریس می‌تواند راه‌حل‌های بهینه را برای مسائل خاص به دست آورد، اما در راه‌حل‌های بهینه محلی راکد می‌ماند. در این تحقیق، یک بهینه‌سازی بهبود یافته شاهین هریس برای حل مسائل بهینه‌سازی پیشنهاد می‌شود. در مرحله اول، ما نقشه آشفته چادر را در مرحله اولیه برای بهبود تنوع جمعیت اولیه معرفی می‌کنیم. روال کلی انتخاب ویژگی برای الگوریـتم شاهین هریس به صورت زیر است.

الگوریتم ۲: شبه کد الگوریتم شاهین هریس برای انتخاب ویژگی

```
Input     Dataset D with n features
1         Initialize the population of hawks Xi (i = 1, 2, ..., N)
2         Set the maximum number of iterations T
3         Evaluate the fitness of each hawk using a fitness function (e.g., classification accuracy + feature count penalty)
4         Find the current best solution X_best
5         for t = 1 to T do
6           for each hawk Xi do
7             Calculate the escaping energy E using = 2 * (1 - t / T) * rand()
8             if |E| ≥ 1 then
9               Perform exploration
10                Update Xi using a random hawk position
```

```
        else
            Perform exploitation
                if rand() ≥ 0.5 then
                    Perform soft besiege
                else
                    Perform hard besiege
                end if
            end if
            Evaluate the fitness of updated Xi
        end for
        Update X_best if a better solution is found
    end for
    Return the feature subset corresponding to X_best
output  Best feature subset F_best
```

شکل۳: شبه کد الگوریتم شاهین هریس



شکل ۴: فلوچارت الگوریتم بهینه‌سازی ویژگی

همان‌طور که در شکل ۳ مشخص است ابتدا باید کدگذاری صورت گیرد، کدگذاری به این صورت است که هر بردار ویژگی به صورت یک آرایه و به صورت یک شاهین است.

F1	F2	F3	F4	…	…	….	..	..	Fn

شکل ۵: بردار ویژگی

همان‌طور که در شکل ۴ مشخص است یک بردار ویژگی مجموعه از ویژگی‌های است که مقادیر صفر و یک (باینری) دارد و بعد از اعمال اگوریـم شاهین هریس بعضی از ویژگی‌ها حذف می‌شوند و ویژگی‌های بهینه برای مدل‌سازی انتخاب می‌شوند.

اگرچه الگوریتـم شاهین هریس می‌تواند راه‌حل‌های بهینه را برای مسائل متنوعی به دست آورد، اما یکی از ایراداتی که به این الگوریتم وارد است این است که در راه‌حل‌های بهینه محلی گیر می‌کند. در نتیجه، یک بهینه‌سازی بهبود یافته شاهین برای حل این چالش لازم است. یکی از راه‌ها برای فرار از این بهبود تنوع جمعیت اولیه است. در واقع در فاز اول، که فاز مقداردهی اولیه است راهی برای ایجاد تنوع در جمعیت اولیه ایجاد شود. در مرحله دوم، یک عامل اکتشاف برای بهینه‌سازی پارامترها برای بهبود توانایی اکتشاف ارایه گردد.

الگوریتم شاهین هریس دارای سه فاز کلی است:

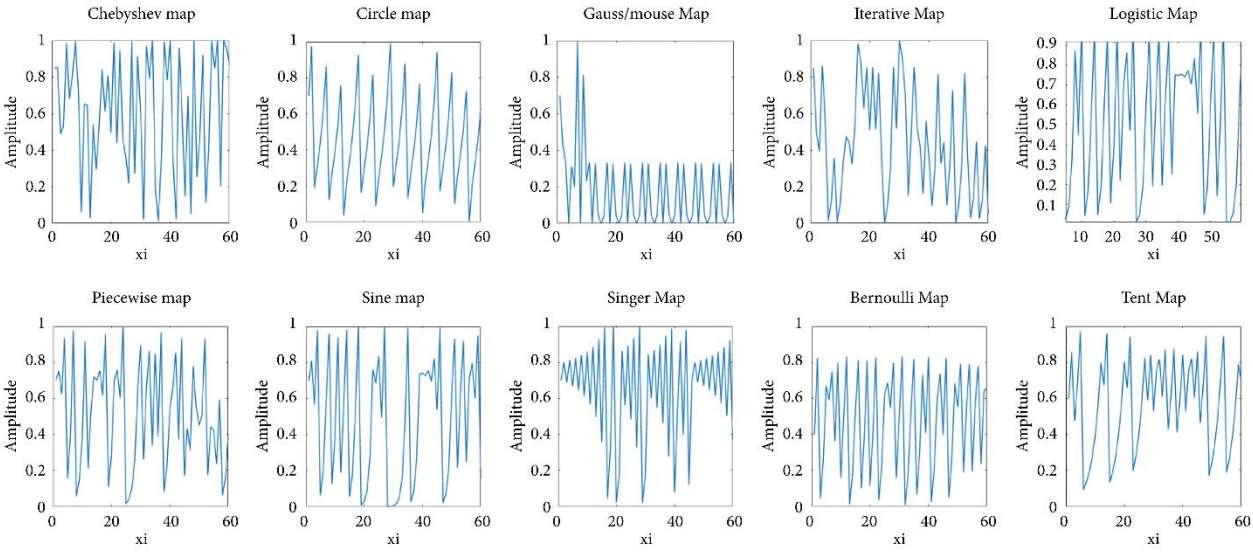
فاز اولیه (مرحله اکتشاف)**:** در این مرحله، شاهین‌ها در مکان‌های تصادفی بر اساس سایر اعضا یا مکان‌های خرگوش نشسته‌اند**.**

فاز دوم (دگرگونی اکتشاف و بهره‌برداری)

فاز سوم–نهایی (مرحله بهره‌برداری)

فاز اول: هدف از این فاز افزایش تنوع جمعیت اولیه است. مقداردهی اولیه مکان تأثیر خاصی بر تنوع جمعیت و پایداری الگوریتم دارد. الگوریتم HHO فقط می‌تواند تصادفی بودن موقعیت جمعیت را در مرحله اولیه‌سازی تضمین کند، اما تصادفی بودن به معنای یکنواختی نیست. الگوریتم آشوب[۱] می‌تواند اعداد تصادفی با توزیع یکنواخت بین ۰ و ۱ ایجاد کند. می‌توان از نقشه‌های مختلف آشوب برای ایجاد جمعیت اولیه متنوع استفاده نمود و براساس هر کدام آزمایش‌های متنوع انجام داد. ویژگی‌ها و تصادفی بودن این نقشه می‌تواند با تغییر موقعیت اولیه شاهین‌ها به طور موثری عملکرد را بهبود بخشد.

