Scientific Journal of Electronical & Cyber Defence
Vol. 10, No. 4, Winter 2023, Serial No. 40

An Optimized Compound Deep Neural Network Integrating With Feature Selection for Intrusion Detection System in Cyber Attacks

J. Mazloum*, H. Bigdeli

* Associate Professor, Faculty of Electrical Engineering, Shahid Sattari University of Aeronautical Sciences and Technology, Tehran, Iran

(Received: 04/01/2022, Accepted: 23/02/2022)

ABSTRACT

Along with the rapid advancement of machine learning (ML) and deep learning (DL) methods in the data world, one of their significant applications in cyber attacks is dedicated to intrusion detection systems (IDSs) using state-of-the-art classification algorithms. Due to the importance of protecting secured and vital information relating to organizations or individuals against intruders' attacks, IDS has been the subject of numerous research to enhance accuracy and reliability. As a consequence, this paper presents a hybrid model integrating feature selection, classification, and hyper-parameters optimization. First, the primary candidate features are selected simultaneously and separately by modified mutual information (MMI), genetic algorithm (GA), and Anova F-value approaches, followed by voting to extract all common features as input variables. Subsequently, a compound CNN and LSTM classifier (CNN-LSTM) is employed, where its hyper-parameters will be determined through a random switch grey wolf-whale optimization algorithm (RS-GWO-WOA). In order to analyze the suggested scheme, a comparison with other strategies in terms of accuracy, precision, recall, F1 score, and time on the NSL-KDD dataset has been accomplished, confirming the superiority of the developed approach.

Keywords: Intrusion Detection System, Feature Selection, Hyper-parameter Optimization, Mutual Information, Genetic Algorithm, Anova F-value, Grey Wolf Optimization Algorithm, Whale Optimization Algorithm.

* Corresponding Author Email: Jalil.mazloum@ssau.ac.ir

نشریه علمی "پدافند الکترونیکی و سایبری"
سال دهم، شماره ۴، زمستان ۱۴۰۱، ص ۵۱- ۴۱

علمی – پژوهشی

شبکه عصبی عمیق ترکیبی بهینه ادغام‌شده با انتخاب ویژگی برای سامانه تشخیص نفوذ در حملات سایبری

جلیل مظلوم[1*]، حمید بیگدلی[2]

۱- دانشیار، دانشکده مهندسی برق، دانشگاه علوم و فنون هوایی شهید ستاری، ۲- استادیار، دانشگاه فرماندهی و ستاد آجا، تهران، ایران

(دریافت: ۱۴۰۰/۱۰/۱۴، پذیرش: ۱۴۰۰/۱۲/۰۴)

چکیده

همراه با پیشرفت سریع روش‌های یادگیری ماشین (ML) و یادگیری عمیق (DL) در دنیای داده، یکی از کاربردهای مهم آن‌ها در حملات سایبری، به سیستم‌های تشخیص نفوذ (IDS) با استفاده از الگوریتم‌های طبقه‌بندی پیشرفته اختصاص می‌یابد. با توجه به اهمیت حفظ اطلاعات ایمن و حیاتی سازمان‌ها یا افراد در برابر حملات نفوذگران، IDS موضوع تحقیقات متعددی جهت افزایش دقت و قابلیت اطمینان بوده است. در نتیجه، این مقاله یک مدل ترکیبی را ارائه می‌کند که به ادغام انتخاب ویژگی، طبقه‌بندی و بهینه‌سازی هایپرپارامترها پرداخته است. ابتدا، ویژگی‌های کاندید اولیه به‌طور هم‌زمان و جداگانه توسط روش‌های اطلاعات متقابل اصلاح‌شده (MMI)، الگوریتم ژنتیک (GA) و آزمون F تحلیل واریانس انتخاب می‌شوند و پس‌ازآن، رأی‌گیری جهت استخراج همه ویژگی‌های مشترک به‌عنوان متغیرهای ورودی صورت می‌پذیرد. در ادامه، یک طبقه‌بند ترکیبی CNN و LSTM (CNN-LSTM) به کار گرفته می‌شود که هایپرپارامترهای آن توسط یک الگوریتم بهینه‌سازی گرگ خاکستری- نهنگ با جابجایی تصادفی (RS-GWO-WOA) تعیین خواهد شد. به‌منظور تجزیه‌وتحلیل طرح پیشنهادی، مقایسه‌ای با سایر روش‌ها از نظر صحت، دقت، یادآوری، امتیاز F1 و زمان در مجموعه‌داده NSL-KDD انجام شده است که برتری رویکرد توسعه‌یافته را تائید می‌نماید.

کلیدواژه‌ها: سیستم تشخیص نفوذ، انتخاب ویژگی، بهینه‌سازی هایپرپارامترها، اطلاعات متقابل، الگوریتم ژنتیک، آزمون F تحلیل واریانس، الگوریتم بهینه‌سازی گرگ خاکستری، الگوریتم بهینه‌سازی نهنگ

۱- مقدمه

نیاز به امنیت شبکه و محافظت در مقابل حملات سایبری با توجه به رشد گسترده اتصالات شبکه‌ها روزبه‌روز در حال افزایش است. در حوزه امنیت شبکه، سیستم تشخیص نفوذ[1] (IDS) به‌عنوان مدافعی جهت تشخیص و مسدودسازی دسترسی متجاوزان به شبکه رفتار می‌کند. زمانی که یک نفوذگر اقدام به حمله می‌کند، وظیفه اصلی IDS خودداری از آن است، به‌طوری‌که این واکنش باید پیش از آسیب یا دسترسی به هرگونه اطلاعات صورت پذیرد. جدای از این، هدف اصلی IDS، ایمن‌سازی دسترسی، یکپارچگی و محرمانگی سیستم شبکه است. در حال حاضر، بهره‌مندی از IDS هوشمند راه‌حلی مؤثر در راستای امنیت شبکه و محافظت در برابر مهاجمان خارجی به‌حساب می‌آید؛ بنابراین، افزایش کارایی سیستم‌های تشخیص نفوذ هوشمند به‌عنوان نقطه مرکزی امنیت شبکه موضوع اصلی بسیاری از پژوهش‌های اخیر تبدیل شده است.

بااین‌حال، مسائل مربوط به امنیت سیستم شبکه با مبحث داده‌های بزرگ سروکار دارد؛ به‌طوری‌که طراحی یک IDS دقیق و قابل‌اعتماد جهت شناسایی تهدیدات امنیتی به دلیل ظرفیت زیاد داده‌های شبکه که حاوی ویژگی‌های اضافی و نامربوط است، چالش‌برانگیز می‌گردد. این مسئله، نه‌تنها روند طبقه‌بندی را کاهش می‌دهد، بلکه مانع از تصمیم‌گیری دقیق طبقه‌بند می‌شود. در نتیجه، تکنیک‌های مختلفی برای شناسایی حملات IDS پیشنهاد شده است که روزبه‌روز درحال‌توسعه است. در این راستا، یکی از عوامل کلیدی در عملکرد IDS، انتخاب ویژگی‌های نماینده از مجموعه‌داده اصلی است که می‌تواند نقش مهمی در بهینه‌سازی عملکرد IDS داشته باشد. افزودن فاز انتخاب ویژگی در فرایند IDS موجب بهبود هزینه محاسباتی، افزایش دقت طبقه‌بندی و اجتناب از نیاز به آموزش مجدد مدل‌ها در صورت اضافه شدن ویژگی‌های جدید است [۱-۳].

از روش‌های یادگیری ماشین[2] (ML) کم‌عمق موجود جهت طبقه‌بندی این حملات می‌توان به تکنیک‌های ماشین بردار

* رایانامه نویسنده مسئول: Jalil.mazloum@ssau.ac.ir

[1] Intrusion Detection System

[2] Machine Learning

ناشر: دانشگاه جامع امام حسین (ع)

پشتیبان[1] (SVM)، درختان تصمیم[2] (DT)، جنگل تصادفی[3] (RF)، درختان مازاد[4] (ET)، تقویت گرادیان[5] (GB) و بیز ساده[6] (NB) اشاره نمود [۴-۶]. این راهکارها نرخ دقت بهبودیافته‌ای را در تشخیص حملات ارائه می‌کنند، اما نیاز به دانش متخصص حوزه نیاز دارند؛ دارای هزینه محاسباتی بالا بوده و همچنین مستعد خطا هستند. به‌طور مثال، شبکه بیز مجموعه‌داده با ویژگی‌های زیاد را بسیار کند طبقه‌بندی می‌کند؛ ماشین بردار پشتیبان فاقد انتخاب مستقیم تابع کرنل است، درحالی‌که فاز آموزش نیز بسیار کند است و به حافظه بیشتری نیاز دارد. درخت تصمیم درصورتی‌که درختان مجدداً هرس نشوند، درگیر بحث بیش برازش می‌شود و از طرف دیگر، نیاز به در نظر گرفتن نوع داده (عددی یا غیرعددی) قبل از ساختن درخت دارد. در جنگل تصادفی، محدودیت اصلی تعداد زیاد درخت‌ها است که می‌تواند روند تست را کند نماید [۷].

