ارائهٔ یک الگوی فراابتکاری تشخیص نفوذ به‌کمک انتخاب ویژگی مبتنی بر بهینه‌سازی گرگ خاکستری بهبودیافته و جنگل تصادفی

شهریار محمدی*، احمد خلعتبری، مهدی باباگلی

دانشکده صنایع، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، ایران

چکیده

پیشرفت سریع در زمینه‌های اینترنت و ارتباطات منجر به رشد چشمگیر شبکه‌های رایانه‌ای، اندازهٔ شبکه و تبادلِ داده شده و همین امر زمینه را برای حملات مختلف فراهم کرده است. سامانه‌های تشخیص نفوذ نقش مهمی در امنیّت شبکه‌های اینترنتی بر عهده دارند و با بازرسی ترافیک‌های شبکه از محرمانگی، یک‌پارچگی و دردسترس‌بودن شبکه محافظت می‌کند. الگوهای تشخیص نفوذ در زمینه امنیّت شبکه، الگوهای پیش‌بینی‌کننده‌ای هستند که در جهت پیش‌بینی داده‌های ترافیکی نفوذ در شبکه‌ها به کار می‌روند و یکی از پرکاربردترین الگوها در سامانه‌های تشخیص نفوذ الگوهای مبتنی بر یادگیری ماشین هستند. نبود توازن و تعادل بین دقّت تشخیص و نرخ هشدار کاذب یکی از چالش‌های بزرگ در این زمینه محسوب می‌شود. در این مقاله برای افزایش قدرت جستجو از الگوریتم‌های فراابتکاری و جهت افزایش قدرت محاسباتی و رده‌بندی از روش یادگیری ماشین استفاده می‌شود. ازاین‌رو، در این پژوهش الگوی کارای مبتنی بر الگوریتم‌های گرگ خاکستری دودویی بهبودیافته و جنگل تصادفی، جهت شناسایی بهترین مجموعه‌ویژگی‌های ترافیک برای تشخیص و پیشگیری از حملات اینترنتی ارائه می‌شود. جهت پیداکردن بهترین زیرمجموعه از الگوریتم گرگ خاکستری، و برای ارزیابی هر زیرمجموعه از جنگل تصادفی استفاده می‌شود. همچنین، به‌منظور بهبود عملکرد گرگ خاکستری، این الگوریتم بهبود داده می‌شود. دقّت حاصل‌شده برای طبقه‌بندی صحیح در روش پیشنهادی در مجموعه‌داده‌ها NSL-KDD در روش گرگ خاکستری سنّتی و بهبودیافته به‌ترتیب برابر با ۹۷,۱۴ و ۹۸,۹۷ درصد است که در مقایسه با روش‌های دیگر دارای دقّت بالاتری است.

واژگان کلیدی: تشخیص نفوذ، انتخاب ویژگی، الگوریتم گرگ خاکستری بهبودیافته، جنگل تصادفی، یادگیری ماشین

Proposing a Meta-heuristic Model of Intrusion Detection Using feature Selection Based on Improved Gray Wolf Optimization and Random Forest

Shahriar Mohammadi*, Ahmad Khalatbari, Mehdi Babagoli

Faculty of Industry, Khwaja Nasiruddin Toosi University of Technology, Tehran, Iran

Abstract

Rapid development in the Internet and communications have been led to dramatic growth in computer networks, network size, and data exchange; hence, this can pose harmful threats to the network. Intrusion detection systems play an important role in Internet networks security, which protects the privacy, integrity, and availability of the network by inspecting network traffic. Intrusion detection models in the field of network security are predictive models that are used to predict malicious data in networks and one of the most widely used models in intrusion detection systems is based on machine learning. The imbalance between the accuracy of detection and false alarm rate is one of the most important challenges in this regard. In this paper, meta-heuristic algorithms are used to increase searchability and machine learning method as well, to increase computational power and classification. Several evaluation models have been developed recently that can consider the merit of a feature subset

* نویسندهٔ عهده‌دار مکاتبات

* Corresponding author

• تاریخ ارسال مقاله: ۱۴۰۰/۱/۳ • تاریخ پذیرش: ۱۴۰۱/۵/۹ • تاریخ انتشار: ۱۴۰۲/۵/۲۰ • نوع مطالعه: پژوهشی

[DOI: 10.61186/jsdp.20.1.133]

instead of evaluating individual features. Stochastic nature and effective search capability of meta-heuristic algorithms play an important role in solving the high dimensional problem. Therefore, in order to detect intrusion and prevent it, an efficient model based on the gray wolf optimization is used to select the most relevant feature and random forest used as an evaluator. Gray wolf optimization (GWO) is a metaheuristic algorithm that inspired from hunting behavior of social hierarchy of grey wolves. According to less randomness and varying numbers of individuals assigned in global and local searching procedures, the GWO algorithm is easier to use and converges more rapidly and the superiority of this algorithm among many metaheuristic algorithms has been proved in many researches. The binary version of this algorithm is used for feature selection method. The procedure of proposed model is described as follows.