[۱] Chaotic Algorithm

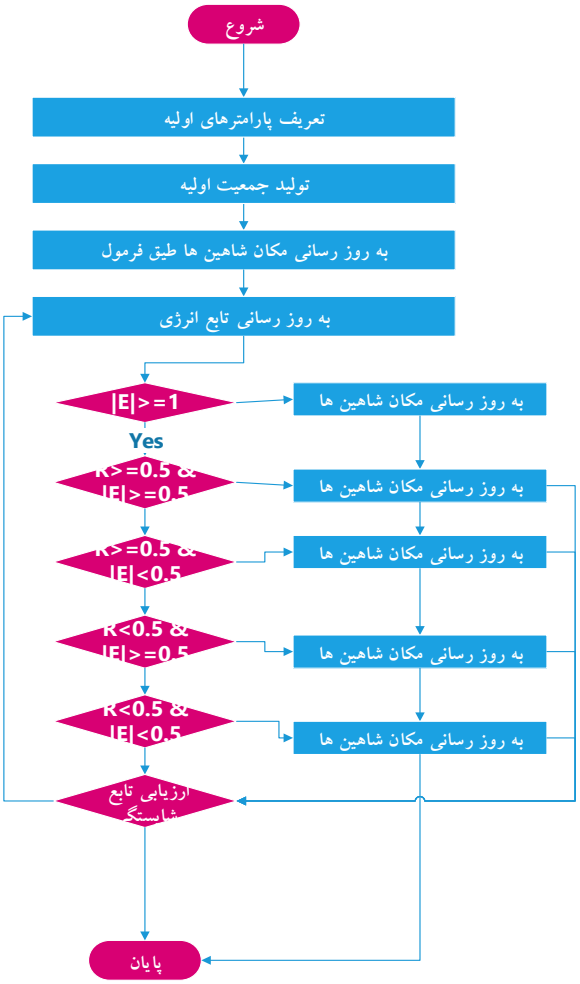


شکل ۶: نقشه‌های مختلف آشوب

ما تنوع جمعیت اولیه را با اصلاح موقعیت‌های اولیه براساس نقشه های فوق افزایش می‌دهیم.

فاز دوم: مثل الگوریـتم اصلی است و تغییری ندارد.

فاز سوم: در مرحله بهره‌برداری الگوریـتم شاهین هریس، شاهین هریس موقعیت خود را از طریق چهار استراتژی (محاصر نرم، محاصر سخت، محاصره نرم با شیرجه‌های سریع پیش رونده، محاصره سخت با شیرجه‌های سریع پیشرونده) به‌روز می‌کند. اگرچه امکان کاوش را افزایش می‌دهد، اما ورود به تکرار بعدی بدون تداخل می‌تواند به راحتی منجر به گیر افتادن الگوریتم در بهینه محلی شود. برای بهبود این مشکل، می‌توان از استراتژی پیاده‌روی تصادفی گاوسی و یا استراتژی حریصانه استفاده کرد. این استراتژی‌ها دارای ویژگی‌های تثبیت در طیفی از مقادیر با احتمال بالا و تغییر شدید مقادیر با احتمال کم هستند. و همه این استراتژی‌ها با ایجاد یک انحراف، روش را بهبود می‌بخشد.



شکل ۷: فلوچارت روش پیشنهادی

۳-۳- تابع شایستگی

بدست آوردن تابع شایستگی مهمترین بخش الگوریتم‌های فراابتکاری برای حل مساله است. باتوجه این که در این تحقیق مساله هدف ساخت مدل و طبقه‌بندی است دقت، فراخوانی و همچنین صحت مدل می‌تواند در انتخاب یک کدگذاری مناسب موثر باشد. در نتیجه استفاده از هر سه معیار برای ساخت تابع شایستگی پیشنهاد شده است. برای ارزیابی عملکرد دسته‌بندی ابتدا باید ماتریس درهم ریختگی برای داده‌های تست ماشین بردار مطابق جدول ۱ تشکیل و مقادیر زیر محاسبه گردد:

جدول۱: ماتریس درهم ریختگی

	حمله	نرمال
حمله	TP	FN
نرمال	FP	TN

TN: تعداد رکوردهایی که دسته واقعی آن ها نرمال بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به درستی نرمال تشخیص داده است.

TP: تعداد رکوردهایی که دسته واقعی آن‌ها حمله بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به درستی حمله تشخیص داده است. (البته قابل ذکر است که با توجه به اینکه در اینجا ۴ نوع حمله وجود دارد منظور جمع تعداد همه آن‌هاست)

FN: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها حمله بوده و الگوریتم دسته‌بندی دسته آن‌ها را به اشتباه نرمال تشخیص داده است.

FP: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها نرمال بوده و الگوریتم دسته‌بندی نیز دسته آن‌ها را به اشتباه حمله تشخیص داده است.

$$\text{Recall}^{\text{نرمال}} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (1)$$

$$\text{Recall}^{\text{حمله}} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (2)$$

$$\text{Precision}^{\text{نرمال}} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (3)$$

$$\text{Precision}^{\text{حمله}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (5)$$

معیار Recall^{x} دقت دسته‌بندی دسته X را با توجه به کل رکوردها با برچسب X نشان می‌دهد. معیار Precision^{x} دقت دسته‌بندی دسته X را با توجه به کل مواردی نشان می‌دهد که برچسب X برای رکوردها مورد بررسی توسط دسته‌بند پیشنهاد شده است. معیار Recall^{x} کارایی دسته‌بندی را با توجه به تعداد رخداد دسته X نشان می‌دهد حال آنکه معیار Precision^{x} اساسا مبتنی بر دقت پیش‌بینی دسته‌بند می‌باشد و نشان می‌دهد به چه میزان می‌توان به خروجی دسته‌بند اعتماد نماییم.

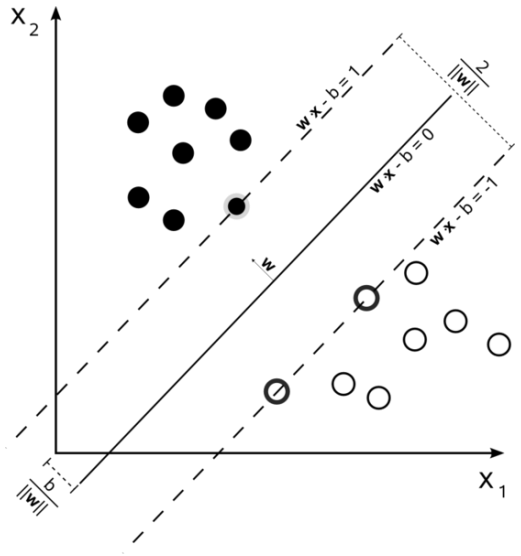
بعد از بدست آوردن این سه معیار باید برای هر کدام از معیارها یک وزن براساس ارزش آن معیار در نظرگرفته‌شده است. با توجه به اینکه ارزش معیار Accuracy بالاتر از آن دو معیار است به آن وزن ۴ و به دو معیار دیگر وزن ۱ داده می‌شود.

$$f\left(\text{زیرمجموعه ویژگی}\right) = \left(\frac{1}{\text{Precision}^{\text{حمله}} + 4 * \text{Accuracy} + \text{Recall}^{\text{حمله}}}\right) \quad (6)$$

با تکرارهای مختلفی که در الگوریتم شاهین هریس انجام می‌شود و با کمک معیارهای Precision، Recall و accuracy و همچنین تابع شایستگی فوق ویژگی‌های زیر به عنوان ویژگی‌های بهینه انتخاب می‌گردند.

