«علوم و فناوری‌های پدافند نوین»
سال پانزدهم، شماره ۲،تابستان۱۴۰۳؛ ص ۹۳-۸۵

علمی- پژوهشی

سیستم امن اینترنت اشیا با استفاده از الگوریتم سنجاقک برای انتخاب ویژگی و تقویت گرادیان برای طبقه‌بندی

علی‌اکبر تجری سیاه مرز کوه

استادیار،دانشگاه گلستان، گرگان، ایران

(دریافت: ۱۴۰۳/۰۲/۲۸، بازنگری: ۱۴۰۳/۰۳/۱۳، پذیرش: ۱۴۰۳/۰۵/۰۷، انتشار: ۱۴۰۳/۰۶/۰۱)
DOR: https://dor.isc.ac/dor/20.1001.1.26762935.1403.15.2.3.2

چکیده

با پیشرفت روزافزون فن آوری اینترنت اشیا و لزوم برقراری ارتباط امن بین دستگاه‌های متصل به یکدیگر، موضوع امنیت داده‌های مبادله شده به یک مسئلهٔ حیاتی تبدیل شده است. مهاجمان و رخنه‌گرهای سایبری بانفوذ به سیستم‌ها و دست‌کاری آنها همواره سعی در ایجاد مشکل در شبکه‌های اینترنت اشیا دارند. ازاین‌رو، تحقیقات زیادی برای مقابله با این تهدیدات توسط بسیاری از محققان انجام شده است. دقت و زمان اجرای روش‌های موجود بهینه نیست و نیاز به یک روش ترکیبی که علاوه بر افزایش دقت، زمان مورد نیاز را نیز کاهش دهد امری ضروری به نظر می‌رسد. در این مقاله، روشی ترکیبی با استفاده از الگوریتم سنجاقک و تقویت گرادیان برای بهبود دقت سیستم اینترنت اشیا ارائه شده است. الگوریتم سنجاقک با بهبود راه‌حل‌های جدید و افزایش تنوع جمعیت منجر به حذف ویژگی‌های غیرمؤثر می‌شود که این امر، داده‌ها را برای دستیابی به دقت‌های بالا در طبقه‌بندی آماده می‌کند، به‌طوری‌که آن دسته از ویژگی‌هایی که منجر به کاهش دقت طبقه‌بندی می‌شوند کنار گذاشته می‌شوند. پس از آن با استفاده از تقویت گرادیان که عاملی با سرعت و دقت بالا برای تشخیص (طبقه‌بندی) است، عملیات طبقه‌بندی داده‌ها که فاز اصلی تشخیص حمله است انجام می‌شود. نتایج شبیه‌سازی نشان می‌دهد که دقت طبقه‌بندی در حالت دودویی و چند رده‌بندی‌شده به ترتیب برابر با ۹۹/۹۹۴ درصد و ۹۹/۹۹۲ درصد است که نشان‌دهندهٔ برتری روش پیشنهادی نسبت به سایر روش‌های قبلی ارائه‌شده است.

کلیدواژه‌ها: تشخیص نفوذ، الگوریتم سنجاقک، یادگیری ماشین، انتخاب ویژگی، تقویت گرادیان.

Internet of Things Secure System Using Dragonfly Algorithm for Feature Selection and Gradient Boosting for Classification

A.A. Tajari Siahmarzkooh
Golestan University

(Received:2024/05/17, Revised: 2024/06/02, Accepted: 2024/07/28, Published: 2024/08/22)

Abstract

With the increasing progress of Internet of Things technology and the need to establish secure communication between devices connected to each other, the security of exchanged data has become a critical issue. Attackers and cyber hackers always try to create problems in Internet of Things networks by infiltrating systems and manipulating them. Therefore, a lot of research has been done by many researchers to deal with these threats. The accuracy and execution time of the existing methods are not optimal, and the need for a combined method that, in addition to increasing the accuracy, also reduces the required time seems essential. In this paper, a hybrid method using the dragonfly algorithm and gradient boosting is presented to improve the accuracy of the Internet of Things system. The dragonfly algorithm leads to the elimination of ineffective features by improving new solutions and increasing population diversity, which prepares the data to achieve high accuracy in classification. So that those features that lead to a decrease in classification accuracy are left out. After that, by using gradient boosting, which is a factor with high speed and accuracy for detection (classification), the data classification operation is performed, which is the main phase of attack detection. The simulation results show that the classification accuracy in binary and multi-class mode is equal to 99.994% and 99.992%, respectively, which indicates the superiority of the proposed method over other previous methods.

Keywords: Intrusion Detection, Dragonfly Algorithm, Machine Learning, Feature Selection, Gradient Boosting.

*Corresponding Author E-mail: a.tajari@gu.ac.ir

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Publisher: Imam Hussein University © Authors

۱. مقدمه

درک مفاهیم شبکهٔ اینترنت اشیا به دلیل گستردگی آن می‌تواند چالش‌برانگیز باشد. این موضوع نشان‌دهندهٔ یک تحول قابل‌توجه در نوآوری دیجیتال است که فراتر از اتصال دستگاه‌های منفرد مانند تبلت‌ها، رایانه‌ها و تلفن‌های همراه است و اکنون تعامل بین اشیای فیزیکی مختلف را امکان‌پذیر می‌کند و مرزهای بین دستگاه‌های ملموس و دنیای مجازی را حذف می‌کند [۱].

اجزای پیچیده و به‌هم‌پیوسته با ضعف‌ها و چالش‌های امنیتی بالقوه مواجه هستند. این اجزا، اهداف اصلی برای نفوذهای سایبری مختلف، به‌ویژه آنهایی هستند که هدف آنها ایجاد اختلال در مسیرهای ارتباطی است. متأسفانه، تنها اتکا به روش‌های رمزگذاری و وارسی، همیشه امنیت این سیستم‌های به‌هم‌پیوسته را تضمین نمی‌کند [۱].

با ادامهٔ پیشرفت دستگاه‌ها، نیاز به امنیت به طور فزاینده‌ای مهم شده است. به طور خاص، سیستم‌های شبکهٔ اینترنت مستقیماً به وب مرتبط هستند و اطلاعات خود را تحت یک سطح اطمینان مفروض به اشتراک می‌گذارند که آن را در برابر تهدیدات آنلاین مختلف از جمله حملاتی مانند DDoS آسیب‌پذیر می‌کند. برای رسیدگی به این مسائل، ادغام یک سیستم تشخیص نفوذ (IDS) برای حفظ امنیت اینترنت اشیا ضروری است [۲]. سیستم تشخیص نفوذ، مکانیزمی است که به‌سرعت به محافظت از داده-های حساس از طریق استفاده از فناوری‌های هوش مصنوعی پاسخ می‌دهد که می‌تواند طیف وسیعی از حملات را شناسایی و از آن جلوگیری کند.