جهت رفع کاستی‌های مذکور، در سال‌های اخیر، یادگیری عمیق[7] (DL)، به‌عنوان زیرمجموعه‌ای از یادگیری ماشین با رویکردی بسیار پیچیده در زمینه IDS نیز متمرکز شده‌اند. یادگیری عمیق برتری خود را با قابلیت یادگیری لایه‌ای خود نشان داده است که می‌تواند بهتر از روش‌های یادگیری کم‌عمق عمل کند. این رویکردها تجزیه‌وتحلیل عمیق‌تری از شبکه به همراه تشخیص سریع‌تر حملات را ارائه می‌دهند. از روش‌های مرسوم آن می‌توان پرسپترون چندلایه[8] (MLP)، شبکه عصبی پیچشی[9] (CNN)، واحد بازگشتی دروازه‌دار[10] (GRU) و حافظه طولانی کوتاه - مدت[11] (LSTM) را نام برد [۸-۱۰].

همان‌طور که بیان گردید، شبکه‌های عصبی عمیق[12] (DNN) به‌عنوان یک روند یادگیری ماشین مفید تکامل‌یافته‌اند که با موفقیت در کاربردهای گوناگونی مورداستفاده قرار گرفته است. بااین‌وجود، عملکرد آن‌ها به‌شدت به مقادیر انتخابی هایپرپارامترها متکی است. ازاین‌رو، نگرانی اساسی هنگام استفاده از یکی از این مدل‌های DL، نحوه تنظیم صحیح هایپرپارامترهای آن است. اساساً، هایپرپارامترهای یک مدل DL مجموعه تنظیمات اساسی هستند که رفتار، معماری و عملکرد مدل را در سیستم اصلی کنترل می‌نمایند. مسئله این است که مقادیر این هایپرپارامترها از حوزه‌ای به حوزه دیگر متفاوت است و برای تنظیم آن‌ها درست قبل از شروع فرایند یادگیری، دانش قبلی موردنیاز است. برای غلبه بر این مشکل، چندین تکنیک خودکار جهت کاهش زمان یادگیری و افزایش قابلیت اطمینان در فضای جستجوی بزرگ و پیچیده همچون الگوریتم‌های فراابتکاری توسعه یافته‌اند [۱۱ و ۱۲].

روش پیشنهادی در این پژوهش بر مبنای محورهای ارائه‌شده در ادامه موردبحث قرار خواهند گرفت:

- روش انتخاب ویژگی ترکیبی جدید مبتنی بر رأی‌گیری اشتراکی، به صورتی که در ابتدا تمامی ویژگی‌های ورودی به‌صورت هم‌زمان وارد الگوریتم‌های انتخاب ویژگی اطلاعات متقابل اصلاح‌شده[13] (MMI)، الگوریتم ژنتیک[14] (GA) و آزمون F تحلیل واریانس[15] می‌شوند و سه زیرمجموعه از ویژگی‌ها را تولید می‌کنند که از طریق واحد اشتراک‌گیری، یک مجموعه نهایی بهینه به‌عنوان متغیرهای ورودی مدل تولید خواهد کرد. این روش انتخاب ویژگی، با گزینش حداقل تعداد مشخصه‌های سیستم با وابستگی زیاد، ابعاد ورودی مدل را کاهش داده و سرعت آموزش را افزایش می‌دهد، درحالی‌که ویژگی‌های اضافی و نامربوط نیز حذف شده‌اند.

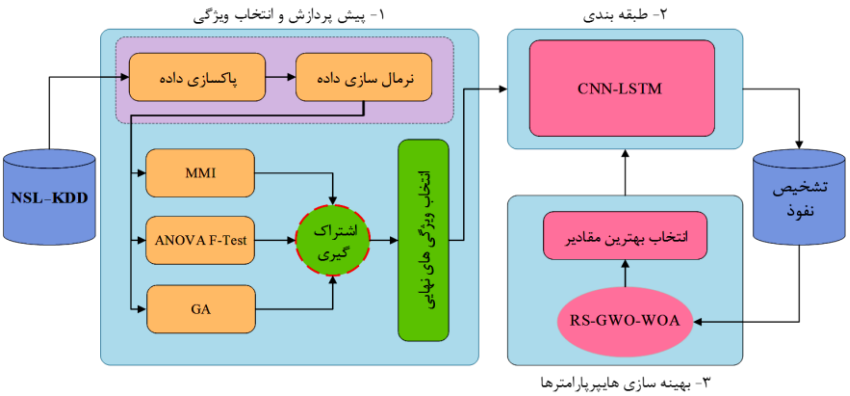
- ارائه یک مدل طبقه‌بند پیشرفته ادغامی از روش‌های طبقه‌بندی CNN و LSTM که CNN-LSTM نام دارد و تمامی مزایای این دو مدل DNN را در خود جای داده است، به‌طوری‌که قابلیت عملکرد IDS را به‌صورت مؤثری افزایش داده و دقت و قابلیت اطمینان آن را بهبود بخشیده است.

- به‌کارگیری یک الگوریتم تلفیقی بهینه‌سازی الهام گرفته از طبیعت RS-GWO-WOA، ترکیبی از الگوریتم‌های گرگ خاکستری[16] (GWO) و نهنگ[17] (WOA) مبتنی بر جابه‌جایی تصادفی است که جهت یافتن مناسب‌ترین هایپرپارامترهای مدل DNN پیشنهادی در طول فرایند طبقه‌بندی بکار گرفته می‌شود و منجر به نرخ همگرایی سریع و افزایش کارایی مدل طبقه‌بندی می‌گردد.

۱-۱- پیشینه پژوهش

با توجه به اهمیت موضوع IDS، تحقیقات گسترده‌ای در این حوزه در سال‌های اخیر صورت گرفته است که می‌توان آن‌ها را در ۳ فاز مختلف انتخاب ویژگی، طبقه‌بندی و بهینه‌سازی هایپرپارامترها بررسی نمود.

[1] Support Vector Machine
[2] Decision Trees
[3] Random Forest
[4] Extra Trees
[5] Gradient Boosting
[6] Naïve Bayes
[7] Deep Learning
[8] Multi-Layer Perceptron
[9] Convolutional Neural Network
[10] Gated Recurrent Unit
[11] Long Short-Term Memory
[12] Deep Neural Network
[13] Modified Mutual Information
[14] Genetic Algorithm
[15] Anova F-test
[16] Grey Wolf Optimizer
[17] Whale Optimizer



شکل (۱). ساختار کلی روش پیشنهادی

کاهش ابعاد ورودی سیستم تحت عنوان انتخاب ویژگی، روش‌های مختلفی را شامل می‌گردد که در سه دسته فیلتر، بسته‌بند و تعبیه‌شده تقسیم می‌شوند. یکی از پرکاربردترین روش‌های مبتنی بر فیلتر، روش اطلاعات متقابل[1] (MI) است. در [۱۳]، یک روش انتخاب ویژگی دوسطحی بر اساس حداقل افزونگی - حداکثر ارتباط[2] (mRMR) بر پایه MI و افزایش اطلاعات[3] (IG) پیشنهاد شده است. در این روش، ابتدا ویژگی‌های نامربوط و زائد برای کاهش ابعاد داده‌ها با استفاده از الگوریتم mRMR فیلتر شدند و ویژگی‌های دارای همبستگی بالا با حملات دارای نرخ تشخیص پایین بر اساس محاسبه بهره اطلاعات به دست آمد که در نهایت این ویژگی‌ها با هم ادغام شدند تا مجموعه ویژگی نهایی به دست آید. زیرمجموعه‌ای از ویژگی‌های مناسب به ترتیب بر اساس روش‌های MI و الگوریتم کرم شب‌تاب توسط [۱۴] انتخاب شده‌اند، سپس یک استراتژی انتخاب ویژگی با ترکیب دو زیرمجموعه ویژگی‌ها پیشنهاد گردیده است. بهره‌جویی از مشخصه سرعت همگرایی بالا در الگوریتم رقابت استعماری و قابلیت جستجوی قدرتمند GA می‌تواند در انتخاب ویژگی‌های بهینه در IDS مؤثر باشد که ترکیب آن‌ها سرعت همگرایی و دقت را بهبود بخشیده است [۱۵]. در [۱۶]، الگوریتم ژنتیک بهبودیافته جهت گزینش ورودی‌های کاندید مؤثر و مطلوب به کار گرفته شده است؛ همچنین، [۱۷] یک ساختار چهارلایه متشکل از تحلیل و طبقه‌بندی ترافیک شبکه، آشکارسازی نفوذ، تصمیم‌گیری و مدیریت ثبت وقایع و هشدارها ارائه کرده است که در بخش پیش‌پردازش داده، از الگوریتم GA جهت انتخاب مشخصه‌های مرتبط بهره می‌جوید. از طرف دیگر، [۱۸] روش‌های انتخاب ویژگی کای دو، ET و تحلیل واریانس را بر روی چهار طبقه‌بند RF، DT، KNN و XGBoost به‌منظور تشخیص زودهنگام حملات DDoS در دستگاه‌های IoT اعمال کرده است. بر اساس نتایج، روش ترکیبی از ANOVA و XGBoost عملکرد برتری را ارائه کرده‌اند. در فرایند انتخاب ویژگی [۱۹]، ابتدا از تکنیک همبستگی زوجی برای حذف بیشترین همبستگی‌ها استفاده شده است. در مرحله بعد، چهار روند انتخاب ویژگی منفرد مانند تحلیل واریانس، VT، کای دو و RFE به هدف شناسایی مجموعه ویژگی‌های خود به‌طور مستقل اجرا شدند. برای یافتن یک زیرمجموعه بهینه، مجموعه‌های مشخص‌شده از ویژگی‌ها با روش‌های جداگانه توسط تکنیک اشتراک‌گیری اقتباس‌شده از نظریه مجموعه‌ها ادغام شده‌اند.