۳-۴- ساخت مدل با استفاده از الگوریتم ماشین بردار پشتیبان

ماشین بردار پشتیبان یکی از روش‌های یادگیری بانظارت است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌کنند. این روش از جمله روش‌های نسبتاً جدیدی است که در سال‌های اخیر کارایی خوبی نسبت به روش‌های قدیمی‌تر برای طبقه‌بندی نشان داده‌است. مبنای کاری دسته‌بندی کننده ماشین بردار پشتیبان دسته‌بندی خطی داده‌ها است و در تقسیم خطی داده‌ها سعی می‌کنیم خطی را انتخاب کنیم که حاشیه اطمینان بیشتری داشته باشد. حل معادله پیدا کردن خط بهینه برای داده‌ها به وسیله روش‌های QP که روش‌های شناخته شده‌ای در حل مسائل محدودیت‌دار هستند صورت می‌گیرد. قبل از تقسیمِ خطی برای اینکه ماشین بتواند داده‌های با پیچیدگی بالا را دسته‌بندی کند داده‌ها را به وسیله‌ی تابعِ phi به فضای با ابعاد خیلی بالاتر می‌بریم. از توابع هسته مختلفی از جمله هسته‌های نمایی، چندجمله‌ای و سیگموید می‌توان استفاده نمود.



شکل ۸ : ماشین بردار پشتیبان

۴- نتایج و ارزیابی

در این بخش روش پیشنهادی در بخش قبل را شبیه‌سازی شده و نتایج آن را مورد ارزیابی قرار می‌گیرد و نتایج بدست آمده از این شبیه‌سازی با روش‌های قبلی مورد مقایسه قرار می‌گیرد تا دقت عملکرد الگوریتم پیشنهادی اثبات شود. این شبیه‌سازی با استفاده از متلب ۲۰۱۷ انجام می‌گیرد. در ادامه نتایج حاصل از این شبیه‌سازی را مورد ارزیابی قرار می‌دهیم.

۴-۱- مرحله پیش‌پردازش داده‌ها

مجموعه داده مورد استفاده در این پژوهش مجموعه داده‌های NSL-KDD می‌باشد که این مجموعه داده یک الگوی بخوبی شناخته شده در پژوهش تکنیک‌های تشخیص نفوذ است. این مجموعه داده شامل ۴۱ ویژگی می‌باشد که وضعیت این ویژگی‌ها در جدول ۲ نشان داده شده است.

جدول ۲: بررسی ویژگی‌های رکوردهای موجود در NSL-KDD

ردیف	نوع ویژگی	توضیحات	نام ویژگی
۱	Numeric	مدت زمان اتصال	Duration
۲	Nominal	نوع پروتکل TCP,UDP,ICMP	protocol_type
۳	Nominal	نوع سرویس شبکه.Telnet,Http,etc	service
۴	Nominal	نرمال یا اشکال داشتن اتصال را مشخص می‌کند	flag
۵	Numeric	تعداد بایت‌های داده از منبع به مقصد	src_bytes
۶	Numeric	تعداد بایت‌های داده از مقصد به منبع	dst_bytes
۷	Nominal	اگر ۱ باشد یعنی اتصال از یک پورت است در غیر این صورت ۰ می‌شود	land
۸	Numeric	تعداد قطعات اشتباه	wrong_fragment
۹	Numeric	تعداد بسته‌های urgent	urgent
۱۰	Numeric	تعداد شاخص‌های hot را نمایش می‌دهد	hot
۱۱	Numeric	تعداد login های دارای نقص	num_failed_logins
۱۲	Nominal	اگر ۱ باشد یعنی موفقیت آمیز بوده درغیر اینصورت ۰ می‌شود	logged_in
۱۳	Numeric	تعداد شرط‌های compromised	num_compromised
۱۴	Numeric	با ۰ و یا ۱ شدن وضعیت root shell را مشخص می-کند	root_shell
۱۵	Numeric	با ۰ و یا ۱ شدن وضعیت su root را مشخص می‌کند	su_attempted
۱۶	Numeric	تعداد دسترسی‌هایی که بهroot انجام گرفته است	num_root
۱۷	Numeric	تعداد فایل‌های عملیاتی ایجاد شده	num_file_creations
۱۸	Numeric	تعداد هسته‌های آماده	num_shells
۱۹	Numeric	تعداد عملیات روی فایل‌های کنترل دستیابی	num_access_files
۲۰	Numeric	تعداد دستورات خارج شده در نشست ftp	num_outbound_cmds
۲۱	Nominal	با ۰ و یا ۱ شدن مشخص می‌کند که آیا login عضو لیست hot شده یا نه	is_hot_login
۲۲	Nominal	با ۰ و یا ۱ شدن وضعیت guest بودن login را	Is_guest_login

		مشخص می‌کند	
۲۳	Numeric	تعداد اتصالاتی که از یک host در یک اتصال جاری بیش از ۲ ثانیه بطول بکشد	count
۲۴	Numeric	تعداد اتصالاتی که از یک سرویس در یک اتصال جاری بیش از ۲ ثانیه بطول بکشد	srv_count
۲۵	Numeric	درصد اتصالاتی که اشکال SYN دارند	serror_rate
۲۶	Numeric	درصد اتصالاتی که اشکال SYN در سرویس دارند	srv_serror_rate
۲۷	Numeric	درصد اتصالاتی که اشکال REJ دارند	rerror_rate
۲۸	Numeric	درصد اتصالاتی که اشکال REJ در سرویس دارند	srv_rerror_rate
۲۹	Numeric	درصد اتصالاتی به سرویس‌های یکسان	same_srv_rate
۳۰	Numeric	درصد اتصالاتی به سرویس‌های مختلف	diff_srv_rate
۳۱	Numeric	درصد اتصالاتی به host های مختلف	srv_diff_host_rate
۳۲	Numeric	تعداد host های مقصد	dst_host_count
۳۳	Numeric	تعداد سرویس host های مقصد	dst_host_srv_count
۳۴	Numeric	درصد اتصالاتی که از یک host با یک سرویس به یک host مقصد در یک بازه زمانی انجام شده است	dst_host_same_srv_rate
۳۵	Numeric	درصد اتصالاتی که از یک host با سرویس‌های مختلف به یک host مقصد در یک بازه زمانی انجام شده است	dst_host_diff_srv_rate
۳۶	Numeric	درصد اتصالاتی که از یک host با یک پورت منبع انجام شده است	dst_host_same_src_port_rate
۳۷	Numeric	درصد اتصالاتی که از یک host به host دیگر با سرویس متفاوت انجام شده است	dst_host_srv_diff_host_rate
۳۸	Numeric	نرخ اشکالات SYN در host منبع	dst_host_serror_rate
۳۹	Numeric	نرخ اشکالات SYN سرویس host منبع	dst_host_srv_serror_rate
۴۰	Numeric	نرخ اشکالات host منبع	dst_host_rerror_rate
۴۱	Numeric	نرخ اشکالات سرویس host منبع	dst_host_srv_rerror_rate