1. **NSL-KDD dataset is a benchmark dataset that is consists of normal and abnormal traffics. despite the oldness, the NSL-KDD dataset is analyzed and used in many recent studies in order to evaluate the effectiveness of the various classification algorithms in intrusion detections. In preprocessing phase, the data normalization is conducted and the class labels are converted to normal and abnormal (0 and 1).**
2. **Binary GWO is used to select best feature subset by exploring and exploiting the search space. The high strength of GWO in finding the optimal feature subset has been originated from three member of grey wolves' pack: Alpha, beta and delta. The random forest technique is used as a classifier in this model. The generated subsets are evaluated with random forest.**
3. **In order to increase the performance of GWO, an improved GWO (IGWO) is proposed. The proposed IGWO is used genetic algorithm nature for making balance between exploitation and exploration. In each iteration, alpha and beta are considered as parents and produced two individuals (child) using uniform crossover. The individuals can add to papulation if they have a good merit. The merit of all individuals is obtained using random forest classifier.**
4. **As shown in the result, the detection accuracy of the traditional and improved gray wolf method is obtained 97.14% and 98.97%, respectively, which is outperformed other methods.**
5.

Keywords:

Intrusion detection system, Feature Selection, improved Gray Wolf Optimization Algorithm, Random Forest, Machine Learning

۱- مقدمه

اینترنت امروزه بخشی از زندگی روزمرّه و یک ابزار ضروری شده و این موضوع به مردم در حوزه‌های زیادی مانند کسب‌وکار، سرگرمی، تحصیل و غیره کمک می‌کند. به‌ویژه، اینترنت به‌عنوان یکی از اجزای مهم الگوهای کسب‌وکار استفاده می‌شود. در عملیّات تجاری، هم شرکت و هم مشتریان از برنامه‌های کاربردی اینترنت نظیر وب‌سایت و پست الکترونیکی در فعّالیّت‌های تجاری بهره می‌گیرند. ازاین‌رو، به امنیّت اطّلاعات در استفاده از اینترنت به‌عنوان یک رسانهٔ گروهی باید با دقّت توجّه شود. تشخیص نفوذ، مسئله پژوهشی مهمی برای شبکه‌های تجاری و شخصی است. سامانه‌های تشخیص نفوذ (IDS[1]) در حال حاضر جزء اصلی‌ترین و کامل‌ترین قسمت‌های یک سامانهٔ پایش یا نظارت شبکه هستند. سامانه‌های تشخیص نفوذ دارای فناوری‌های تقریباً جدیدی هستند و این نوید را می‌دهند که به ما در جهت شناسایی نفوذهایی که به شبکه انجام می‌شود، کمک خواهند کرد. تشخیص نفوذ[2]، درواقع، فرآیندی است که در آن رویدادها و رخدادهای یک سامانه، یا شبکه پایش شده و بر اساس آن پایش‌ها، وقوع نفوذ به آن شبکه یا سامانه تشخیص داده می‌شود. یک نفوذ در واقع، فعّالیّت یا عملی است که توسط آن محرمانگی، صحّت، تمامیت و یا دسترسی به منابع دچار اختلال و یا تعرض می‌شود[۱]. در چند دههٔ گذشته، در حوزهٔ اینترنت و سامانه‌های رایانه‌ای مشکلات امنیّتی متعددی با توجّه به حجم استفاده از شبکه‌ها مطرح شده است. به گزارش آماری مرکز CERT[3]، مقدار رسوخ در سامانه‌ها بیش‌از حد تصور، سال‌به‌سال افزایش یافته است. هرگونه نفوذ مخرّب و یا حمله باعث آسیب‌پذیری شبکه، سامانه‌های رایانه‌ای، یا اطّلاعات می‌شود و ممکن است منجر به حوادث جدی از جمله نقض سیاست‌های امنیّتی رایانه، و به‌عنوان مثال، محرمانگی، یکپارچگی و در دسترس بودن آن‌ها شود[۲]. نفوذگرها به‌طور عموم، از عیوب نرم‌افزاری، شکستن کلمات رمز، استراق سمع ترافیک شبکه و نقاط ضعف طراحی در شبکه، سرویس‌ها و یا رایانه‌های شبکه برای نفوذ به سامانه‌ها و شبکه‌های رایانه‌ای بهره می‌برند. به‌منظور مقابله با نفوذگران به سامانه‌ها و شبکه‌های رایانه‌ای، روش‌های بی‌شماری تحت عنوان روش‌های تشخیص نفوذ ایجاد شده است که عمل نظارت بر وقایع

[1] Intrusion detection system
[2] Intrusion Detection
[3] Computer Emergency Readiness Team (CERT)

[DOI: 10.61186/jsdp.20.1.133]