به‌عنوان یک سیستم هوشمند، داده‌کاوی در طول فرایند IDS جهت کسب بهترین نتایج اعمال می‌گردد. طبقه‌بند DT در کنار الگوریتم‌های رقابت استعماری و GA برای انتخاب ویژگی در [۱۵]، تحت معیارهای نرخ تشخیص، نرخ هشدار غلط و افزایش سرعت همگرایی روی مجموعه‌داده KDD99 پیاده‌سازی شده است. مرجع [۲۰]، مسئله افزایش سرعت یادگیری و تست و همچنین دقت مدل RF را به کمک ترکیب آن با GA تقویت کرده است که روی KDD Cup 99 آنالیز شده است درحالی‌که روش بهره اطلاعات به کاهش ویژگی‌های ورودی پرداخته است. زمان ساخت مدل و دقت آن روی مجموعه‌داده NSL-KDD با طبقه‌بندهای SVM، CART و NB توسط [۲۱] مورد مقایسه قرار گرفت که پیش از آن در فاز پیش‌پردازش داده، الگوریتم DT هرس شده با C4.5 و گسسته‌سازی به اجرا درآمده است. در سال‌های اخیر، ترکیب روش‌های داده‌کاوی از قبیل ML و DL به‌منظور رفع محدودیت مدل‌ها، افزایش کارایی، جلوگیری از برازش و بهبود نرخ همگرایی بسیار موردتوجه قرار گرفته است. به‌عنوان نمونه، یک چارچوب یکپارچه IDS برای شبکه‌های برق مبتنی بر SCADA پیشنهاد شده است [۲۲]؛ این طرح ترکیبی RFE-XGBoost نامیده شد، به‌طوری‌که RFE ویژگی‌ها را به‌صورت بازگشتی بر اساس امتیازات WFI در طول فرایند آموزش انتخاب می‌کند، درحالی‌که روش مجموعه آرای اکثریت، برچسب خروجی را بر اساس مجموع نه طبقه‌بند ناهمگن پیش‌بینی می‌کند - سه گروه بسته‌بندی، یعنی RF، ET و DT، سه گروه تقویت‌کننده، یعنی XGB، GB و AdB-DT، همراه با ANN، NB و KNN. این روند منجر به یک راه‌حل دقیق در نتیجه ترکیبی از مفیدترین ویژگی‌ها و پیش‌بینی از طبقه‌بندهای

[1] Mutual Information
[2] Maximum Relevance-Minimum Redundancy
[3] Information Gain

ناهمگن چندگانه بر اساس دقت، صحت، یادآوری، امتیاز F1 می‌شود. بر اساس [۲۳]، یک IDS شامل سه مرحله اصلی، ۱) انتخاب ویژگی مبتنی بر mRMR، 2) تشخیص نفوذ موازی با استفاده از MLP و RBF و ۳) تجزیه‌وتحلیل نهایی ارائه شده است؛ برای مرحله تشخیص نفوذ موازی، چهار IDS مستقل اجرا شده‌اند که هرکدام مسئول تجزیه‌وتحلیل یک زیر کلاس از نفوذها هستند. در مرحله آخر، یک رمزگذار تقدم اتخاذ شده است که خروجی را بر اساس نتایج IDSهای موازی تولید می‌کند. در [۲۴]، یک ساختار جدید DL مبتنی بر رأی‌گیری، به نام VNN، معرفی شده است تا با در نظر گرفتن چندین مدل ایجادشده، توانایی جمع‌آوری بهترین مدل‌ها را به‌منظور ایجاد نتایج دقیق‌تر و قوی‌تر فراهم نماید. نتایج تجربی بر روی KDD CUP 99 و CTU-13، به‌عنوان دو مجموعه‌داده پرکاربردتر در حوزه شبکه‌های کامپیوتری، نشان داد که هشدارهای کاذب تا ۷۵ درصد در مقایسه با مدل‌های DL اصلی، از جمله DNN، LSTM CNN و GRU کاهش یافته است.

در واقع، پارامترهای تنظیم مدل‌های DL با توجه به کاربرد متفاوت است؛ بنابراین نحوه دستیابی یا استخراج پارامترهای دقیق از اهمیت بالایی برخوردار است و به‌شدت موردتوجه محققان قرار گرفته است. بهترین روش خودکار برای حل این موضوع، تبدیل مسئله استخراج پارامتر به یک مسئله بهینه‌سازی توسط الگوریتم‌های بهینه‌سازی همچون تکنیک‌های فراابتکاری است. این الگوریتم‌ها باید به‌طور مؤثر بین بهره‌وری محلی و اکتشاف جهانی تعادل برقرار نمایند تا موجب افزایش عملکرد گردند. گروهی از این الگوریتم‌های فراابتکاری پرکاربرد به‌عنوان روش‌های مبتنی بر طبیعت از قبیل الگوریتم‌های GWO و WOA به‌حساب می‌آیند. همان‌طور که در [۲۵] بیان شده است، مقداردهی اولیه و تنظیم بردار وزن ANN برای دستیابی به حداقل میانگین مربعات خطا توسط الگوریتم WOA اعمال گشته است، به‌طوری‌که WOA قادر است ANN را برای یافتن وزن‌های بهینه آموزش دهد. تنظیم پارامترهای یک مدل IDS نیمه نظارتی توسط الگوریتم GWO ابری (CGWO) در [۲۶] توضیح داده شده است که تعادل توانایی‌های اکتشاف و بهره‌برداری را به‌طور هم‌زمان در نظر می‌گیرد. با توجه به محدودیت‌های موجود در هر الگوریتم که مانع رسیدن به تعادل مناسب بین فاز بهره‌برداری و اکتشاف می‌گردد، ایده ترکیب روش‌های بهینه‌سازی جهت رسیدن به این تعادل ارائه شده است. هنگامی‌که انتخاب پارامتر غیرعلمی باشد، منجر به دقت طبقه‌بندی ضعیف می‌شود؛ برای حل این مشکل، [۲۷] یک الگوریتم GWO مبتنی بر الگوریتم بهینه‌سازی ازدحام ذرات (PSOGWO) با هدف بهبود IDS ارائه کرده است تا بتوان تابع کرنل و پارامترهای مدل طبقه‌بندی SVM را به بهترین نحو ممکن تخمین نمود.

۲- روش پیشنهادی

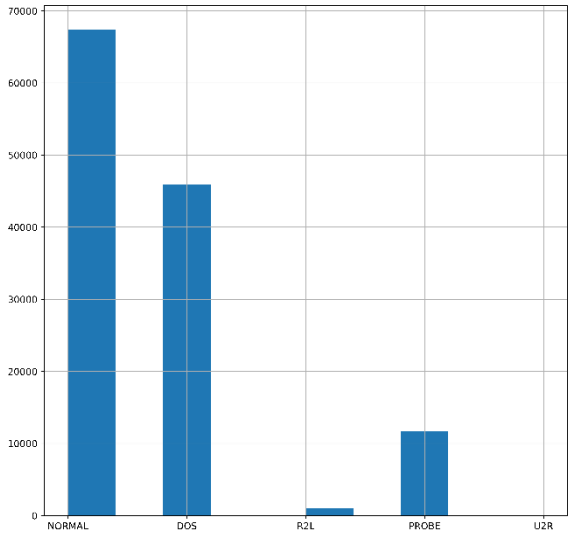
در این پژوهش، یک مدل جدید IDS مبتنی بر ترکیب معرفی شده است. همان‌طور که در شکل (۱) نشان داده شده است، این مطالعه دقت شناسایی نفوذ را از طریق افزودن ماژول انتخاب ویژگی بهبودیافته، به‌کارگیری ترکیب مدل‌های CNN و LSTM به‌عنوان طبقه‌بند و استفاده از الگوریتم فراابتکاری RS-GWO-WOA به‌منظور بهینه‌سازی هایپرپارامترهای مدل طبقه‌بندی تقویت نموده است. لازم به ذکر است که مجموعه‌داده اولیه پیش از ورود به سه گام مذکور، ابتدا وارد مرحله پیش‌پردازش می‌گردد که در آنجا مقادیر معیوب و ازدست‌رفته داده با مقدار متوسط ستون مربوطه به‌عنوان فاز پاک‌سازی داده جایگزین می‌شوند. در ادامه، نرمال‌سازی داده‌ها به کمک روش MinMaxScaler از کتابخانه Keras صورت می‌پذیرد تا داده‌ها را جهت ورود به گام مهندسی ویژگی آماده نمایند.