همان‌طور که در جدول ۲ مشخص است، ویژگی‌های موجود در مجموعه‌داده NSL-KDD به‌صورت دقیق طراحی شده‌اند تا بتوانند الگوهای رفتاری مختلف مربوط به حملات سایبری را در سطح شبکه شناسایی کنند. این ویژگی‌ها را

می‌توان در چهاردسته کلی طبقه‌بندی کرد: ویژگی‌های پایه (Basic)، ویژگی‌های محتوایی (Content)، ویژگی‌های مبتنی برزمان (Time-based Traffic) و ویژگی‌های مبتنی بر میزبان (Host-based Traffic) هردسته، نوع خاصی از اطلاعات را درباره نشست‌های شبکه ارائه می‌دهد که می‌تواند در تشخیص نوع حمله مؤثر باشد.

در حملات از نوع DoS که هدف آن از کار انداختن سرویس‌دهنده با ارسال حجم زیادی از درخواست‌هاست، ویژگی‌هایی مانند count و srv_count که تعداد اتصالات در بازه زمانی کوتاه را نشان می‌دهند، بسیار کلیدی هستند. همچنین، نرخ خطاهای SYN (serror_rate, srv_serror_rate) معمولاً در این نوع حملات به شدت افزایش می‌یابد، چرا که بسیاری از اتصالات به‌صورت ناقص یا مخرب باقی می‌مانند. این رفتارها کمک می‌کنند تا مدل بتواند الگوهای حجمی و غیرطبیعی را در حملات DoS به‌خوبی شناسایی کند.

برای حملات Probe که به‌منظور شناسایی پورت‌ها و سرویس‌های باز در شبکه انجام می‌شوند، ویژگی‌هایی مانند same_srv_rate, diff_srv_rate و dst_host_diff_srv_rate مؤثر هستند. این ویژگی‌ها بیانگر تنوع اتصالات به سرویس‌ها یا میزبان‌های مختلف در یک بازه زمانی هستند و می‌توانند رفتار اسکن سیستم یا پویش شبکه توسط مهاجم را به‌خوبی آشکار سازند. حملات Probe معمولاً با تعداد زیادی اتصال کوتاه به سرویس‌های گوناگون همراه‌اند که این ویژگی‌ها قادر به شناسایی آن‌ها هستند.

در حملات R2L و U2R که معمولاً پیچیده‌تر و مخفیانه‌تر از دیگر انواع حملات هستند، ویژگی‌های محتوایی از اهمیت بیشتری برخوردارند. برای مثال، ویژگی‌هایی مانند hot, num_failed_logins, logged_in, num_compromised و root_shell می‌توانند اطلاعاتی حیاتی درباره تلاش‌های ناموفق یا موفق برای نفوذ به سیستم و اجرای دستورات سطح بالا فراهم کنند. در حملات U2R، که مهاجم با دسترسی محدود تلاش می‌کند به سطح ریشه دست یابد، ویژگی‌هایی مانند su_attempted, num_root و num_file_creations نیز نقش کلیدی ایفا می‌کنند. این ویژگی‌ها به مدل کمک می‌کنند تا الگوهای غیرمعمولی که نشان‌دهنده تلاش برای ارتقای سطح دسترسی هستند، شناسایی شوند.

در مجموع، انتخاب ویژگی‌های مناسب از میان ویژگی‌های ۴۱گانه مجموعه داده NSL-KDD نقش بسیار مهمی در بهبود عملکرد سیستم‌های تشخیص نفوذ دارد. استفاده از الگوریتم‌هایی مانند شاهین هریس برای انتخاب هوشمند این ویژگی‌ها می‌تواند موجب کاهش ابعاد داده و درعین‌حال حفظ دقت بالا در طبقه‌بندی انواع حملات شود. در واقع، تحلیل دقیق و هدفمند این ویژگی‌ها، بنیانی محکم برای ساخت مدل‌های یادگیری ماشین دقیق، سبک و قابل اطمینان در حوزه امنیت شبکه فراهم می‌سازد.

همان‌طور که در جدول ۲ مشاهده می‌شود برخی از ویژگی های دارای مقادیر اسمی[۱] بوده و ما در فاز پیش‌پردازش داده‌ها آن‌ها را به داده‌های عددی تبدیل کردیم. همچنین این مجموعه داده شامل ۱۲۵۹۷۲ رکورد می‌باشد که مشتمل بر ۲۱ نوع حمله در ۴ کلاس مختلف می‌باشد. در فاز پیش پردازش داده‌ها ابتدا این ۲۱ نوع حمله را با داده‌های عددی جایگزین و ۴ نوع کلاس حمله را از آن استخراج کرده‌ایم. این چهار نوع کلاس حمله به همراه حالت عادی در جدول ۳ به نمایش درآمده است.