همان‌طور که گفته شد، سیستم‌هایی که برای شناسایی فعالیت‌های غیرمجاز در شبکه‌های اینترنت اشیا طراحی شده‌اند، به نام IDS شناخته می‌شوند که به شناسایی هرگونه تهدید بالقوه کمک می‌کنند. این سیستم‌ها می‌توانند در سطح شبکه یا میزبان عمل کنند و پوشش امنیتی جامعی را ارائه دهند. در پاسخ به خطرات امنیتی فزایندهٔ مرتبط با دستگاه‌ها، ابزارهای IDS برای نظارت بر الگوهای ارتباطی این دستگاه‌ها و شناسایی هرگونه رفتار مشکوکی که ممکن است نشان‌دهندهٔ حمله سایبری باشد، مستقر می‌شوند. دو روش اصلی تشخیص حمله که توسط IDS استفاده می‌شوند روش‌های تشخیص مبتنی بر امضا و مبتنی بر ناهنجاری هستند [۲].

روش مبتنی بر امضا: در این روش، مجموعه‌ای از خطرات شناخته‌شده و نشانه‌های مشکوک آنها از قبل تنظیم شده است، توالی بایت‌های شناسایی‌شده، شناسایی فایل، فضای اینترنتی مضر یا اقدامات خاصی که مرتباً با حمله به شبکهٔ متخاصم همراه است، همگی می‌توانند نشانه‌هایی از نفوذ باشند. یک سیستم تشخیص نفوذ با استفاده از یک روش مبتنی بر امضا، بسته‌ها را با پایگاه داده‌ای از نام‌های حملهٔ شناخته‌شده در هنگام حرکت در سراسر شبکه جمع‌آوری می‌کند و مراقب هرگونه رفتار مشکوکی است [۳].

روش مبتنی بر ناهنجاری: در سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری، الگوی استاندارد رفتارهای معمول با استفاده از دستورالعمل‌های یادگیری ماشین شناسایی می‌شود. عملیات شبکه متعاقباً بر اساس الگوی رفتار استاندارد ارزیابی می‌شود. سیستم‌های تشخیص نفوذ مبتنی بر ناهنجاری به طور مجزا هرگونه فعالیت غیرمعمول را به‌عنوان ناهنجاری تشخیص می‌دهند [۳].

IDS لایه‌های امنیتی پیشرفته‌ای را برای دستگاه متصل در اینترنت اشیا ارائه می‌دهد. به لطف سیستم تشخیص نفوذ، مدیران و سرپرستان سیستم می‌توانند اطلاعات انتقال‌یافته توسط این ساختارهای داده را تنظیم و درک کنند. این سیستم، مشکلات احتمالی را قبل از اینکه آسیبی وارد شود به طور کارآمد نشان می‌دهد. IDS می‌تواند انواع مختلفی از حملات را شناسایی کند، از جمله حملات Denial of Service (DoS)، حملهٔ از راه دور محلی (R2L)، کاربر به ریشه (U2R) و حملات Probing. هدف حملات DoS ایجاد اختلال در سرویس‌ها و ممانعت از دسترسی کاربران قانونی به آنها و ایجاد ازدحام شبکه است. حملات R2L شامل دسترسی غیرمجاز از راه دور به یک سیستم هدف است. حملات U2R زمانی اتفاق می‌افتد که از آسیب‌پذیری‌های برنامه‌نویسی برای به‌دست‌آوردن امتیازات سیستم سوءاستفاده می‌شود. حملات کاوشگر شامل قراردادن یک کاوشگر در نزدیکی یک قطعه مدار الکترونیکی برای نظارت بر رفتار الکتریکی آن است [۴].

بسیاری از محققان از روش‌های فراابتکاری و یادگیری ماشین برای تقویت سیستم‌های تشخیص نفوذ که قادر به شناسایی حملات در محیط شبکه هستند استفاده می‌کنند [۶-۴]. در این مقاله، برای حذف ویژگی‌هایی که در شناسایی داده‌های ناهنجار مؤثر نیستند از الگوریتم سنجاقک استفاده شده است. این الگوریتم به دلیل برقراری تعادل بهتر بین اکتشاف و استخراج ویژگی‌ها منجر به ارائهٔ راه‌حل‌های بهتر با پیچیدگی زمانی کمتر می‌شود و ازاین‌رو در فاز انتخاب ویژگی‌ها از آن استفاده می‌شود. در ادامه، از روشی مبتنی بر یادگیری ماشین به نام تقویت گرادیان برای بهبود دقت طبقه‌بندی سیستم پیشنهادی استفاده شده است. همچنین، یک ارزیابی مقایسه‌ای با چند روش شناخته‌شدهٔ مبتنی بر یادگیری ماشین انجام شده است تا اثباتی بر کارآیی روش پیشنهادی ارائه شود. لازم به ذکر است که متداول‌ترین روش‌های یادگیری ماشین را روی دو مجموعهٔ داده استاندارد ارزیابی شده است. انتظار می‌رود نتایج آزمایش باتکیه‌بر قدرت الگوریتم‌های استفاده‌شده از لحاظ دقت و زمان اجرا، به‌مراتب بهتر از سایر روش‌های مبتنی بر یادگیری ماشین عمل نماید.

بخش‌های بعدی این مقاله به شرح زیر دسته‌بندی شده‌اند. در بخش ۲ روش‌های ارائه‌شده توسط محققان برای اجرای استراتژی‌های یادگیری ماشین در IDS تشریح شده است. بخش ۳ شامل رویکرد پیشنهادی مقاله است که در آن طیفی از روش‌های فراابتکاری و یادگیری ماشین در مجموعه داده‌های مشابه تحت طبقه‌بندی‌های دودویی و چند کلاسه اعمال شده است. بخش ۴ شامل نتایج آزمایش است و در نهایت نتیجه‌گیری ارائه شده است.

۲. پیشینه تحقیق

چالش‌های امنیتی به دلیل افزایش قابل‌توجه جریان داده‌ها از طریق دستگاه‌های مختلف شبکهٔ اینترنت اشیا و پروتکل‌های ارتباطی افزایش‌یافته است که بر ضرورت وجود سیستم‌های تشخیص نفوذ کارآمد و سفارشی‌شده تأکید می‌کند. محققان به ارائه روش‌های یادگیری ماشین متنوع و روش‌های یادگیری عمیق برای IDS پرداخته‌اند. این بخش به مطالعات اخیری می‌پردازد که در سال‌های گذشته برای ارائهٔ راه‌حل‌های نوآورانهٔ IDS فرموله-سازی شده‌اند.