۲-۱- توصیف مجموعه داده

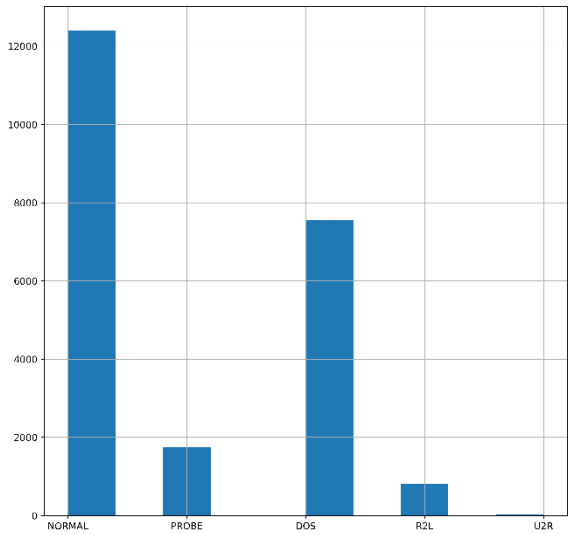
NSL-KDD یک نسخه به‌روز شده از مجموعه‌داده KDD Cup 99 به‌حساب می‌آید که معیار مؤثری برای محققان جهت مقایسه انواع روش‌های IDS، ساخت یک IDS (مبتنی بر میزبان یا مبتنی بر شبکه) و انجام برخی آزمایش‌ها در حوزه امنیت سایبری است. در این مجموعه‌داده، هیچ رکورد تکراری در زیرمجموعه تست وجود ندارد و تعداد کافی از رکوردها در مجموعه‌داده‌های آموزش و تست موجود است. در هر رکورد ۴۱ نوع ویژگی وجود دارد که به انواع حمله یا نرمال اختصاص داده شده‌اند. هر ویژگی در ۳ دسته از انواع مقداردهی اسمی، باینری و عددی طبقه‌بندی شده است. این همه انواع کلاس‌های حمله اساساً در ۴ بخش طبقه‌بندی می‌شوند که به‌عنوان DoS، Probing، U2R و R2L شناخته شده‌اند (جدول ۱) [۲۸]. همچنین، توزیع رکوردها در کلاس‌های مذکور برای زیرمجموعه داده‌های آموزش و تست به ترتیب در شکل‌های (۲) و (۳) ارائه شده‌اند.

جدول (۱). کلاس‌بندی حملات مجموعه‌داده NSL-KDD

کلاس حمله	۲۲ نوع حمله
DOS	back, land, neptune, pod, smurf, teardrop
R2L	ftp_write, guess_passwd, imap, multhop, phf, spy, warezclient, warezmaster
U2R	buffer_overflow, perl, loadmodule, rootkit
Probing	ipsweep, nmap, portsweep, satan



شکل (۲). توزیع رکوردها در کلاس‌های حمله برای مجموعه‌داده آموزش NSL-KDD



شکل (۳). توزیع رکوردها در کلاس‌های حمله برای مجموعه‌داده تست NSL-KDD

۲-۲- ماژول انتخاب ویژگی

تکنیک انتخاب ویژگی پیشنهادی از طریق حداکثرسازی وابستگی و حداقل‌سازی افزونگی اقدام به بهبود مؤثر کارایی IDS کرده است. اجرای این تکنیک در ۲ مرحله صورت می‌گیرید؛ ابتدا، ۳ روش انتخاب ویژگی کارا به‌صورت موازی ویژگی‌های بهینه منفرد خود را در سه زیرمجموعه استخراج می‌نمایند. سپس، یک الگوریتم رأی‌گیری مبتنی بر اشتراک پیاده‌سازی می‌گردد تا ویژگی‌های مشترک این سه زیرمجموعه را به‌عنوان ورودی‌های بهینه مدل گزینش کند.

گام اول: در این گام سه روش انتخاب ویژگی به‌صورت موازی و هم‌زمان اجرا می‌شوند. روش اول با بهره‌گیری از مفهوم اطلاعات متقابل (MI) [۲۹] به انتخاب ویژگی‌های بهینه تحت عنوان روش اطلاعات متقابل اصلاح‌شده (MMI) می‌پردازد که در آن اطلاعات MI بین ویژگی‌ها با کلاس و میانگین متحرک کلاس به‌عنوان امتیاز مثبت جهت حداکثرسازی و MI بین ویژگی‌های کاندید به‌عنوان امتیاز منفی جهت حداقل‌سازی در نظر گرفته می‌شود. امتیاز مثبت ذکرشده با فرض ویژگی، کلاس و میانگین متحرک از طریق رابطه زیر محاسبه می‌گردد:

$$MI\left(E,C,MA\right)=$$
$$\sum_{i}\sum_{j}\sum_{l}p\left(E_i,C_j,MA_l\right)\log\left(\frac{p\left(E_i,C_j,MA_l\right)}{p\left(E_i\right)p\left(C_j\right)p\left(MA_l\right)}\right) \quad (۱)$$

به‌طوری‌که $p\left(F_i,C_j,MA_l\right)$ بیانگر چگالی احتمال مشترک و $p\left(F_i\right)$، $p\left(C_j\right)$ و $p\left(MA_l\right)$ توابع چگالی احتمال حاشیه‌ای هستند. از طرف دیگر، با تعریف افزونگی به‌عنوان MI بین ویژگی‌های کاندید، زیرمجموعه نهایی ویژگی‌ها خواهد بود:

$$S=S\cup\arg\max_{E_i\in F-S}\left[MI\left(E,C,MA\right)-\underset{E_j\in S}{MI}\left(E_i,E_j\right)\right] \quad (۲)$$

در این حین، الگوریتم ژنتیک به‌عنوان روش دوم انتخاب ویژگی با بهره‌گیری از تکامل طبیعی ژن‌های موجودات زنده توسعه می‌یابد تا زیرمجموعه دوم از ویژگی‌های منتخب را به دست آورد [۱۷]. روش سوم انتخاب ویژگی، آزمون F تحلیل واریانس نامیده می‌شود که از مفهوم نسبت واریانس بین میانگین نمونه‌ها طبق رابطه زیر بهره می‌گیرد تا به انتخاب زیرمجموعه‌ای مناسب از ویژگی‌ها دست یابد [۳۰]:

$$F=\frac{variation\_between\_sample\_means}{variotion\_within\_the\_samples} \quad (۳)$$

گام دوم: در این بخش از یک روند رأی‌گیری مبتنی بر اشتراک استفاده می‌گردد به صورتی که اگر زیرمجموعه‌های منتخب از گام پیشین $F_{s1}=\{f_{m1},...,f_{mn}\}$، $F_{s2}=\{f_{a1},...,f_{an}\}$ و $F_{s3}=\{f_{g1},...,f_{gn}\}$ اقتباس‌شده از مجموعه ویژگی‌های اولیه باشند، ویژگی‌های منتخب نهایی با بالاترین امتیاز حاصل از رأی‌گیری به‌عنوان ویژگی‌های مشترک در این سه زیرمجموعه برگزیده خواهند شد؛ یعنی:

$$F_{s4}=F_{s1}\bigcap F_{s2}\bigcap F_{s3} \quad (۴)$$

به‌طوری‌که $\cap$ بیانگر اشتراک‌گیری بین زیرمجموعه‌ها است. شبه کد روند انتخاب ویژگی مطرح‌شده در شکل (۴) بیان شده است.

```
Input: Historical dataset
Output: A subset of features
        //Preprocessing dataset
        Prep_DS = Replace missing and defective
            values with the average value
        //Normalizing dataset
        Prep_DS = MinMaxScaler( Prep_DS )
      SelectedFeatures _GA =
            Genetic_Algorithm(Prep_DS)
      SelectedFeatures _Anova = Anova(Prep_DS)
      SelectedFeatures _MMI =
            Argmax(MI(Prep_DS,Class,Movong_Av
            erage)-
            MI(Prep_DS,Candidate_Feature))
      SelectedFeatures = INTERSECTION(
                        SelectedFeatures_GA,
                        SelectedFeatures _Anova,
                        SelectedFeatures _MMI)
    Return  SelectedFeatures
end
```

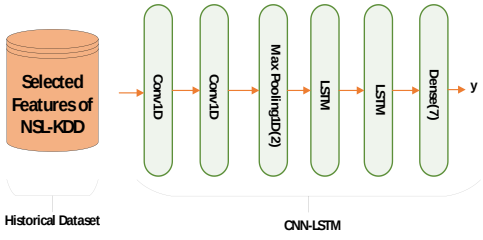
شکل (۴). شبه‌کد روند انتخاب ویژگی پیشنهادی

۲-۳- ماژول طبقه‌بندی

هدف این کار، دستیابی به IDS با بالاترین دقت با ارائه یک مدل قدرتمند جهت طبقه‌بندی داده‌ها است؛ بنابراین، یک مدل ترکیبی جدید با استفاده از الگوریتم‌های CNN و LSTM [۱۰] معرفی شده است.