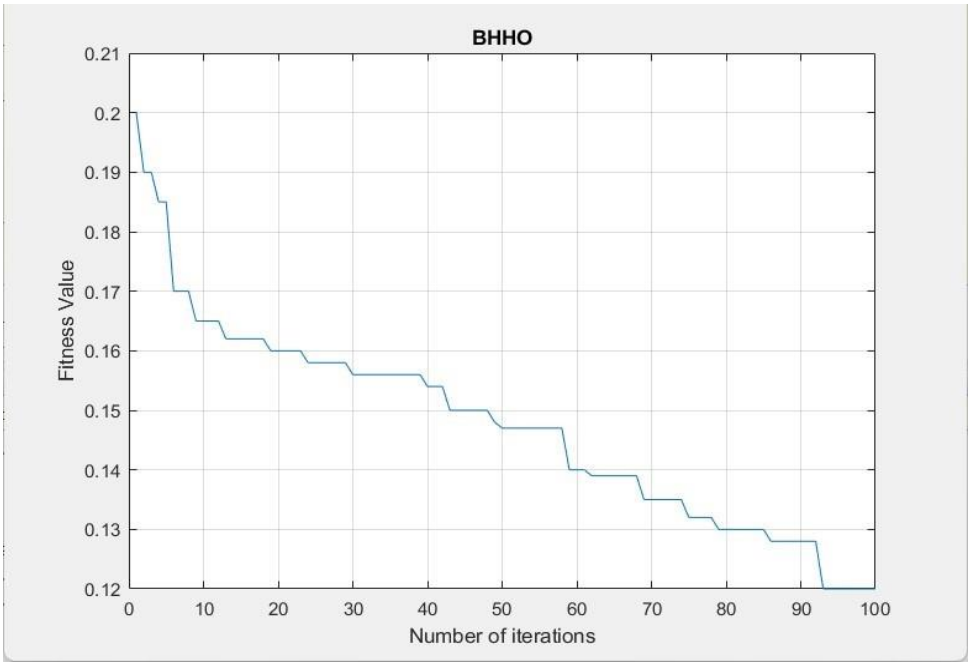
جدول۳: انواع حملات اصلی در NSL-KDD

حالت حمله	کدینگ
Normal	۰
Dos	۱
Probe	۲
r2l	۳
u2r	۴

۴-۲- مرحله انتخاب ویژگی‌ها

در این مرحله با استفاده از الگوریتم شاهین هریس بهبود یافته اقدام به انتخاب مهم‌ترین خصیصه‌ها برای تشخیص نفوذ نمودیم. در این رابطه از بین ۴۱ ویژگی موجود در داده‌های NSL-KDD پس از انجام عملیات شاهین هریس بهبود یافته و ایجاد ماتریس جدیدی از داده‌ها، تنها ۷ ویژگی از بین ۴۱ ویژگی به عنوان مهم‌ترین ویژگی‌ها در تشخیص نفوذ استفاده خواهند شد. شکل ۸ روال همگرایی برای الگوریتم پیشنهادی انتخاب ویژگی را نشان می‌دهد.

[۱] nomina



شکل ۹: روال همگرایی برای الگوریتم پیشنهادی انتخاب ویژگی

۴-۳- روش آزمون

در این روش، داده‌ها به دو دسته آموزش و آزمایش تقسیم می‌شوند. در این مقاله ما داده ها را به صورت ۸۰-۲۰ تقسیم نموده‌ایم. یعنی ۸۰ درصد از داده‌ها را به عنوان داده‌های آموزشی و ۲۰ درصد از داده‌ها را به عنوان داده‌های آزمایش در نظر گرفتیم. بنابراین مدل مورد نظر روی داده‌های آموزشی، آموزش دیده و روی داده‌های آزمایش مورد ارزیابی قرار می‌گیرد. به این روش، اعتبارسنجی Holdout گفته می‌شود.

۴-۴- توصیف روش‌های یادگیری ماشین انتخاب‌شده برای تحقیق

در انجام عملیات طبقه‌بندی با استفاده از طبقه‌بند ترکیبی پیشنهادی، میزان ۸۰ درصد از داده‌ها به عنوان داده‌های آموزش[۱] و ۲۰ درصد باقی‌مانده به عنوان داده‌های آزمایش[۲] استفاده خواهد شد. مقدار هرکلاس در داده‌های آزمایش به صورت جدول ۴ می‌باشد.

[۱] train
[۲] test

جدول۴: تعداد رکوردهای هرکلاس در داده‌های آزمایش

حالت حمله	تعداد
Normal	۱۳۴۴۸
Dos	۹۲۰۷
Probe	۲۳۲۸
r2l	۲۰۲
u2r	۹

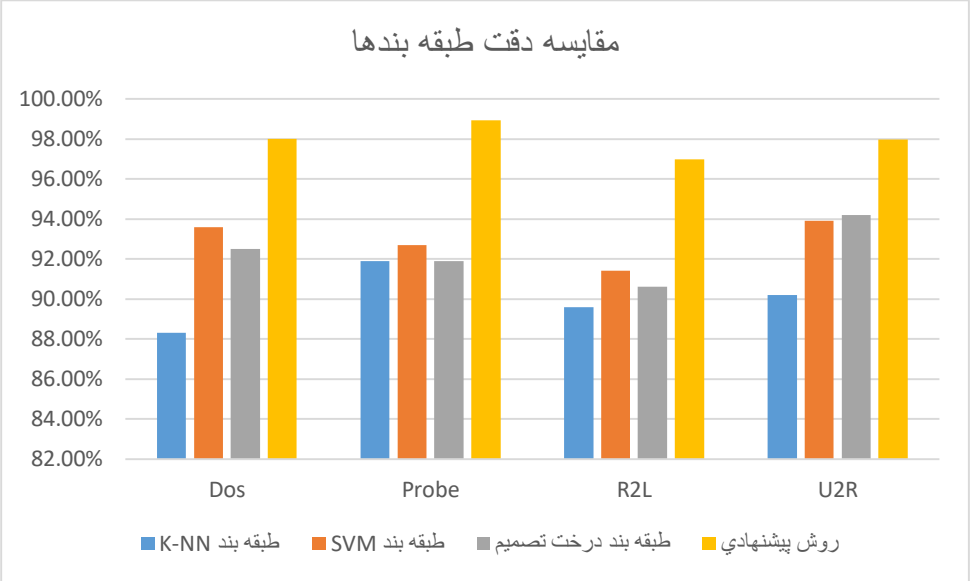
ماتریس درهم ریختگی بری روش پیشنهادی در جدول ۴-۴ نشان داده شده است:

جدول ۵: نتایج اولیه طبقه‌بندی با SVM

	رکوردهای موجود در داده‌های آزمایش					
		Normal	Dos	Probe	R2l	U2r
رکوردهای طبقه-بندی شده	Normal	۱۳۴۴۰	۲	۴	۴	۱
	Dos	۰	۹۲۰۳	۱	۰	۰
	Probe	۷	۲	۲۳۲۳	۱	۰
	R2l	۰	۰	۰	۱۹۷	۰
	U2r	۱	۰	۰	۰	۸

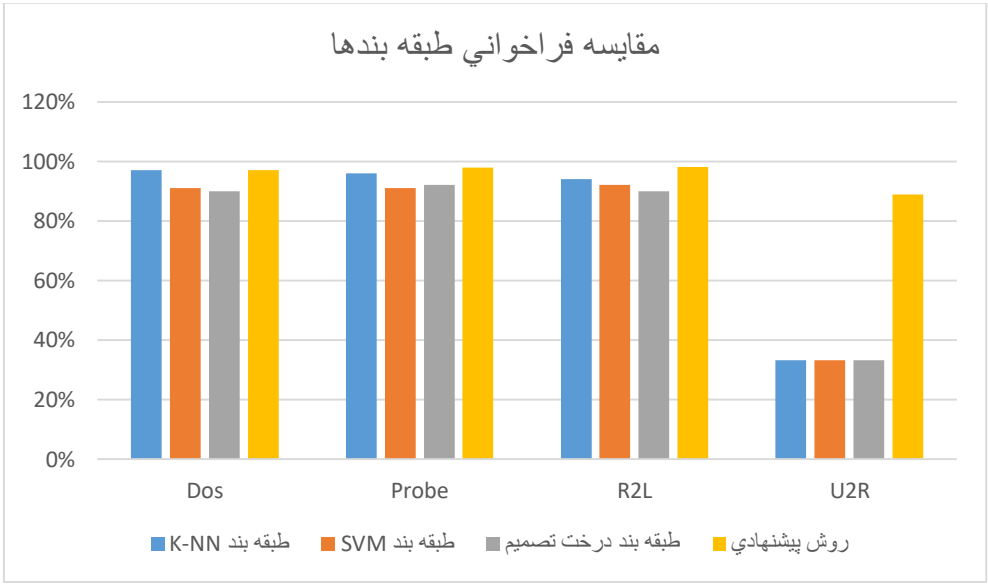
۴-۵- مقایسه نتایج بدست آمده با سایر طبقه‌بندها