چندین روش یادگیری عمیق برای تشخیص نفوذ در حوزه امنیت اینترنت اشیا به‌کارگرفته‌شده است. وان هونگ و همکاران [۷]، روشی ارائه کردند که در آن داده‌های گزارش‌شده از یک شبکه شامل جزئیاتی مانند آدرس، مکان و سرویس در یک مجموعه‌داده ثبت می‌شود. پس از پیش‌پردازش و تبدیل داده‌ها به یک ماتریس شبیه به یک تصویر، یک شبکهٔ عصبی کانولوشن (CNN) برای آموزش بادقت متوسط ۹۸/۹٪ ساخته می‌شود. وو و همکاران [۸،]روش جدید تشخیص نفوذ را با استفاده از CNN معرفی کردند. این مدل CNN سطوح دقت را با کمترین خطا تقویت می‌کند و هم‌زمان نرخ هشدار کاذب را کاهش می‌دهد. یک رویکرد مبتنی بر یادگیری عمیق برای تشخیص نفوذ شبکهٔ اینترنت اشیا در کار گارا نگ و همکاران باهدف تمایز بین اتصالات عادی و نفوذ ارائه شده است [۹]، جایی که سیستم از یک شبکهٔ عمیق برای آموزش بر روی الگوهای ناهنجاری استفاده کرده است. رویکرد آنها همچنین تلاش می‌کند تا نرخ هشدار نادرست را تا حد ممکن برای مجموعه‌داده NSL-KDD بادقت ۸۷/۲٪ به حداقل برساند. تسوگباتر و همکاران [DeL-IoT۱۰] را به‌عنوان یک ساختار تشخیص نفوذ با استفاده از شبکه‌های تعریف‌شده با نرم-افزار (SDN) تشریح کردند. ویژگی‌های قابل‌تشخیص از طریق رمزگذارهای خودکار عمیق و انباشته استخراج شدند. مدل ارائه‌شده، قابلیت اطمینان بالایی را در شناسایی حملات اینترنت اشیا نشان داده است و دامنهٔ دقت بین ۹۹/۵٪ الی ۹۹/۹٪ و محدودهٔ پارامتر ارزیابی همبستگی ماتیو (MCC) بین ۹۱,۰۴٪ الی ۹۹,۹۵٪ را نشان می‌دهد.

اولاه و همکاران [۱۱]، اثربخشی استفاده از الگوریتم‌های مختلف یادگیری عمیق را در تشخیص ناهنجاری‌ها در شبکه‌های اینترنت اشیا بررسی کردند. تمرکز آنها در درجهٔ اول، حول استفاده از روش‌های پیشرفتهٔ CNN بود. برای ارزیابی سیستم تشخیص نفوذ پیشنهادی، مجموعه‌های مختلف IDS مرتبط با اینترنت اشیا مانند BoTIoT، IoT-DS-2، IoT-23، و MQTT-IoT-IDS2020 مورداستفاده قرار گرفتند. این چارچوب پیچیده با موفقیت بالایی، حملات بالقوه را شناسایی نموده است.

در مقاله دیگری توسط روپاک و همکاران [۱۲]، یک رویکرد جدید برای تشخیص نفوذ با ترکیب مدل‌های CNN با حافظه طولانی کوتاه‌مدت (LSTM) در مجموعه‌داده CISIDS2017 معرفی شده است. آنها برای بهبود انتخاب ویژگی، روش بهینه‌سازی الگوریتم ژنتیک مرتب‌سازی بدون مالکیت (NSGA) را انتخاب کردند. علاوه بر این، سایر محققان به توسعهٔ سیستم‌های تشخیص نفوذ مبتنی بر یادگیری عمیق برای تجزیه‌وتحلیل ترافیک چندوجهی پرداختند. مدل‌های آنها نرخ‌های دقت قابل‌توجهی را مانند ۹۶/۶۹٪ و ۹۵/۹۲٪ به ترتیب برای مجموعه‌داده‌های UNSW-NB15 و CICIDS2017 نشان داد که بر قابلیت‌های تشخیص چند کلاسه قوی تأکید داشت.

بسیاری از محققان بر روی توسعهٔ IDS متمرکز شده‌اند که در این راستا بر روی ابزارها برای شناخت روش‌ها سرمایه‌گذاری کردند. به‌عنوان‌مثال، ساهید و همکاران [۱۳]به استفاده از سیستم تشخیص نفوذی که کاملاً مبتنی بر شناخت ابزار برای درک تهدیدهای جامعه اینترنت اشیا در مجموعه‌داده UNSW-NB15 است پرداختند که در نتیجه‌گیری‌ها به‌دقت ۹۹/۹۹٪ و ۹۹/۹۷٪ در پارامتر MCC دست یافتند.

بهات و همکاران [۱۴]، یک روش تشخیص حمله را ارائه دادند که کاملاً بر اساس الگوریتم‌های یادگیری ماشین برای تشخیص ناهنجاری در یک ماژول است. جنگل تصادفی، نقشه خودسازمان-دهی (SOM)، ماشین بردار پشتیبان یک کلاسه (OCSVM Gaussian) و مدل‌سازی مخلوط (GMM)، اجزای این روش ترکیبی هستند که دقت ۹۸ درصدی را ارائه می‌کند. شفیق و همکاران [۱۵]، یک روش انتخاب مشخصهٔ کاملاً جدید به نام CorrAUC را توسعه دادند که اساساً بر پایهٔ سطح زیر منحنی (AUC) و روش بسته‌بندی بنا شده است. درخت C4.5، بیزین ساده، جنگل تصادفی و ماشین بردار پشتیبان بر روی یک مجموعه‌داده اعمال شده‌اند. روش پیشنهادی آنها منجر به بیش از ۹۶٪ دقت روی مجموعه‌داده Bot-IoT شده است.

سرور و همکاران [۱۶]، یک روش پیچیده برای طراحی IDS با طبقه‌بندی‌کنندهٔ جنگل تصادفی (RF) و الگوریتم بهینه‌سازی ازدحام ذرات (PSO) برای انتخاب تابع ارائه کردند. این روش به‌دقت ۹۸٪ برای دسته‌بندی دودویی و ۸۳٪ برای دسته چند کلاسه در مجموعه‌داده IoTID20 دست می‌یابد.

باگا و همکاران [۱۷]سه مرحلهٔ کلیدی را اجرا کردند: پیش‌پردازش، انتخاب تابع، مدل‌سازی و ارزیابی IDS پیشنهادی. نویسندگان یک رویکرد یادگیری سیستم کاملاً جدید را معرفی کردند که بر روی یک ساختار حفاظتی ساخته شده است که به طور مکانیکی به مسائل حفاظتی درحال‌توسعه مرتبط با دامنهٔ اینترنت اشیا می‌پردازد. برای کاهش آسیب‌پذیری‌های مختلف، این چارچوب از شبکه‌های تعریف‌شدهٔ نرم‌افزار و فعال‌کننده‌های مجازی‌سازی عملکرد شبکه (NFV) استفاده می‌کند. این ساختار از یک IDS برای تشخیص ناهنجاری اطلاعات حسگر تشکیل شده است و دقت ۹۸٪ را به دست آورده است.

کازونگو و همکاران [۱۸]یک سیستم تشخیص نفوذ ارائه کردند که از جنگل تصادفی برای تنظیم پارامترهای الگوریتم ژنتیک برای انتخاب تابع استفاده می‌کند. نتایج نشان داده که برای فرآیند دسته‌بندی دودویی، دقت ۹۹،۹۷٪ به‌دست‌آمده است. سو و همکاران [۱۹]از یک شبکهٔ عصبی مصنوعی (ANN) استفاده کردند تا آسیب‌پذیری‌های DDoS را پیدا کنند. برای مقابله با مشکل عدم توازن در داده‌ها، رویکرد روش نمونه‌برداری حداقلی (SMOTE) مورد استفاده قرار گرفته است. قبل از انتقال رکوردها به شبکهٔ عصبی مصنوعی، نویسندگان از نرمال‌سازی ویژگی‌ها در رکوردها استفاده کردند که بر روی مجموعه‌داده Bot-IoT، دقت حدود ۱۰۰٪ به‌دست‌آمده است.