ثابت شده است که CNN می‌تواند دقت طبقه‌بندی داده‌ها را بهبود بخشد به‌طوری‌که قادر به استخراج ویژگی‌های محلی و عمیق و همچنین وابستگی‌های طولانی‌مدت از طریق شبکه عصبی بازگشتی (RNN) است، بنابراین برای مدل‌سازی داده‌های متوالی مناسب بوده است. از طرفی، قابلیت پردازش و یادگیری الگوهای پیچیده شبکه‌های LSTM به دلیل حافظه داخلی بسیار سودمند و کاربردی است.

شبکه عصبی عمیق در نظر گرفته شده، در شکل (۵) نشان داده شده است. ویژگی‌های انتخاب‌شده از مرحله قبل به‌عنوان ورودی به DNN تغذیه می‌شوند. این شبکه از دو لایه CNN یک‌بعدی تشکیل شده است، خروجی‌های CNN به لایه MaxPooling وارد می‌شوند. در ادامه، دو لایه LSTM وجود دارد. لایه متراکم به‌عنوان آخرین لایه شامل ۷ نورون و یک تابع فعال‌سازی SoftMax است که عملیات طبقه‌بندی را انجام می‌دهد.



شکل (۵). شمای کلی مدل CNN-LSTM پیشنهادی

در ادامه به هدف یافتن بهترین مقادیر پارامترها برای تعداد فیلترهای CNN در لایه‌های اول و دوم، تعداد نورون‌های LSTM در هر دو لایه و تابع فعال‌سازی LSTM، یک روش بهینه‌سازی فراابتکاری ترکیبی معرفی خواهد شد.

۲-۴- ماژول بهینه‌سازی هایپرپارامترها

در سال‌های اخیر، الگوریتم‌های فراابتکاری ترکیبی در حوزه‌های مختلف جهت حل مسائل بهینه‌سازی توجه زیادی را به خود جلب کرده‌اند. به‌منظور تنظیم هایپرپارامترهای مدل CNN-LSTM، شناساگر پیشنهادی یک الگوریتم فراابتکاری ترکیبی را در نظر گرفته است. این الگوریتم RS-GWO-WOA نام دارد که ترکیبی از الگوریتم‌های GWO [۲۶] و WOA [۲۵] است. استراتژی‌های ترکیب مختلف با عملکردهای گوناگونی وجود دارند که در اینجا، رویکرد جابه‌جایی تصادفی (RS) پیشنهاد شده است. همچنین، ازآنجایی‌که هدف DNNها به حداقل رساندن هزینه در هر تکرار است، لازم است تابع هزینه در طول بهینه‌سازی حداقل شود.

الگوریتم GWO-WOA، با ادغام GWO و WOA از مزایای هر دو بهره می‌برد. WOA، موقعیت عوامل جستجو را با توجه به راه‌حلی که به‌صورت تصادفی تولیدشده بهروز می‌کند. WOA، از این استراتژی برای جلوگیری از گرفتار شدن در راه‌حل‌های بهینه محلی استفاده می‌نماید. در مقایسه، GWO موقعیت هر راه‌حل را بر اساس موقعیت سه تا از بهترین راه‌حل‌ها در جمعیت بهروز می‌کند. این استراتژی باعث می‌شود GWO بیشتر بهره‌برداری محور باشد. در نتیجه، WOA پتانسیل اکتشاف بالاتری نسبت به GWO دارد.

اکتشاف، قابلیت کاوش در مناطق مختلف جستجو است. یک اکتشاف بهتر باعث می‌شود که رویکرد حل به مقدار بهینه برسد. بهره‌برداری، توانایی تمرکز بر تشخیص راه‌حل دقیق‌تر در یک منطقه امیدوارکننده است. بهره‌برداری بهتر می‌تواند زمان اجرایی یک الگوریتم را کاهش دهد، اما به‌ندرت یک الگوریتم می‌تواند این دو معیار را به‌خوبی اصلاح کند زیرا در تضاد با یکدیگر هستند؛ بنابراین، یک تعادل مناسب بین اکتشاف و بهره‌برداری مطمئناً می‌تواند عملکرد الگوریتم را افزایش دهد و ما را به سمت تکنیک‌های ترکیبی با بهترین تناسب بین این دو مورد هدایت کند. ازاین‌رو، برای به دست آوردن این تعادل و استفاده از هر دو، این الگوریتم‌ها با جابه‌جایی تصادفی ترکیب شده‌اند، به‌طوری‌که نسخه RS-GWO-WOA توسعه داده شده است. RS-GWO-WOA در واقع به‌طور تصادفی بین GWO و WOA در هر تکرار جابه‌جا می‌شود؛ بنابراین زمان محاسباتی یافتن راه‌حل بهینه کاهش می‌یابد. به‌عنوان‌مثال، یک عدد تصادفی بین

۰ و ۱ تولید می‌شود. اگر عدد کوچک‌تر از ۰/۵ باشد، GWO اجرا می‌شود. در غیر این صورت، WOA بدون در نظر گرفتن اطلاعات عملکردی هر الگوریتم اعمال خواهد شد.

۳- نتایج تجربی و بحث

در این بخش، ابتدا به معرفی معیارهای ارزیابی و تنظیمات پیکربندی پرداخته می‌شود. در ادامه، پیاده‌سازی گام‌های مختلف روش پیشنهادی شامل انتخاب ویژگی، طبقه‌بندی و بهینه‌سازی هایپرپارامترها بر روی مجموعه‌داده موردنظر به همراه مقایسه با سایر تکنیک‌های طبقه‌بندی با هدف ارزیابی عملکرد رویه و دقت آن، توضیح داده می‌شود.

۳-۱- معیارهای ارزیابی عملکرد

برای هر مجموعه‌داده‌ای که در روش‌شناسی (آموزش، اعتبارسنجی و تست) استفاده می‌گردد، معیارهای ارزیابی برای درک رفتار مدل و توانایی پیش‌بینی آن موردمحاسبه قرار می‌گیرند. تنوع زیادی در انتخاب معیارهای ارزیابی برای سیستم‌های طبقه‌بندی وجود دارد. معیارهای مورداستفاده برای یک مطالعه تحلیلی باید مناسب حوزه مسئله باشد. یک ماتریس درهم‌ریختگی، می‌تواند برای توصیف عملکرد یک سیستم طبقه‌بندی بر اساس داده‌های تست مورداستفاده قرار گیرد (جدول ۲).

با افزایش تعداد کلاس‌ها، درک عملکرد کلی مدل با ماتریس درهم‌ریختگی ممکن است آسان نباشد؛ بنابراین، صحت به‌عنوان ساده‌ترین شکل ارزیابی جهت تعیین اینکه چند وقت یک‌بار طبقه‌بند صحیح رفتار کرده است به کار می‌رود. در برخی موارد، صحت نیز به‌تنهایی می‌تواند گمراه‌کننده باشد. این معیار بهتر است همراه با سایر معیارها استفاده گردد. دقت و یادآوری، عدم تعادلی که می‌تواند در یک مجموعه‌داده رخ دهد را برطرف می‌نمایند. دقت به این سؤال پاسخ می‌دهد که وقتی طبقه‌بند یک نمونه مثبت را پیش‌بینی می‌کند، چند وقت یک‌بار درست است. از سوی دیگر، معیار یادآوری پاسخگوی این است که هنگام ارزیابی تمام نمونه‌های مثبت در یک مجموعه‌داده، چه کسری از مجموعه‌داده را طبقه‌بند شناسایی کرده است. اگر طبقه‌بند به‌گونه‌ای تنظیم شود که یادآوری بسیار بالایی ارائه دهد، نرخ دقت کاهش می‌یابد. از طرف دیگر، طبقه‌بندهایی که برای نرخ دقت بالا بهینه‌سازی شده‌اند، از نرخ یادآوری ضعیف رنج می‌برند. به‌منظور ایجاد تعادل، معیار امتیاز F1 به‌عنوان میانگین وزنی هماهنگ دقت و یادآوری معرفی شده است.