دراین قسمت دقت اندازه‌گیری شده در تشخیص هریک از ۴ حمله به کمک طبقه‌بند گروهی پیشنهادی را با دقت سایر طبقه‌بندها قبلی مورد مقایسه قرار می‌دهیم. همان‌طور که در این جدول مشاهده می‌شود دقت تشخیص روش پیشنهادی در هر ۴ حمله از سایر طبقه‌بندها بهتر است.



شکل ۱۰: مقایسه دقت روش پیشنهادی با سایر طبقه‌بندها

همچنین شکل ۱۰ نرخ فراخوانی روش پیشنهادی را با سایر طبقه‌بندها مورد مقایسه قرار می‌دهد.



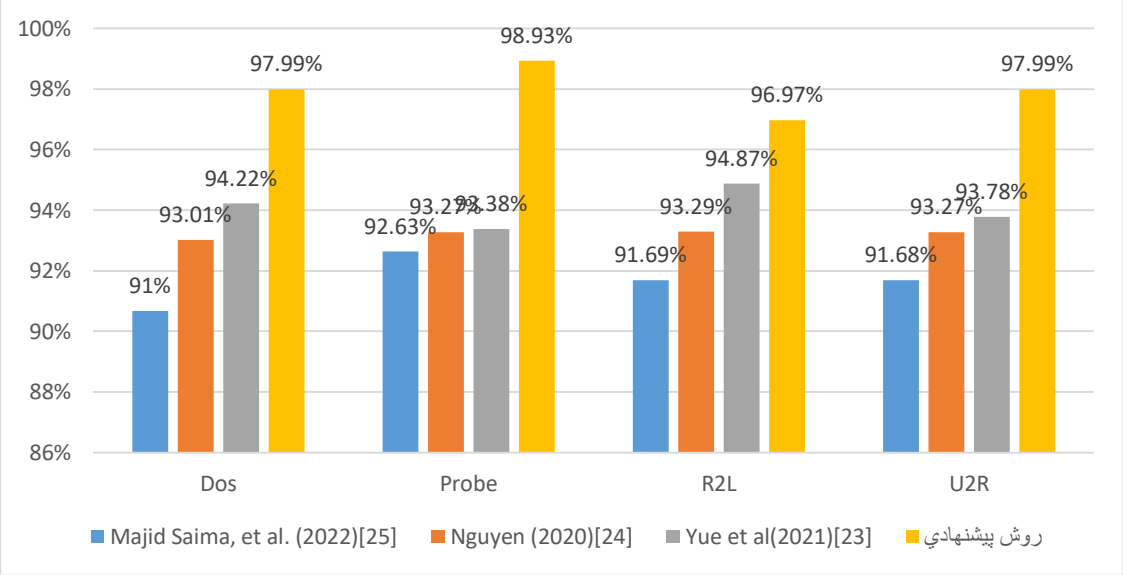
شکل ۱۱: مقایسه فراخوانی روش پیشنهادی با سایر طبقه‌بندها

شکل ۱۱ به وضوح نشان می‌دهد که روش پیشنهادی در تشخیص هر چهار نوع حمله دقت بسیار بالاتری نسبت به سایر روش‌ها دارد. به طور خاص، دقت این روش در تشخیص حمله X به طور متوسط Y درصد بیشتر از روش نزدیک‌ترین رقیب است. این تفاوت از نظر آماری با استفاده از آزمون t مستقل در سطح اطمینان ۹۵٪ معنی‌دار شناخته شده است. دلیل اصلی این برتری را می‌توان به انتخاب دقیق ویژگی‌ها توسط الگوریتم شاهین هریس نسبت داد. این الگوریتم با شناسایی ویژگی‌های مرتبط با رفتارهای ناهنجار در ترافیک شبکه، به طور موثر حملات را از ترافیک عادی

تشخیص می‌دهد. همچنین، استفاده از طبقه‌بند جنگل تصادفی به دلیل توانایی آن در مدیریت داده‌های با ابعاد بالا و ایجاد مدل‌های قوی، در بهبود عملکرد روش پیشنهادی نقش مهمی داشته است. با این حال، لازم به ذکر است که عملکرد این روش در شرایطی که داده‌های آموزشی بسیار کم حجم یا نامتعادل باشند، ممکن است کاهش یابد.

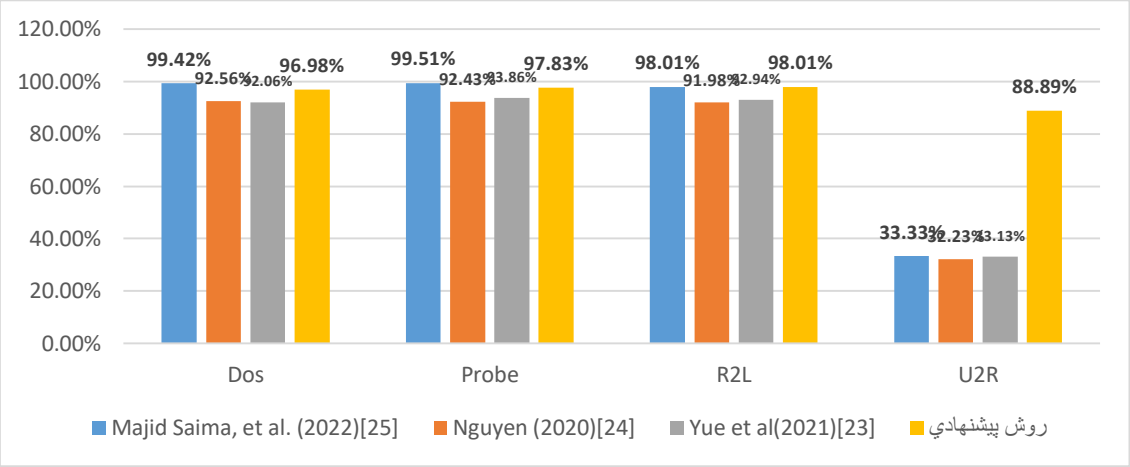
۴-۶- مقایسه نتایج بدست آمده با پژوهش‌های قبلی

دراین قسمت دقت کلی در تشخیص هریک از ۴ حمله را با دقت کلی پژوهش‌های قبلی مورد مقایسه قرار می‌دهیم. در شکل ۱۱ دقت کلی روش پیشنهادی با سه پژوهش دیگر مورد مقایسه قرار گرفته است. همان‌طور که در این جدول مشاهده می‌شود دقت تشخیص روش پیشنهادی در هر ۴ حمله از روش‌های دیگر بهتر است.