باتوجه‌به اینکه روش‌های پیشین، یا در دقت نهایی طبقه‌بندی ضعف دارند و یا اینکه به دلیل استفاده از الگوریتم‌های شبکهٔ عصبی و یادگیری عمیق، همگرایی پایینی در مدت‌زمان کم دارند زمان اجرای بالایی دارند و در دسته‌بندی چند کلاسه چندان موفقیت‌آمیز عمل نمی‌کنند، نیاز به روشی ترکیبی که بتواند این چالش‌ها را رفع کند ضروری به نظر می‌رسد. از طرفی، در روش پیشنهادی در این مقاله از برخی از کارهای صورت‌گرفته در مقالات ذکر شده مانند نرمال‌سازی ویژگی‌ها، حذف ویژگی‌های زاید و داده‌های پرت، دسته‌بندی داده‌ها و همچنین طبقه‌بندی آنها به دو دستهٔ کلی عادی و ناهنجار استفاده شده است.

۳. روش تحقیق

در این مقاله، اثربخشی روش‌های فراابتکاری رایج در شناسایی نفوذهای بالقوه بررسی شده است. با تجزیه و تحلیل گرادیان تقویتی (Gradient Boosting (GB [۲۰]، تعیین مناسب‌ترین پاسخ‌ها برای تشخیص ناهنجاری محقق می‌شود. تجزیه و تحلیل، بر اساس تکنیک انتخاب ویژگی مبتنی بر الگوریتم فراکتشافی به نام الگوریتم سنجاقک [۲۱] تعبیه شده است، که برای افزایش کارآیی مدل استفاده می‌شود. مجموعه‌داده مورد استفاده برای این تحقیق، مجموعه‌داده NSL-KDD و UNSW-NB15 است. عملکرد مدل پیشنهادی با استفاده از معیارهایی مانند دقت (Accuracy)، صحت (Precision)، یادآوری (Recall)، امتیاز F1 (F1-score) و ضریب همبستگی ماتیو (MCC) برای ارزیابی جامع اندازه‌گیری می‌شود.

۳-۱. مجموعه داده‌ها

در این مقاله، از دو مجموعه‌داده شناخته‌شده شامل مجموعه‌داده NSL-KDD و UNSW-NB15 برای رسیدگی به چالش‌ها استفاده می‌کنیم. مجموعه‌داده NSL-KDD [۲۲]دارای ۱۴۸۵۱۷ ورودی است که هر کدام شامل ۱۴ ویژگی هستند. این مجموعه‌داده شامل ۳۹ نوع مختلف حمله است. این حملات را می‌توان به دو صورت دسته‌بندی کرد: دودویی (حمله یا عادی) و چند کلاسه (Normal، DoS، Probe، Privilege و Access).

علی‌رغم محدودیت‌های موجود در مجموعه‌داده‌های IDS مبتنی بر شبکه و نقص در نسخهٔ فعلی مجموعه‌داده KDD، محققان بر این باورند که همچنان می‌توانند به‌عنوان یک معیار ارزشمند برای ارزیابی روش‌های تشخیص نفوذ عمل کنند [۲۲]. مجموعه‌داده‌های آموزشی و آزمایشی NSL-KDD مقدار قابل‌توجهی از داده‌ها را ارائه می‌کنند که امکان تجزیه و تحلیل جامع را در کل مجموعه‌داده فراهم می‌کنند. این امر مقایسه و سازگاری نتایج حاصل از پروژه‌های تحقیقاتی مختلف را تسهیل می‌کند.

UNSW-NB15 [۲۳]مجموعه‌داده نفوذ با ۴۳ ویژگی است که شامل ۹ حمله DoS، کرم‌ها، Backdoors و Fuzzers است. مجموعه‌داده شامل ۱۷۵۳۴۱ رکورد آموزشی و ۸۲۳۳۲ رکورد آزمایشی از انواع حمله و داده‌های عادی است.

۳-۲. پیش‌زمینه

در این بخش، الگوریتم‌های مورد استفاده در روش پیشنهادی توضیح داده شده است.

۳-۲-۱. الگوریتم سنجاقک

موقعیت حرکت سنجاقک‌ها شامل ۵ رفتار است که عملیات آن‌ها را شرح داده می‌شود:

الف- جدایی: می‌توانیم جدایی سنجاقک‌های مجاور را به‌صورت رابطه ۱ محاسبه کنیم:

$$S_i = -\sum_{j=1}^{N} X - X_j \quad (۱)$$

که در آن X مکان هر سنجاقک را نشان می‌دهد، j همسایگان را نشان می‌دهد و N تعداد افراد همسایه است.

ب- تراز: از رابطه ۲ برای تعیین تراز سنجاقک‌ها استفاده می‌شود:

$$A_i = \frac{\sum_{j=1}^{N} V_j}{N} \quad (۲)$$

جایی که V_j نرخ همسایگی j-امین فرد را توصیف می‌کند.

ج- انسجام: انسجام از رابط ۳ به‌دست می‌آید:

$$C_i = \frac{\sum_{j=1}^{N} X_j}{N} - X \quad (۳)$$

جایی که X مکان فرد ساکن را نشان می‌دهد، j فرد همسایه را نشان می‌دهد و N طیف گسترده‌ای از افراد همسایه است.

د- جاذبه: با استفاده از رابطه ۴، عملیات جذب را برای تأمین وعده‌های غذایی محاسبه می‌کنیم:

$$F_i = X^{+} - X \quad (۴)$$

جایی که X محل قرارگیری فرد ساکن و X^{+} محل قرارگیری منبع وعده‌های غذایی را توصیف می‌کند.

ه- حواس‌پرتی: می‌توانیم میزان حواس‌پرتی از دشمن را با رابطه ۵ تعیین کنیم:

$$R_i = X^{-} - X \quad (۵)$$

جایی که X مکان فرد ساکن و X^{-} نشان‌دهنده دشمن غذایی است.

برای جایگزینی قراردادن سنجاقک‌های مصنوعی در ناحیه جستجو و انعکاس حرکت آنها، بردارهایی در نظر گرفته می‌شود: بردار گام ΔX و بردار تابع X.

بنابراین، بردار گام، مسیر حرکت سنجاقک‌ها را ارائه می‌دهد. می-توانیم این بردار را از طریق رابطه ۶ ترسیم کنیم:

$$\Delta X_t + 1 = (sS_i + aA_i + cC_i + fF_i + rR_i) + w\Delta X_t \quad (۶)$$

که در آن s، a، c، و f وزن‌هایی برای فازهای جداسازی، هم‌ترازی، انسجام، جاذبه و حواس‌پرتی هستند. همچنین، w وزن اینرسی را فراهم می‌کند و t شمارنده جدید است.

$$X_t + 1 = X_t + \Delta X_t + 1 \quad (۷)$$

محدوده‌های اکتشاف و بهره‌برداری ممکن است از طریق ابزار تبدیل مقادیر آن وزن‌ها اجرا شود. بردار نقش بدون زحمت با کمک بردار گام تعیین می‌شود، همان‌طور که در معادله ۷ نشان داده شده است:

در واقع، وزن توافقی و وزن تراز بیش از حد به مرحله اکتشاف اختصاص داده می‌شود. به طور مشابه، وزن هم‌ترازی کم و وزن بیش از حد به مرحله بهره‌برداری اختصاص داده می‌شود. بردار تابع، مکان سنجاقک را نشان می‌دهد. بااین‌حال، اگر هیچ پاسخی در اطراف وجود نداشته باشد سنجاقک‌ها باید در یک فضای جستجوی تصادفی پرواز کنند و عملکرد آنها باید با استفاده از معادله ۸ به‌روز شود:

$$X_t + 1 = X_t + X_t . levy(d) \quad (۸)$$

مکانیزم پرواز اجباری برای سنجاقک‌های مصنوعی استفاده می‌شود تا وقتی همسایه‌ای نداشته باشند، در منطقه جستجو حرکت کنند. این یک‌قدم زدن تصادفی است که برای ایجاد یک نقش تصادفی برای سنجاقک‌هایی که هیچ همسایه‌ای ندارند استفاده می‌شود.