جدول (۲). ماتریس درهم‌ریختگی

		پیش‌بینی‌شده	
		نرمال	حمله
واقعی	نرمال	TP	FN
	حمله	FP	TN

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (۵)$$

$$\text{Pr}ecision = \frac{TP}{TP + FP} \quad (۶)$$

$$\text{Re}call = \frac{TP}{TP + FN} \quad (۷)$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (۸)$$

۳-۲- تنظیمات پیاده‌سازی

در طول پیاده‌سازی شبکه‌های یادگیری، استفاده از متغیرهای تصادفی منجر به تولید نتایج متنوعی در هر اجرا می‌گردد. در این راستا، شبیه‌سازی موردنظر به تعداد ۱۰ بار اجرا شده است که نتیجه نهایی میانگین مقدار این دفعات است. مدل پیشنهادی در پایتون با ورژن ۳/۸ با استفاده از کتابخانه Keras برای اعمال شبکه‌های عصبی عمیق و Scikit-Learn، NiaPy و EvoPreprocess برای الگوریتم‌های فراابتکاری پیاده‌سازی شده است، درحالی‌که مقادیر batch_size=128 و epochs=50 در نظر گرفته شده‌اند.

۳-۳- ارزیابی مهندسی ویژگی

پس از آماده‌سازی اولیه مجموعه‌داده در گام پیش‌پردازش، نوبت به انتخاب ویژگی‌ها توسط روش پیشنهادی می‌رسد که در فاز اول آن تکنیک‌های MMI، GA و آزمون F تحلیل واریانس به ترتیب ۲۲، ۱۶ و ۱۶ ویژگی از ۴۱ ویژگی اولیه را برگزیده‌اند که در ادامه این سه زیرمجموعه وارد فاز رأی‌گیری اشتراکی شدند و در آنجا ۱۴ ویژگی به‌عنوان موارد برتر طبق جدول (۳) برگزیده شد.

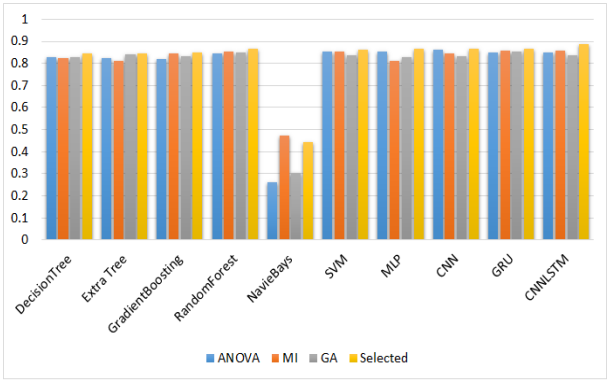
جدول (۳). ویژگی‌های برگزیده نهایی

ویژگی‌های انتخاب‌شده
src_bytes, dst_bytes, count,same_srv_rate,dst_host_count, dst_host_srv_count,dst_host_same_srv_rate, dst_host_diff_srv_rate,dst_host_same_src_port_rat, dst_host_srv_diff_host_rate,dst_host_srv_serror_rate, rerror_rate, dst_host_srv_rerror_rate, Flag

به‌منظور ارزیابی عملکرد انتخاب‌گر ویژگی بهبودیافته پیشنهادی، چهار زیرمجموعه ویژگی‌های منتخب در فاز مقایسه وارد طبقه‌بندهای DT، ET، GB، RF، NB، SVM، MLP، CNN، GRU و همچنین طبقه‌بند پیشنهادی (CNN-LSTM) شده‌اند که نتایج آن در جدول (۴) بر اساس معیار صحت لیست شده است. همان‌طور که در جدول (۴) مشخص است، تکنیک انتخاب ویژگی پیشنهادی دارای عملکرد بالاتری نسبت به سایر روش‌های منفرد است که به‌خصوص ادغام آن با طبقه‌بند پیشنهادی CNN-LSTM به‌عنوان یک مدل ترکیبی DL صحتی معادل ۰/۸۸۸۹۶ به‌عنوان بیشترین مقدار در میان سایرین را به خود اختصاص داده است. تمامی نتایج حاصل از جدول (۴) به‌صورت یک نمودار میله‌ای نیز در شکل (۶) مشاهده می‌گردد.

جدول (۴). مقایسه عملکرد تکنیک‌های انتخاب ویژگی

	MMI	GA	ANOVA	پیشنهادی
Decision Tree	۰/۸۲۲۰۲	۰/۸۲۶۱۹	۰/۸۳۰۲۷	**۰/۸۴۵۱۲**
Extra Tree	۰/۸۱۳۴۱	۰/۸۳۸۹۶	۰/۸۲۱۹۳	**۰/۸۴۵۵۷**
Gradient Boosting	۰/۸۴۳۶۷	۰/۸۳۴۱۷	۰/۸۲۱۳۵	**۰/۸۴۹۷۸**
Random Forest	۰/۸۵۲۰۵	۰/۸۴۸۶۳	۰/۸۴۶۷۷	**۰/۸۶۷۶۶**
Navie Bayes	۰/۴۷۲۰۰	۰/۳۰۵۶۵	۰/۲۶۰۰۹	**۰/۴۴۲۰۳**
SVM	۰/۸۵۳۷۴	۰/۸۳۵۵۹	۰/۸۵۳۲۵	**۰/۸۶۰۹۲۶**
MLP	۰/۸۱۰۳۵	۰/۸۲۹۵۱	۰/۸۵۴۰۹	**۰/۸۶۶۹۵**
CNN	۰/۸۴۶۱۵	۰/۸۳۲۹۷	۰/۸۶۲۰۳	**۰/۸۶۴۶۰۱**
GRU	۰/۸۵۶۱۳	۰/۸۵۳۱۱	۰/۸۵۱۰۳	**۰/۸۶۷۲۴**
CNN-LSTM	۰/۸۵۶۱۸	۰/۸۳۶۶۱	۰/۸۴۸۷۲	**۰/۸۸۸۹۶**



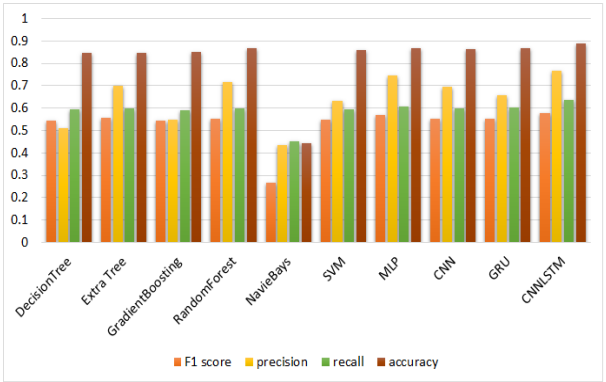
شکل (۶). نمودار میله‌ای ارزیابی عملکرد الگوریتم‌های انتخاب ویژگی

۳-۳- ارزیابی مدل طبقه‌بندی

در این گام، وارد ارزیابی روند طبقه‌بندی پیشنهادی CNN-LSTM با سایر طبقه‌بندهای رقیب منفرد ML و DL از قبیل DT، ET، GB، RF، NB، SVM، MLP، CNN، GRU شده‌ایم. البته، ویژگی‌های ورودی به این مدل‌ها حاصل از روش انتخابی پیشنهادی به‌عنوان الگوریتم برتر فاز پیشین است. این مقایسه بر اساس چهار معیار شناخته‌شده و کاربردی در زمینه IDS تحت عناوین صحت، دقت، یادآوری و امتیاز F1 انجام شده است که نتایج نهایی آن در جدول (۵) و همچنین شکل (۷) قابل‌مشاهده است. بر اساس این نتایج، مدل CNN-LSTM پیشنهادی با مقادیر ۰/۵۷۷۰۱۸، ۰/۷۶۵۲۷۶، ۰/۶۳۴۳۹۶ و ۰/۸۸۸۹۶ به ترتیب مربوط به امتیاز F1، دقت، یادآوری و صحت به بالاترین عملکرد در میان سایرین دست یافته است که می‌توان آن را ناشی از برتری مدل‌های ترکیبی یادگیری عمیق نسبت به سایر روش‌ها دانست.