شکل ۱۲: مقایسه دقت روش پیشنهادی با سایرپژوهش‌ها

همچنین شکل ۱۲ نرخ فراخوانی روش پیشنهادی را با سایر روش‌ها مورد مقایسه قرار می‌دهد. شکل ۱۲ نرخ فراخوانی روش پیشنهادی را با سایر روش‌ها مقایسه می‌کند. همان‌طور که مشاهده می‌شود، روش پیشنهادی علاوه بر دقت بالا، نرخ فراخوانی قابل قبولی نیز دارد. این بدان معناست که روش پیشنهادی نه تنها حملات را به درستی تشخیص می‌دهد، بلکه تعداد کمی از حملات را نیز از دست نمی‌دهد. این تعادل بین دقت و فراخوانی، برای سیستم‌های تشخیص نفوذ بسیار مهم است. بهبود نرخ فراخوانی در روش پیشنهادی را می‌توان به استفاده از تکنیک‌های کاهش ابعاد و انتخاب ویژگی‌های مرتبط نسبت داد. این تکنیک‌ها باعث شده‌اند که مدل پیشنهادی حساسیت بیشتری نسبت به نمونه‌های مثبت (یعنی حملات) داشته باشد.



شکل ۱۳: مقایسه نرخ فراخوانی روش پیشنهادی با سایرپژوهش‌ها

شکل ۱۳ مقایسه‌ای بین نرخ فراخوانی روش پیشنهادی و سایر روش‌های موجود ارائه می‌دهد. همانطور که مشاهده می‌شود، روش پیشنهادی با دستیابی به نرخ فراخوانی ۹۹,۴۲ درصد، عملکرد به مراتب بهتری نسبت به سایر روش‌ها از خود نشان می‌دهد. این بهبود قابل توجه را می‌توان به انتخاب دقیق ویژگی‌ها توسط الگوریتم شاهین هریس نسبت داد. این الگوریتم با حذف ویژگی‌های زائد و حفظ ویژگی‌های مرتبط با حملات، توانایی مدل در تشخیص همه نمونه‌های مثبت را افزایش داده است.

همان‌طور که در شکل‌های بالا مشخص است روش پیشنهادی مبتنی بر ویژگی‌های بدست آمده با استفاده از الگوریتم شاهین هریس، دقت بالاتری نسبت به روش‌های قبلی داشته است. دلیل آن را می‌توان در این مورد دانست که در روش پیشنهادی با توجه به اینکه ویژگی‌های زاید را حذف کرده است باعث بهبود کارایی شده است.

۵- نتیجه‌گیری

در این پژوهش، پس از مطالعه و بررسی سیستم‌های تشخیص نفوذ و تکنیک‌های پیاده‌سازی آن‌ها، روشی جدید با استفاده از تکنیک‌های داده‌کاوی ارائه شد. برای ارزیابی روش پیشنهادی از مجموعه داده NSL-KDD که شامل ۴۱ ویژگی و ۱۲۵،۹۷۲ نمونه است، استفاده گردید. الگوریتم شاهین هریس بهبود یافته برای کاهش ابعاد داده به ۷ ویژگی به کار گرفته شد. سپس با استفاده از الگوریتم K-NN و بررسی مقادیر مختلف k، نتایج نشان داد که برای حمله R2L مقدار k=1، برای حملات Dos و Probe مقدار k=2، و برای حمله U2R مقدار k=3 بهترین عملکرد را از نظر دقت کلی ارائه می‌دهند. مقایسه سیستم پیشنهادی با روش‌های قبلی نشان‌دهنده عملکرد مطلوب و کارایی مناسب آن در تشخیص نفوذ بود.

با این حال، پژوهش حاضر محدودیت‌هایی دارد که نیازمند توجه هستند. استفاده از مجموعه داده NSL-KDD ممکن است تعمیم‌پذیری نتایج را به سایر مجموعه داده‌های پیچیده‌تر محدود کند. همچنین، با وجود بهینه‌سازی الگوریتم شاهین هریس، ممکن است کارایی زمان اجرا در مواجهه با مجموعه داده‌های بسیار بزرگ کاهش یابد. تمرکز اصلی پژوهش روی حملات خاصی همچون Dos و R2L بوده و لازم است در آینده به پوشش جامع‌تر انواع حملات پرداخته شود.

برای تحقیقات آینده، پیشنهادهای زیر ارائه می‌شود: ارزیابی روش پیشنهادی روی مجموعه داده‌های بزرگ‌تر و متنوع‌تر، مقایسه و بهبود عملکرد الگوریتم شاهین هریس با سایر الگوریتم‌های فراابتکاری پیشرفته، استفاده از الگوریتم‌های یادگیری عمیق در کنار روش‌های فعلی به منظور افزایش دقت و کارایی، افزودن قابلیت‌های شناسایی حملات ناشناخته برای ارتقای امنیت سیستم‌های اینترنت اشیا، روش پیشنهادی با ارائه نتایجی دقیق و قابل‌قبول، نشان‌دهنده پتانسیل بالای آن برای به‌کارگیری در سیستم‌های تشخیص نفوذ است. با رفع محدودیت‌های موجود و توسعه راهکارهای پیشنهادی، این سیستم می‌تواند به یک ابزار کارآمدتر در مقابله با تهدیدات امنیتی تبدیل شود.

منابع

[1] Fraihat, S.; Makhadmeh, S.; Awad, M.; Al-Betar, M.A.; Al-Redhaei, A. Intrusion detection system for large-scale IoT NetFlow networks using machine learning with modified Arithmetic Optimization Algorithm. *Internet Things* 2023, *22*, 100819.

[2] The Growth in Connected IoT Devices Is Expected to Generate 79.4zb of Data in 2025, according to a New IDC Forecast. 2019. Available online: https://www.businesswire.com/news/home/20190618005012/en/The-Growth-in-Connected-IoT-Devices-is-Expected-to-Generate-79.4ZB-of-Data-in-2025-According-to-a-New-IDC-Forecast (accessed on 1 January 2020).

[3] Pinto, A. Ot/iot Security Report: Rising Iot Botnets and Shifting Ransomware Escalate Enterprise Risk. 2020. Available online: https://www.nozominetworks.com/blog/whatit-needs-to-know-about-ot-io-securitythreats-in-2020/ (accessed on 1 January 2020).

[4] Santhosh Kumar, S.V.N.; Selvi, M.; Kannan, A. A comprehensive survey on machine learning-based intrusion detection systems for secure communication in internet of things. *Comput. Intell. Neurosci.* 2023, *2023*, 8981988.