۳-۲-۲. گرادیان تقویتی

گرادیان تقویتی، شامل ایجاد یک سری مدل است که با همکار می‌کنند تا خطاهای مدل قبلی را بهبود ببخشند. این یک فرآیند تکراری است که در آن هر مدل برای داده‌ها تنظیم می‌شود و وزن نقاط داده بر اساس دقت پیش‌بینی آنها تنظیم می‌شود. اصل اساسی گرادیان تقویتی، استفاده از گرادیان تابع ضرر برای هدایت به‌روزرسانی‌های مدل است. این رویکرد برای انواع مختلفی از وظایف یادگیری تحت نظارت، از جمله طبقه‌بندی، پیش‌بینی و ترتیب اثربخش است [۲۰].

یکی از راه‌هایی که می‌توان گرادیان تقویتی را به‌صورت ریاضی به تصویر کشید نشان‌دادن ماهیت فرمولی آن است که در ادامه مشاهده می‌کنید:

الف- پیش‌بینی $F_0(x) = \text{argmin}\gamma \sum i = 1^n L(y_i, \gamma)$ را مقداردهی اولیه کنید که در آن $L(y_i, \gamma)$ ویژگی از دست‌دادن است که خطای پیش‌بینی γ را به‌عنوان نمونه i اندازه‌گیری می‌کند.

ب- در هر دور r = 1,... ,R

- گرادیان ویژگی از دست‌دادن را با پیش‌بینی پیشرفته محاسبه کنید، یا به عبارتی دیگر $T_{r,i} = -\frac{\partial L(y_i, F_{r-1}(x_i))}{\partial F_{r-1}(x_i)}$.

- به یادگیرندۀ حساس h_r در باقی‌مانده‌های $T_{(r,i)}$ استفاده از سوابق آموزشی X را آموزش دهید.

- طول گام باکیفیت بالا y_t را محاسبه کنید به‌گونه‌ای که تابع تلفات را به حداقل برساند، یا به عبارتی دیگر $y_t = \text{argmin}\gamma \sum i = 1^n L(y_i, F_{r-1}(x_i) + \gamma h_r(x_i))$.

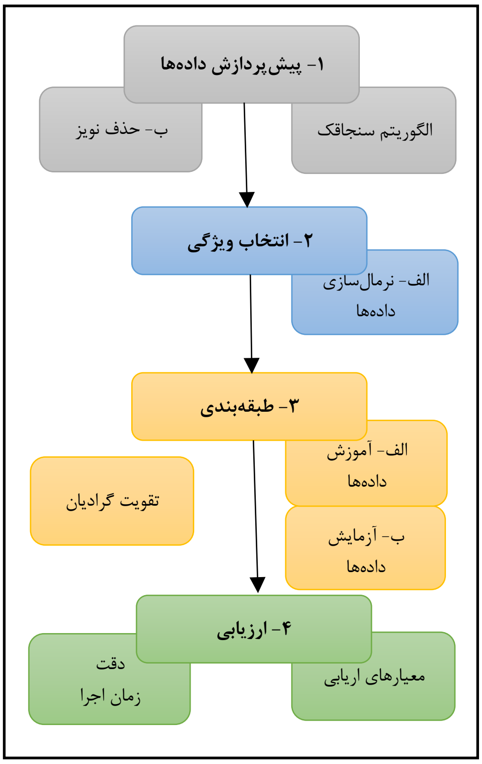
- پیش‌بینی $F_r(x) = F_{r-1}(x) + \gamma_r h_r(x)$ را به‌روزرسانی کنید.

ج- آخرین پیش‌بینی $F_r(x)$ مجموعه را محاسبه کنید.

این روش، یک تکنیک مؤثر و به طور گسترده پذیرفته شده است. در واقع، ابزاری همه‌کاره برای رسیدگی به مشکلات مختلف در حوزه یادگیری ماشین مانند رگرسیون و طبقه‌بندی است. گرادیان تقویتی، به دلیل توانایی مدیریت روابط پیچیده و غیرخطی بین متغیرها با پیش‌بینی‌های مبتنی بر خطا با عملکرد پایین‌تر متمایز است.

۳-۳. الگوریتم پیشنهادی

رویکرد پیشنهادی نشان‌داده‌شده در شکل (۱) شامل یک فرآیند ساختاریافته است. در ابتدا، هدف مرحله پیش‌پردازش داده‌ها رسیدگی به مقادیر از دست رفته و به حداقل رساندن تداخل نویز است. در مرحله بعد، الگوریتم سنجاقک برای انتخاب و استخراج ویژگی برای افزایش کیفیت داده‌ها و شناسایی مؤثرترین ویژگی‌ها برای آزمایش استفاده می‌شود. مراحل بعدی شامل مدل‌سازی و ارزیابی، با اجرای الگوریتم گرادیان تقویتی است. عملکرد این مدل‌ها بر اساس دقت، صحت، یادآوری، امتیاز F1 و MCC ارزیابی می‌شود.



شکل ۱. راهکار پیشنهادی برای تشخیص نفوذ

آماده‌سازی داده‌ها: آماده‌سازی داده‌ها به‌عنوان مرحله‌ای حیاتی در پالایش و ساختاردهی اطلاعات برای اطمینان از مناسب‌بودن آن برای توسعه و آموزش مدل‌ها عمل می‌کند. پس از به‌دست‌آوردن مجموعه‌داده، یک متغیر جدید برای طبقه‌بندی دسته ورودی‌های استاندارد به‌عنوان صفر و تمام دسته‌های دیگر به‌عنوان ۱، باهدف ایجاد یک نمایش دودویی ساده ایجاد می‌شود که بین حملات و دسترسی‌های معمولی تمایز قائل شود. به‌منظور استانداردسازی داده‌ها، تابع نرمال‌ساز ارائه شده در رابطه ۹ را پیاده‌سازی شده است.

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (۹)$$

انتخاب ویژگی: به دنبال مراحل پیش‌پردازش داده‌ها، تعداد ویژگی‌های مجموعه‌داده به ترتیب در NSL-KDD و UNSW-NB15 به ۱۲ و ۷ می‌رسد. انتخاب ویژگی، ما را قادر می‌سازد تا مرتبط‌ترین زیرمجموعه ویژگی را برای ساخت مدل شناسایی کنیم. در این مرحله، الگوریتم سنجاقک را برای استخراج ویژگی‌های مؤثرتر ارزیابی می‌کنیم.