جدول (۵). مقایسه عملکرد مدل‌های طبقه‌بندی

	صحت	یادآوری	دقت	امتیاز F1
Decision Tree	۰/۸۴۵۱۲	۰/۵۹۳۰۱	۰/۵۱۱۶۹	۰/۵۴۳۹۴
Extra Tree	۰/۸۴۵۵۷	۰/۶۰۰۱۲	۰/۷۰۰۰۳	۰/۵۵۸۴۰
Gradient Boosting	۰/۸۴۹۷۸	۰/۵۸۹۶۶	۰/۵۴۹۲۶	۰/۵۴۳۳۶
Random Forest	۰/۸۶۷۶۶	۰/۵۹۸۹۰	۰/۷۱۶۷۲	۰/۵۵۴۳۵
Navie Bayes	۰/۴۴۲۰۳	۰/۴۵۰۴۴	۰/۴۳۶۲۶	۰/۲۶۶۹۹
SVM	۰/۸۶۰۹۲	۰/۵۹۴۰۷	۰/۶۳۲۴۵	۰/۵۴۷۱
MLP	۰/۸۶۶۹۵	۰/۶۰۸۶۸	۰/۷۴۴۵۴	۰/۵۶۷۵۳
CNN	۰/۸۶۴۶۰	۰/۵۹۹۰۵	۰/۶۹۶۲۵	۰/۵۵۰۷۸
GRU	۰/۸۶۷۲۴	۰/۶۰۴۶۴	۰/۶۵۸۱۵	۰/۵۵۲۷۷
CNN-LSTM	**۰/۸۸۸۹**	**۰/۶۳۴۳**	**۰/۷۶۵۲**	**۰/۵۷۷۰۱**



شکل (۷). نمودار میله‌ای ارزیابی عملکرد مدل‌های طبقه‌بندی

۳-۵- ارزیابی الگوریتم بهینه‌سازی هایپرپارامترها

همان‌طور که پیش‌ازاین بیان شد، بهینه‌سازی تکنیک‌های طبقه‌بندی از طریق یافتن مناسب‌ترین مقادیر تنظیم آن‌ها توسط

روشی خودکار امکان‌پذیر است که در اینجا طبق جدول (۶)، این هایپرپارامترها توسط الگوریتم‌های منفرد GWO، WOA و روش پیشنهادی RS-GWO-WOA تنظیم شده‌اند و بر اساس معیار صحت در فاز قیاس قرار گرفته‌اند. با توجه به جدول (۶)، بهینه‌ساز فراابتکاری ترکیبی پیشنهادی با اختصاص صحت ۰/۹۲۸۱۶۴ به خود، مدل موردنظر را به بهینه‌ترین حالت ممکن نسبت به بقیه روش‌های رقیب تبدیل کرده است.

جدول (۶). مقایسه نتایج با الگوریتم‌های بهینه‌سازی فراابتکاری

	تعداد فیلتر لایه اول	تعداد فیلتر لایه دوم	تعداد نورون لایه اول	تعداد نورون لایه دوم	تابع فعال‌سازی	صحت
GWO	۴۰	۳۰	۲۰	۱۵	selu	۰/۸۹۹۹۶
WOA	۳۰	۲۰	۲۰	۲۰	tanh	۰/۹۰۸۹۶
RS-GWO-WOA	۳۰	۳۰	۲۵	۱۵	selu	۰/۹۱۸۱۶

۳-۶- ارزیابی زمان آموزش و تست

این بخش از ارزیابی با اندازه‌گیری مدت‌زمان آموزش و تست برای مدل‌های ساخته شده از تکنیک‌های انتخاب ویژگی و طبقه‌بندی مذکور در گام‌های پیشین بر روی مجموعه‌داده آموزش و تست صورت گرفته است. نتایج مربوط به آموزش و تست به ترتیب در جدول (۷) و (۸) بیان شده‌اند که طبق آن‌ها مدل یادگیری عمیق ترکیبی توسعه‌یافته دارای مدت‌زمان یادگیری و تست کمتری نسبت به سایر الگوریتم‌های یادگیری عمیق منفرد است. همچنین، مدل پیشنهادی دارای نتایج قابل‌مقایسه‌ای نسبت به سایر روش‌های یادگیری ماشین است، درحالی‌که عملکرد آن به‌مراتب بهتر از آن‌ها بوده است.

جدول (۷). مقایسه عملکرد تکنیک‌ها بر اساس زمان آموزش

	MMI	GA	ANOVA	پیشنهادی
Decision Tree	۱/۰۲۸۸۵	۰/۴۴۳۰۵	۱/۰۷۴۴۳	**۰/۳۷۴۵۱۱**
Extra Tree	۲/۶۰۵۳۵	۱/۳۶۶۶	۲/۵۰۹۵۹	**۱/۱۶۶۸۵۱**
GradientBoosting	۱۰۷/۱۸۲	۵۷/۲۳۷۸	۱۳۶/۳۴۳	**۴۰/۰۲۹۵۹**
Random Forest	۰/۱۳۷۹۲	۱/۷۸۰۲۷	۴/۰۶۴۹۸	**۰/۷۴۳۱۳**
Navie Bayes	۰/۱۵۰۸۸	۰/۱۱۴۱۳	۰/۱۴۱۶۰	**۰/۱۰۷۲۰۸**
SVM	۹۴/۵۲۶۴	۱۱۱/۷۲۵	۱۱۰/۰۲۱	**۷۱/۱۷۰۰۶**
MLP	۲۰۵/۴۷۱	۱۲۱/۹۰۳	۲۰۹/۵۰۳	**۱۰۴/۲۷۳**
CNN	۱۳۲۳/۷۳	۱۲۳۶/۵۸	۱۱۹۳/۷۵	**۸۴۸/۹۵۱**
GRU	۱۳۹۴/۲۵	/۰۹۶ ۲۱۴۲	۲۰۳۳/۴۵	**۱۲۶۷/۵۷**
CNN-LSTM	۸۳/۷۶۱۳	۹۴/۱۸۶۵	۱۱۳/۶۵۰	**۶۶/۰۲۹۷۱**

جدول (۸). مقایسه عملکرد تکنیک‌ها بر اساس زمان تست

	MMI	GA	ANOVA	پیشنهادی
Decision Tree	۰/۰۲۱۵۷	۰/۰۰۷۹۷	۰/۰۰۷۹۷	**۰/۰۰۶۶**
Extra Tree	۰/۱۵۱۳۰	۰/۰۸۰۷۸	۰/۱۰۷۹۸	**۰/۰۷۲۹۴۷**
GradientBoosting	۰/۲۰۳۹۰	۰/۱۲۵۷۰	۰/۲۶۰۱۴	**۰/۰۶۶۷۵**
Random Forest	۰/۱۱۰۶۸	۰/۰۶۶۸۷	۰/۱۲۱۶۷	**۰/۰۷۴۹۴۱**
Navie Bayes	۰/۰۷۴۸۱	۰/۰۵۰۴۳	۰/۰۴۴۴۸	**۰/۰۴۷۸۱۱**
SVM	۱۰/۱۶۲۹	۱۰/۵۲۸۶	۱۴/۳۴۰۰	**۷/۴۹۲۳۲**
MLP	۲۰۶/۱۷۷	۱۲۲/۳۱۹	۲۱۰/۳۷۹	**۱۰۴/۸۱۰۷**
CNN	۱/۸۵۴۱۶	۲/۲۸۹۲۵	۱/۴۴۲۷۸	**۱/۲۱۸۲۹۹**
GRU	۴/۱۳۶۳۴	۳/۶۴۱۱۳	۴/۵۲۹۵۰	**۳/۵۹۴۵۸**
CNN-LSTM	۰/۷۷۳۸۹	۰/۹۵۲۴۰	۱/۱۵۲۹۸	**۰/۶۳۱۷۵**

۴- نتیجه گیری

این پژوهش یک چارچوب یادگیری عمیق مبتنی بر ترکیب جدید متشکل از انتخاب ویژگی، طبقه‌بندی و بهینه‌سازی هایپرپارامترها فراهم کرده است تا در عین کاهش هشدارهای نادرست، موجب افزایش عملکرد سیستم شود. نوآوری کلیدی این پژوهش در سه گام مجزا ارائه شده است؛ در گام اول یک الگوریتم رأی‌گیری اشتراکی جهت بهره‌وری از سه تکنیک انتخاب ویژگی MMI، GA و آزمون F تحلیل واریانس معرفی شده است که با انتخاب ویژگی‌های مربوطه و غیرزائد موجب کاهش ابعاد ورودی و زمان آموزش و تست شده است. در ادامه، ارائه یک مدل یادگیری عمیق ترکیبی (CNN-LSTM) جهت طبقه‌بندی داده‌ها منجر به بهبود چشمگیر قابلیت اطمینان سیستم بر اساس معیارهای ارزیابی شده است، درحالی‌که زمان اجرای آن نیز از سایر الگوریتم‌های یادگیری عمیق کمتر بوده و با مدل‌های یادگیری ماشین قابل‌مقایسه است. سپس، از یک روند بهینه‌سازی فراابتکاری ترکیبی (RS-GWO-WOA) جهت بهینه‌سازی هایپرپارامترهای مدل طبقه‌بند استفاده شده است که تأثیر مثبت خود را در بهینه‌سازی مدل بر اساس معیار صحت نشان داده است.