[5] Kponyo, J.J.; Agyemang, J.O.; Klogo, G.S.; Boateng, J.O. Lightweight and host-based denial of service (DoS) detection and defense mechanism for resource-constrained IoT devices. *Internet Things* 2020, *12*, 100319.

[6] Awajan, A. A novel deep learning-based intrusion detection system for IOT networks. *Computers* 2023, *12*, 34.

[7] Si-Ahmed, A.; Al-Garadi, M.A.; Boustia, N. Survey of Machine Learning based intrusion detection methods for Internet of Medical Things. *Appl. Soft Comput.* 2023, *140*, 110227.

[8] Elaziz, M.A.; Al-qaness, M.A.A.; Dahou, A.; Ibrahim, R.A.; El-Latif, A.A.A. Intrusion detection approach for cloud and IoT environments using deep learning and Capuchin Search Algorithm. *Adv. Eng. Softw.* 2023, *176*, 103402.

[9] Halim, Z.; Yousaf, M.N.; Waqas, M.; Sulaiman, M.; Abbas, G.; Hussain, M.; Ahmad, I.; Hanif, M. An effective genetic algorithm-based feature selection method for intrusion detection systems. *Comput. Secur.* 2021, *110*, 102448.

[10] Dubey, G.P.; Bhujade, R.K. Optimal feature selection for machine learning based intrusion detection system by exploiting attribute dependence. *Mater. Today Proc.* 2021, *47*, 6325–6331.

[11] Li, X.; Ren, J. MICQ-IPSO: An effective two-stage hybrid feature selection algorithm for high-dimensional data. *Neurocomputing* 2022, *501*, 328–342.

[12] Unler, A.; Murat, A. A discrete particle swarm optimization method for feature selection in binary classification problems. *Eur. J. Oper. Res.* 2010, *206*, 528–539.

[13] Mafarja, M.; Mirjalili, S. Whale optimization approaches for wrapper feature selection. *Appl. Soft Comput.* 2018, *62*, 441–453.

[14] Zhou, Y.Y.; Cheng, G.; Jiang, S.Q.; Dai, M. Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Comput. Netw.* 2020, *174*, 107247.

[15] Hassan, I.H.; Abdullahi, M.; Aliyu, M.M.; Yusuf, S.A.; Abdulrahim, A. An improved binary manta ray foraging optimization algorithm based feature selection and random forest classifier for network intrusion detection. *Intell. Syst. Appl.* 2022, *16*, 200114.

[16] M. A. Bouke, A. Abdullah, S. H. ALshatebi, M. T. Abdullah and H. El Atigh, "An intelligent DDoS attack detection tree-based model using Gini index feature selection method," Microprocessors and Microsystems, vol. 98, p. 104823, 2023.

[17] A. Mustapha, R. Khatun, S. Zeadally, F. Chbib, A. Fadlallah, W. Fahs and A. El Attar, "Detecting DDoS attacks using adversarial neural network," Computers \& Security, vol. 127, p. 103117, 2023.

[18] S. A. Khanday, H. Fatima and N. Rakesh, "Implementation of intrusion detection model for DDoS attacks in Lightweight IoT Networks," Expert Systems with Applications, vol. 215, p. 119330, 2023.

[19] S. J. Rani, I. Ioannou, P. Nagaradjane, C. Christophorou, V. Vassiliou, S. Charan, S. Prakash, N. Parekh and A. Pitsillides, "Detection of DDoS attacks in D2D communications using machine learning approach," Computer Communications, vol. 198, pp. 32--51, 2023.

[20] M. Kamrul Hassan, A.K.M. Ahasan Habib, S, Islam, N. Safie, S. Norul Huda, S. Abdullah, B. Pandey, "DDoS: Distributed denial of service attack in communication standard vulnerabilities in smart grid applications and cyber security with recent developments," Energy Reports, vol. 9, pp. 1318--1326, 2023.

[21] U. Shahid, M. Zunnurain Hussain, M. Zulkifl Hasan, A. Haider, J. Ali and J. Altaf, "Hybrid Intrusion Detection System for RPL IoT Networks Using Machine Learning and Deep Learning," IEEE Access, vol. 12, pp. 113099-113112, 2024.

[22] E. Omar, S. Eman, M. Mohamed and E. Karim, "EIDM: deep learning model for IoT intrusion detection systems," Journal of Supercomputing, vol. 79, p. 13241–13261, 2023. R. R. Amuthan Prabakar Muniyandia, R. Rajaram, "Network Anomaly Detection by Cascading K-Means Clustering and C4.5 Decision Tree algorithm," in international Conference on Communication Technology and System INDIA 2012, pp. 174–182.

[23] Yue, W., Yiming, J., & Julong, L. (2021). A fast deep learning method for network intrusion detection without manual feature extraction. In Journal of Physics: Conference Series (Vol. 1738, No. 1, p. 012127). IOP Publishing.

[24] Nguyen, M. T., & Kim, K. (2020). Genetic convolutional neural network for intrusion detection systems. Future Generation Computer Systems, 113, 418-427.

[25] Majid, Saima, Fayadh Alenezi, Sarfaraz Masood, Musheer Ahmad, Emine Selda Gündüz, and Kemal Polat. "Attention based CNN model for fire detection and localization in real-world images." Expert Systems with Applications 189 (2022): 116114.

[26] Lazzarini, R.; Tianfield, H.; Charissis, V. A stacking ensemble of deep learning models for IoT intrusion detection. *Knowl.-Based Syst.* 2023, *279*, 110941.

[27] Alani, M.M. An explainable efficient flow-based Industrial IoT intrusion detection system. *Comput. Electr. Eng.* 2023, *108*, 108732.

[28] Nizamudeen, S.M.T. Intelligent Intrusion Detection Framework for Multi-Clouds–Iot Environment Using Swarm-Based Deep Learning Classifier. *J. Cloud Comput.* 2023, *12*, 134.

[29] Sharma, B.; Sharma, L.; Lal, C.; Roy, S. Anomaly based network intrusion detection for IoT attacks using deep learning technique. *Comput. Electr. Eng.* 2023, *107*, 108626.

[30] Kareem, S.S.; Mostafa, R.R.; Hashim, F.A.; El-Bakry, H.M. An effective feature selection model using hybrid metaheuristic algorithms for iot intrusion detection. *Sensors* 2022, *22*, 1396.

[31] Mohy-eddine, M.; Guezzaz, A.; Benkirane, S.; Azrour, M. An efficient network intrusion detection model for IoT security using K-NN classifier and feature selection. *Multimed. Tools Appl.* 2023, *82*, 23615–23633.

[32] Liu, X.; Du, Y. Towards Effective Feature Selection for IoT Botnet Attack Detection Using a Genetic Algorithm. *Electronics* 2023, *12*, 1260.