تجزیه‌وتحلیل روابط داده‌ها یک عامل اساسی در حوزه آمار است که معمولاً از ۱- تا ۱ متغیر است. این ضریب فاصله بین متغیرها را با در نظر گرفتن نسبت کل انحرافات آنها اندازه‌گیری می‌کند که به طور مؤثر نقاط پرت را بین مجموعه داده‌های مختلف ثبت می‌کند. در مجموعه‌داده NSL-KDD، بر روی ۸ ویژگی با یک همبستگی قوی برای طبقه‌بندی دودویی تمرکز کردیم، در حالی‌که برای چند کلاسه، بر روی ۵ ویژگی تمرکز کردیم. همچنین، در UNSW-NB15، این اعداد به ترتیب ۷ و ۶ هستند. هنگامی‌که استراتژی پیشنهادی به روش انتخاب ویژگی می‌رسد، الگوریتم سنجاقک به‌عنوان بهینه‌ترین تکنیک برجسته می‌شود. این رویکرد ارتباط ویژگی‌ها را با متغیر هدف در نظر می‌گیرد تا مهم‌ترین ویژگی‌ها را تعیین کند.

ازآنجایی‌که روش بهینه‌سازی سنجاقک بر پایهٔ حرکات ایستا و پویای سنجاقک‌ها کار می‌کند، آن دسته از سنجاقک‌هایی که وزن و اثر بیشتری در شکل‌گیری همگرایی دارند نسبت به سایرین مشخص است و ازاین‌جهت، امکان تشخیص مواردی که اهمیت کمتری در همگرایی دارند امکان‌پذیر است. تعیین ویژگی‌های غیرمفید در مجموعه‌داده نیز با درنظرگرفتن همگرایی سنجاقک‌ها حاصل می‌شود. در واقع ازیک‌طرف، امکان رسیدن به همگرایی سریع‌تر و از طرفی دیگر، امکان انتخاب بهترین دسته‌ها با اجرای عملیات اصلی الگوریتم سنجاقک، این روش را نسبت به سایر روش‌ها برای انتخاب ویژگی مناسب‌تر می‌سازد.

طبقه‌بندی: در این مرحله از گرادیان تقویتی برای شناسایی کلاس داده‌ها استفاده می‌شود که یکی از روش‌های یادگیری ماشین است. تکنیک‌های یادگیری ماشین با استفاده از رفتار جمعی در داخل یک گروه تحریک می‌شوند. رفتارهای یادگیری ماشین دارای برخی ویژگی‌ها هستند: خود تصدی‌گری و کارفرمایی؛ بنابراین به روابط بین افراد و رفتارهای مشارکتی و محیط برنده آنها توجه می‌کنند.

ارزیابی عملکرد مدل: پس از اجرای مدل، یک ارزیابی جامع برای ارزیابی اثربخشی آن در شناسایی ریسک‌ها و بررسی رویکردهای جایگزین انجام می‌شود. معیارهای عملکردی مختلفی مانند دقت، صحت، یادآوری، امتیاز F1 و ضریب همبستگی ماتیو مورد استفاده قرار گرفت. این معیارها بر اساس معیارهایی مانند مثبت واقعی (TP)، منفی واقعی (TN)، مثبت کاذب (FP) و منفی کاذب (FN) محاسبه شدند. نماد TP نشان‌دهنده نمونه‌های مثبت شناسایی شده درست، FP نشان‌دهنده نمونه‌های منفی با برچسب اشتباه به‌عنوان مثبت، TN نشان‌دهنده نمونه‌های منفی بادقت شناسایی شده است و FN نشان‌دهنده نمونه‌های مثبت به‌اشتباه به‌عنوان منفی طبقه‌بندی شده است.

۴. نتایج و بحث

در این مقاله، تمامی شبیه‌سازی‌ها با استفاده از برنامه‌نویسی پایتون انجام شده است. تمامی وظایف عملیاتی این کار در محیط سخت‌افزاری زیر اجرا شده است: CPU Intel core i7-3400U (3.70 گیگاهرتز)، ۸ گیگابایت رم و سیستم‌عامل ویندوز ۱۰.

اثربخشی یک مدل در تشخیص نفوذ بستگی به مقادیر به‌دست‌آمده از معیارهای ارزیابی آن دارد (معرفی شده در بخش ۴-۱). برای مشاهده بهتر، نتایج کلی این مقاله در دو مرحله مجزا نمایش داده می‌شود. مرحله اول، عملکرد روش پیشنهادی برای طبقه‌بندی دودویی را نشان می‌دهد، در حالی‌که مرحله دوم عملکرد معیارهای مختلف برای طبقه‌بندی چند کلاسه را نشان می‌دهد. در هر دو مرحله برای نشان دادن تأثیر انتخاب بهترین ویژگی‌ها در بهبود نتایج، آزمایش‌ها یک بار با کل مجموعه‌داده و بار دیگر با بخشی از مجموعه‌داده که شامل ویژگی‌های انتخاب شده است انجام شده است.

۴-۱. معیارهای ارزیابی

دقت: یک معیار کلیدی که نشان‌دهنده نسبت پیش‌بینی‌های صحیح نسبت به همه مشاهدات است (رابطه ۱۰).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (۱۰)$$

دقت: نشان می‌دهد که روش چقدر صحیح است: از بین این موارد مثبت پیش‌بینی‌شده، چه تعداد مثبت واقعی هستند (رابطه ۱۱).

$$Precision = \frac{TP}{FP+TP} \quad (۱۱)$$

فراخوانی: این معیار بیان می‌کند که روش، چه تعداد مثبت را از طریق کل مثبت واقعی پیش‌بینی کرده است (رابطه ۱۲).

$$Recall = \frac{TP}{TP+FN} \quad (۱۲)$$

امتیاز F1: صحت و یادآوری باظرافت درون امتیاز F1 ترکیب شده‌اند (رابطه ۱۳). در سناریوهای داده‌های نامتعادل، برتری منفی‌های واقعی به طور گسترده دیدگاه ما را نسبت به عملکرد الگوریتم تغییر می‌دهند. اکنون مقدار زیادی از منفی‌های واقعی (TN) دیگر تأثیری بر امتیاز F1 نخواهند داشت.

$$F1 - Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (۱۳)$$

نرخ مثبت کاذب (FPR): این معیار نشان‌دهندهٔ نسبت بین نمونه‌های عادی که برچسب ناهنجار دارند بر تعداد کل نمونه‌های عادی است که از رابطهٔ ۱۴ به دست می‌آید:

$$FPR = \frac{FP}{FP+TN} \quad (۱۴)$$

پارامتر AUC بیانگر سطح زیر نمودار ROC است. هرچه سطح زیر نمودار ROC بیشتر باشد مقدار AUC نیز بالاتر است و نشان‌دهندهٔ قدرت بالای روش در تشخیص کلاس داده‌ها است.