به‌عنوان پژوهش آتی، با توجه به مزایای روش‌های مبتنی بر ترکیب بر دقت و عملکرد سیستم‌های IDS می‌توان به سراغ ساخت سایر روش‌های ترکیبی رفت. همچنین، با توجه به قابلیت عملکرد بالای مدل پیشنهادی می‌توان آن را روی سایر مجموعه‌داده‌های موجود و کاربردهای طبقه‌بندی دیگر توسعه داد.

۶- مراجع

[1] Di Mauro, M; Galatro, G; Fortino, G; & Liotta, A, "Supervised Feature Selection Techniques in Network Intrusion Detection: A Critical Review," Engineering Application of Artificial Intelligence, Vol. 101, pp. 104-216, 2021.

[2] Thakkar, A. & Lohiya, R. "A Survey on Intrusion Detection System: Feature Selection, Model, Performance Measures, Application Perspective, Challenges, and Future Research Directions," Artifical Intellignce Review, Vol. 55, No. 1, pp. 453-563, 2022.

[3] Almasoudy, F. H. Al-Yaseen, W. L. & Idrees, A. K. "Differential Evolution Wrapper Feature Selection for Intrusion Detection System," Procedia Computer Science, Vol. 167, pp. 1230-1239, 2020.

[4] Musa, U. S. Chakraborty, S. Abdullahi, M. M. & Maini, T. "A Review on Intrusion Detection System using Machine Learning Techniques," International Conference on Computing, Communication, and Intelligence Systems (ICCCIS). IEEE, pp. 541-549, 2021.

[5] Panigrahi, R. Borah, S. Bhoi, A. K. Ijaz, M. F. Pramanik, M. Jhaveri, R. H. & Chowdhary, C. L. "Performance Assessment of Supervised Classifiers for Designing Intrusion Detection Systems: A Comprehensive Review and Recommendations for Future Research," Mathematics, Vol. 9, No. 6, pp. 690, 2021.

[6] Ozkan-Okay, M. Samet, R. Aslan, Ö. & Gupta, D. "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," IEEE Access, Vol. 9, pp. 157727-157760, 2021.

[7] Hodo, E. Bellekens, X. Hamilton, A. Tachtatzis, C. & Atkinson, R. "Shallow and Deep Networks Intrusion Detection System: A Taxonomy and Survey," arXiv preprint, arXiv:1701.02145, 2017.

[8] Lansky, J. Ali, S. Mohammadi, M. Majeed, M. K. Karim, S. H. T. Rashidi, S. ... & Rahmani, A. M. "Deep Learning-Based Intrusion Detection Systems: A Systematic Review," IEEE Access, Vol. 9, pp. 101574-101599, 2021.

[9] Lee, S. W. Mohammadi, M. Rashidi, S. Rahmani, A. M. Masdari, M. & Hosseinzadeh, M. "Towards Secure Intrusion Detection Systems using Deep Learning Techniques: Comprehensive Analysis and Review," Journal of Network Computer Application, Vol. 187, pp. 103-111, 2021.

[10] Ahmad, Z. Shahid Khan, A. Wai Shiang, C. Abdullah, J. & Ahmad, F. "Network Intrusion Detection System: A Systematic Study of Machine Mearning and Deep Learning Approaches," Transaction on Emerging Telecommunications Technologies, Vol. 32, No. 1, pp. 41-50, 2021.

[11] Akay, B. Karaboga, D. & Akay, R. "A Comprehensive Survey on Optimizing Deep Learning Models by Metaheuristics," Artifical Intelligence Review, pp. 1-66, 2022.

[12] Abd Elaziz, M. Dahou, A. Abualigah, L. Yu, L. Alshinwan, M. Khasawneh, A. M. & Lu, S. "Advanced Metaheuristic Optimization Techniques in Applications of Deep Neural Networks: a Review," Neural Computing and Applications, pp. 1-21, 2021.

[13] Wang, C. Ye, X. He, X. Tian, Y. & Gong, L. "Two-Level Feature Selection Method for Low Detection Rate Attacks in Intrusion Detection," Internation Conference, Security and Privacy in New Computing Environments, Springer, Cham, pp. 689-696, 2019.

[14] Wang, Z. Tang, M. Deng, J. Wang, Y. Qian, J. & Chen, X. "A New Feature Selection Method for Intrusion Detection," IEEE International Conference on Ubiquitous Computing and Communication (IUCC) Data Science Computational Intelligence (DSCI) Smart Computing Networking and Services (SmartCNS), pp. 298-304 IEEE, 2019.

[15] Najafi, M. & Rafeh, R. "A New Light Weight Intrusion Detection Algorithm for Computer Networks," Advance Defence Science Technology, Vol. 8, No. 29, pp. 191-200, 2017. (In Persian)

[16] Keshavarzi, M.; & Momenzadeh, H. "Improving Intrusion Detection Systems by Feature Reducing Based on Genetic Algorithm and Data Mining Techniques," Communication Engineering, Vol. 8, No. 32, pp. 1-13, 2019. (In Persian)

[17] Parsa, S.; & Aarabi, S. H. R. "A New Approach to Network Intrusion Detection Based on Hybrid Methods," Electronic and Cyber Defence, Vol. 5, No. 3, pp. 79-93, 2017. (In Persian)

[18] Gaur, V.; & Kumar, R. "Analysis of Machine Learning Classifiers for Early Detection of DDoS Attacks on IoT Devices," Arabian Journal for Science Engineerig, Vol. 47, No. 2, pp. 1353-1374, 2022.

[19] Hooshmand, M. K. "Using Ensemble Learning Approach to Identify Rare Cyber-Attacks in Network Traffic Data," International Conference on Advanced Computer science and Information Systems (ICACSIS) IEEE, pp. 141-146, 2020.

[20] Kazemitabar, J.; Taheri, R.; & Kheradmandian, GH. " A Novel Technique for Improvement of Intrusion Detection via Combining Random Forrest and Genetic Algorithm," Advanced Defence Science Technology, Vol. 10, No. 37, pp. 287-296, 2019. (In Persian)

[21] Taheri, R. Parsaei, M. R. & Javidan, R. "Real-Time Intrusion Detection System Using a Combination of Discretization and Feature Selection," Advanced Defence Science Technology, Vol. 8, No. 29, pp. 251-263, 2017. (In Persian)

[22] Upadhyay, D. Manero, J. Zaman, M. & Sampalli, S. "Intrusion Detection in SCADA Based Power Grids: Recursive Feature Elimination Model with Majority Vote Ensemble Algorithm," IEEE Transaction on Network Science and Engineering, Vol. 8, No. 3, pp. 2559-2574, 2021.

[23] Hassan Nataj Solhdar, M. Janinasab Solahdar, M. & Eskandari, S. "An Intrusion Detection System with a Parallel Multi-Layer Neural Network," Journal of Mathematical Modeling, Vol. 9, No. 3, pp. 437-450, 2021.

[24] Haghighat, M. H. & Li, J. "Intrusion Detection System using Voting-based Neural Network," Tsinghua Science and Technology, Vol. 26, No. 4, pp. 484-495, 2021.

[25] Haghnegahdar, L. & Wang, Y. "A Whale Optimization Algorithm-Trained Artificial Neural Network for Smart Grid Cyber Intrusion Detection," Neural Computing and Applications, Vol. 32, No. 13, pp. 9427-9441, 2020.

[26] Yang, H. & Zhou, Z. "A Novel Intrusion Detection Scheme using Cloud Grey Wolf Optimizer," 37th Chinese Control Conference (CCC) IEEE, pp. 8297-8302, 2018.

[27] Chen, C. Song, L. Bo, C. & Shuo, W. "A Support Vector Machine with Particle Swarm Optimization Grey Wolf Optimizer for Network Intrusion Detection, " International Conference on Big Data Analysis Computer Science (BDACS) IEEE, pp. 199-204, 2021.

[28] Bala, R. & Nagpal, R. "A Review on Kdd cup99 and Nsl Nsl-kdd Dataset," International Journal Advanced Reserarch in Computer Science, Vol. 10, No. 2, 2019.

[29] Amiri, F. Yousefi, M. R. Lucas, C. Shakery, A. & Yazdani, N."Mutual Information-based Feature Selection for Intrusion Detection Systems," Journal of Network and Computer Applications, Vol. 34, No. 4, pp. 1184-1199, 2011.

[30] Shakeela, S. Shankar, N. S. Reddy, P. M. Tulasi, T. K. & Sai, M. M. "Optimal Ensemble Learning Based on Distinctive Feature Selection by Univariate ANOVA-F Statistics for IDS," International Journal of Electronics and Telecommunications, Vol. 67, No. 2, pp. 267-275, 2021.