۴-۲. دسته‌بندی دودویی (با/بدون انتخاب ویژگی)

کار آیی هر مدل تشخیص نفوذ را می‌توان بر اساس نتایج معیارهای ارزیابی مختلف آن بررسی کرد، به‌طوری‌که مقادیر بالاتر دقت، امتیاز F1، صحت و یادآوری عملکرد طبقه‌بندی بهتری را نشان می‌دهد. نتایج کلی طبقه‌بندی دودویی در مجموعه‌داده‌های UNSW-NB15 و NSL-KDD در جدول (۱) نشان داده شده است. برای تجزیه‌وتحلیل با رویکرد مقایسه‌ای، نتایج معیارهای به‌دست آمده با و بدون استفاده از الگوریتم سنجاقک بر روی مجموعه‌داده نشان داده شده است.

جدول ۱. نتایج دسته‌بندی دودویی

Criterion	NSL – KDD Dataset		UNSW – NB15 Dataset	
	Without feature selection	*With feature selection*	*Without feature selection*	*With feature selection*
Accuracy	99.994	99.874	99.992	99.985
Precision	99.983	98.773	99.746	99.894
Recall	99.971	98.473	99.377	98.841
F1-Score	99.974	98.553	99.622	99.027
FPR	0.016	0.010	0.016	0.007
AUC	99.628	99.884	99.702	99.945

جدول ۲. زمان مصرفی در دسته‌بندی دودویی

Criterion	NSL – KDD Dataset		UNSW – NB15 Dataset	
	Testing time (s)	*Training time (s)*	*Testing time (s)*	*Training time (s)*
Wihout feature selection	785	2184	632	1855
With feature selection	**467**	1638	**442**	1021

هدف اصلی استفاده از الگوریتم سنجاقک در مدل ارائه شده، کاهش پیچیدگی آن در سطح بالاتر با حفظ الگوهای داده اصلی است. ازآنجایی‌که مزیت روش سنجاقک یافتن متغیرهای نامرتبط

در داده‌ها است، نتیجه گرفته شده است که استفاده از آن در مجموعه‌داده مؤثرتر است. از جدول (۱) می‌توان دریافت که همه معیارهای ارزیابی برای مجموعه داده‌های UNSW-NB15 و NSL-KDD تا حدودی افزایش‌یافته است. همان‌طور که در جدول ۱ مشاهده می‌کنیم، دقت در مجموعه داده‌های UNSW-NB15 و NSL-KDD به ترتیب ۹۹/۹۹۲٪ و ۹۹/۹۹۴٪ است. علاوه بر کاهش پیچیدگی، به دلیل کاهش اطلاعات، بهبود عملکرد در هنگام استفاده از سنجاقک وجود دارد، از سویی دیگر، میزان زمان محاسبات نیز کاهش می‌یابد (جدول ۲ را ببینید).

۴-۳. دسته‌بندی چند کلاسه (با/ بدون انتخاب ویژگی)

خلاصه‌ای از عملکرد روش پیشنهادی برای مجموعه‌داده‌های UNSW-NB15 و NSL-KDD در جدول‌های (۳) و (۴) نشان داده شده است.

جدول (۳) تمام معیارهای ارزیابی را برای مجموعه داده‌های UNSW-NB15 و NSL-KDD نشان می‌دهد که با استفاده از الگوریتم سنجاقک افزایش یافته‌اند (۹۹/۹۸۷٪ برای UNSW-NB15 و ۹۹/۹۸۱٪ برای NSL-KDD). علاوه بر کاهش پیچیدگی، به دلیل کاهش اطلاعات، افزایش عملکرد در هنگام استفاده از سنجاقک وجود دارد، از سوی دیگر، میزان زمان محاسبات نیز کاهش می‌یابد (جدول ۴).

جدول ۳. نتایج دسته‌بندی چند کلاسه

Criterion	NSL – KDD Dataset		UNSW – NB15 Dataset	
	Without feature selection	With feature selection	Without feature selection	With feature selection
Accuracy	**99.987**	96.548	**99.981**	97.637
Precision	98.520	97.551	97.542	96.367
Recall	97.682	95.733	98.037	96.772
F1-Score	97.551	96.253	97.403	96.738
FPR	0.019	0.014	0.018	0.011
AUC	99.015	99.302	99.547	99.930

جدول ۴. زمان مصرفی در دسته‌بندی چند کلاسه

Criterion	NSL – KDD Dataset		UNSW – NB15 Dataset	
	Testing time (s)	Training time (s)	Testing time (s)	Training time (s)
Wihout feature selection	914	2377	728	2421
With feature selection	**537**	2028	**486**	1467

۴-۴. مقایسه

در طول مقایسه (جدول ۵)، استراتژی‌های طبقه‌بندی مختلفی را اجرا کردیم. در ابتدا، یک مجموعه دستورالعمل دو حالته برای یک هدف دودویی (خوش‌خیم یا ناهنجاری) با و بدون گنجاندن انتخاب ویژگی به‌کار گرفته شد. داده‌های جدول‌های (۱) و (۳) به ترتیب کار آیی کلی الگوریتم‌های گرادیان تقویتی و سنجاقک را برای مدل دودویی و چند کلاسه نشان می‌دهد که مشاهده شد که روش پیشنهادی بادقت ۹۹/۹۸۵٪ بدون انتخاب ویژگی برتر بود، در حالی‌که الگوریتم سنجاقک بالاترین میزان دقت را در ۹۹/۹۹۴٪ داشت.

در سیستم طبقه‌بندی چند کلاسه، نتیجه آزمایش به نرخ دقت ۹۷/۹۳۷٪ بدون انتخاب ویژگی (جدول ۳) و دقت ۹۹/۹۸۷٪ با اعمال انتخاب ویژگی دست پیدا کرد؛ بنابراین، رویکرد پیشنهادی با استفاده از الگوریتم سنجاقک به حداکثر دقت ۹۹/۹۸۴٪ و زمان اجرای ۴۴۲ ثانیه دست پیدا کرده است. در روش دیگر، درخت تصمیم با زمان اجرای ۴۸۳ ثانیه و دقت ۹۸/۶۵٪ دست‌یافته است؛ بنابراین، راه‌حل پیشنهادی ما از این لحاظ بهتر از سایر رویکردها است.

همان‌طور که در جدول (۵) مشاهده می‌شود راه‌حل پیشنهادی در بهترین حالت به‌دقت ۹۹/۹۹۴٪ در زمان ۴۴۲ ثانیه دست می‌یابد، درحالی‌که بهترین نتیجه برای سایر روش‌ها ۹۹/۹۵۷٪ و مدت‌زمان ۴۸۶ ثانیه است.

جدول ۵. مقایسه روش پیشنهادی با برخی روش‌های دیگر

Method	Accuracy	Time (s)
CNN + LSTM [12]	99.957	486
NFV + SVM [17]	99.943	491
SMOTE [19]	99.938	512
RF + GA [18]	99.879	526
NSGA [12]	99.844	533
OCSVM Gaussian [14]	99.437	541
Proposed approach	99.994	442

۵. نتیجه‌گیری

در این مقاله، مطالعه‌ای بر روی تکنیک‌های یادگیری ماشین برای تشخیص نفوذ شبکه‌های اینترنتی ارائه شد. برخی از مقالات تحقیقاتی برای مقایسه روش‌های مختلف یادگیری ماشین، مجموعه‌داده‌ها، رویکردهای مهندسی ویژگی و معیارهای عملکرد مورد بررسی قرار گرفتند. علاوه بر این، آزمایشی برای ارزیابی اثربخشی یک الگوریتم فراابتکاری با تکنیک انتخاب ویژگی مانند الگوریتم سنجاقک برای افزایش عملکرد مدل انجام شد. هدف اولیهٔ این آزمایش تشخیص حملات بدخیم و خوش‌خیم (طبقه‌بندی دودویی) بود.

۵. مرجع‌ها

[1] Zhang, C.; Jia, D.; Wang, L.; Wang, W.; Liu, F.; Yang, A. "Comparative Research on Network Intrusion Detection Methods Based on Machine Learning"; Comput. Secur. 2022, 121, 102861. doi: 10.1016/j.cose.2022.102861.

[2] Dina, A. S.; Manivannan, D. "Intrusion Detection Based on Machine Learning Techniques in Computer Networks"; Internet Things 2021, 16, 100462. doi: 10.1016/ j.iot.2021.100462.

[3] Lansky, J.; Ali, S.; Mohammadi, M.; Majeed, M. K.; Karim, S. H. T.; Rashidi, S. "Deep Learning-Based Intrusion Detection Systems: A Systematic Review", IEEE. Access. 2021, 9, 101574-101599. doi: 10.1109/ACCESS.2021. 3097247.

[4] Ahmad, Z.; Shahid Khan, A.; Wai Shiang, C.; Abdullah, J.; Ahmad, F. "Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches"; T. Emerg. Telecommun. T. 2021, 32, e4150. doi: 10.1002/ett.4150.

[5] Chou, D.; Jiang, M. "A Survey on Data-Driven Network Intrusion Detection"; Acm. Comput. Surv. 2021, 54, 1-36. doi: 10.1145/3472753.

[6] Heidari, A.; Jamali, M.A. "Internet of Things Intrusion Detection Systems: a comprehensive review and future directions"; CLUSTER. COMPUT. 2023, 26, 3753-3780. doi: 10.1007/s10586-022-03776-z.

[7] Huong, P.; Hung, D.V. "Intrusion Detection in IoT Systems Based on Deep Learning Using Convolutional Neural Network"; 6th NAFOSTED Conference on Information and Computer Science (NICS), 2019, 448-453. doi: 10.1109/NICS48868.2019.9023871.

[8] Baich, M.; Hamim, T.; Sael, N.; Chemlal, Y. "Machine Learning For IoT Based Networks Intrusion Detection: A Comparative Study"; PROCEDIA. COMPUT. SCI. 2022, 215, 742-751. doi: 10.1016/j.procs.2022.12.076.

[9] Gurung, S.; Ghose, M. K.; Subedi, A. "Deep Learning Approach on Network Intrusion Detection System Using NSL-KDD Dataset"; Int. J. Comput. Net. Inf. Sec. 2019, 11, 8-14. doi: 10.1109/TETCI.2017.2772792.

[10] Tsogbaatar, E.; Bhuyan, M. H.; Taenaka, Y.; Fall, D.; Gonchigsumlaa, K.; Elmroth, E.; Kadobayashi, Y. "DeL-IoT: A Deep Ensemble Learning Approach to Uncover Anomalies in IoT"; Internet Things 2021, 14, 100391. doi: 10.1016/j.iot.2021.100391.

[11] Ullah, I.; Mahmoud, Q. H. "Design and Development of a Deep Learning-Based Model for Anomaly Detection in IoT Networks"; IEEE. Access, 2021, 9, 103906-103926. doi: 10.1109/ACCESS.2021.3094024.

[12] Roopak, M.; Tian, G.Y.; Chambers, J. "An Intrusion Detection System Against DDoS Attacks in IoT Networks"; 10th Annual Computing and Communication Workshop and Conference, 2020, 3, 0562-0567. doi: 10.1109/CCWC47524.2020.9031206.

[13] Saheed, Y. K.; Abiodun, A. I.; Misra, S.; Holone, M. K.; Colomo-Palacios, R. "A Machine Learning-Based Intrusion Detection for Detecting Internet of Things Network Attacks"; Alex. Eng. J. 2020, 61, 9395-9409. doi: 10.1016/j.aej.2022.02.063.

[14] Bhatt, P.; Morais, A. "HADS: Hybrid Anomaly Detection System for IoT Environments"; International Conference on Internet of Things, Embedded Systems and Communications, 2018, 3, 191-196. doi: 10.1109/IINTEC.2018.8695303.

[15] Shafiq, M.; Tian, Z.; Bashir, A. K.; Du, X.; Guizani, M. "CorrAUC: A Malicious BoT-IoT Traffic Detection Method in IoT Network Using Machine Learning Techniques"; IEEE Internet Things J. 2020, 8, 3242-3254. doi: 10.1109/JIOT.2020.3002255.

[16] Sarwar, A.; Hasan, S.; Khan, W. U.; Ahmed, S.; Marwat, S.N.K. "Design of an Advance Intrusion Detection System For IoT Networks"; 2nd International Conference on Artificial Intelligence, 2022, 3, 46-51. doi: 10.1109/ICAI55435.2022.9773747.

[17] Bagaa, M.; Taleb, T.; Bernabe, J. B.; Skarmeta, A. "A Machine Learning Security Framework For IoT Systems"; IEEE. Access. 2020, 8, 114066-114077. doi: 10.1109/ACCESS.2020.2996214.

[18] Kasongo, S. M. "An Advanced Intrusion Detection System For IIoT Based on GA and Tree-Based Algorithms"; IEEE. Access. 2021, 9, 113199-113212. doi: 10.1109/ACCESS.2021.3104113.

[19] Soe, Y. N.; Santosa, P. I.; Hartanto, R. "DDoS Attack Detection Based on Simple ANN with SMOTE for IoT Environment"; Fourth International Conference on Informatics and Computing 2019, 1-5. doi: 10.1109/ICIC47613.2019.8985853.

[20] Raza, A.; Rustam, F.; Siddiqui, H. U. R.; Diez, I. D. l. T.; Garcia-Zapirain, B.; Lee, E.; Ashraf, I. "Predicting Genetic Disorder and Types of Disorder Using Chain Classifier Approach"; Genes. 2023, 14, 71. https://www.mdpi.com/2073-4425/14/1/71.

[21] Meraihi, Y.; Ramdane-Cherif, A.; Acheli, D.; Mahseur, M. "Dragonfly Algorithm: A Comprehensive Review and Applications"; Neural Comput. Appl. 2020, 32, 16625-16646. doi: 10.1007/s00521-020-04866-y.

[22] Meena, G.; Choudhary, R. R. "A Review Paper On IDS Classification Using KDD99 and NSL-KDD Dataset in WEKA"; International Conference on Computer, Communications and Electronics, 2017, 553-558. doi: 10.1109/COMPTELIX.2017.8004032.

[23] Meftah, S.; Rachidi, T.; Assem, N. "Network Based Intrusion Detection Using the UNSW-NB15 Dataset"; Int. J. Comput. Digital. Sys. 2019, 478-487. doi: 10.1109/ICAEEE54957. 2022.9836404.