اتفاق افتاده در یک سامانه یا شبکهٔ رایانه‌ای را بر عهده دارند. در یک دسته‌بندی مهم روش‌های تشخیص نفوذ به سه دستهٔ مبتنی بر امضا[1]، مبتنی بر ناهنجاری[2] و ترکیبی تقسیم‌بندی می شوند. در روش‌های مبتنی بر ناهنجاری با استفاده از روش‌هایی نظیر یادگیری ماشین، هوش مصنوعی، الگوریتم‌های فراابتکاری به ایجاد یک پروفایل از ترافیک‌های نرمال شبکه پرداخته می‌شود تا هر عمل غیرنرمال به‌عنوان نفوذ تشخیص داده شود. از این رو، ایجاد یک الگوی کارا در ایجاد پروفایل بسیار مهم و حیاتی‌ست[۳]. تا به امروز، تهدید در شبکه و امنیّت اطّلاعات هنوز هم از مسائل مهم و قابل توجّه پژوهش‌ها هستند و ادبیات زیادی در بررسی و طبقه‌بندی روش‌های تشخیص نفوذ وجود دارد. به دلیل پیچیدگی ترافیک‌های شبکه و ابعاد بالای ویژگی‌ها بسیاری از پژوهش‌ها از روش انتخاب ویژگی جهت حذف ویژگی‌های زاید استفاده می‌کنند. از طرفی در دهه‌های اخیر، الگوریتم‌های فراابتکاری عملکرد شایسته‌ای در زمینه‌های حل مسائل پیچیدهٔ مختلف از خود نشان داده‌اند[۴]. الگوریتم‌های فراابتکاری را می‌توان باتوجّه‌به‌این‌که این الگوریتم‌ها مبتنی بر یک جواب هستند یا مبتنی بر جمعیّت، دسته‌بندی کرد. الگوریتم‌های مبتنی بر یک جواب در حین فرآیند جستجو یک جواب را تغییر می‌دهند؛ درحالی‌که الگوریتم‌های مبتنی بر جمعیّت در حین جستجو یک جمعیّت از جواب‌ها را در نظر می‌گیرند. الگوریتم‌های مبتنی بر یک جواب بر روی مناطق محلی جستجو تمرکز دارند؛ در مقابل الگوریتم‌های مبتنی بر جمعیّت، می‌توانند جستجو را به‌طور هم‌زمان، در مناطق مختلفی از فضای جواب انجام دهند[۵]. در این پژوهش از الگوریتم فراابتکاری مبتنی بر جمعیّت جهت انتخاب بهترین ویژگی‌ها و حذف ویژگی‌های اضافی استفاده شده است که به‌منظور بهبود عملکرد این الگوریتم روند جستجوی آن بهبود داده شده است. در گام بعدی به‌منظور رده‌بندی و ارزیابی دقیق‌تر ترافیک‌ها و افزایش دقّت تشخیص از الگوریتم جنگل تصادفی(RF[3]) استفاده شده است. الگوریتم جنگل تصادفی یکی از پرکاربردترین الگوریتم‌های یادگیری ماشین است که در بیشتر مواقع بهترین عملکرد را دارد. این الگوریتم نسخهٔ پیشرفتهٔ درخت تصمیم گیری(DT[4]) است که جزء یادگیری جمعی[5] محسوب می‌شود و از روش bagging برای آموزش استفاده می‌کند[۶]. ممکن است سامانهٔ تشخیص نفوذ، برخی از دسترسی‌های مجاز را نیز به‌عنوان دسترسی غیرمجاز و نفوذ تلقی کرده و هشدار نادرستی صادر کند؛ بنابراین، یکی از ایرادات مهمی که در سامانه‌های تشخیص نفوذ وجود دارد، نرخ هشدار نادرست ایجادشده آن‌هاست. ارائهٔ روشی که ضمن تشخیص درست و کامل نفوذها و دسترسی‌های غیرمجاز در شبکه و افزایش دقّت، بتواند هشدارهای نادرست کمتری را تولید کند، بسیار حیاتی است.

ساختار این مقاله به‌این‌ترتیب است: در بخش دو کارهای انجام‌شده معرفی می‌شود. در بخش سه روش پیشنهادی بیان می‌شود. نتایج، مقایسه‌ها و ارزیابی روش پیشنهادی در بخش چهار به‌تفصیل شرح داده می‌شود و در نهایت در بخش پنج، نتیجه‌گیری کلی بیان می‌شود.

۲- مرور کارهای گذشته

همان‌طورکه اشاره شد، مطالعات زیادی در زمینهٔ تشخیص نفوذ با به‌کارگیری روش‌های یادگیری ماشین، هوش مصنوعی و الگوریتم‌های فراابتکاری شده است. در روش‌های موجود به‌طور عموم، از الگوریتم‌های فراابتکاری به‌منظور رده‌بندی و آموزش و انتخاب ویژگی استفاده‌شده است. در [۷] با استفاده از روش یادگیری گروهی به تشخیص نفوذ و رده‌بندی انواع مختلف حملات و متمایز کردن آن‌ها با ترافیک‌های (عادی/طبیعی؟) نرمال پرداخته شده است. جهت ارزیابی الگوی پیشنهادی از دیتاست NSL-KDD استفاده شده است. الگوی پیشنهادی کارایی خوبی از خود نشان داده است. یکی از حملات مهم در شبکه‌های رایانه‌ای حمله سیل HTTP[6] است که از جمله راه‌کارهای ارائه‌شده برای تشخیص این حمله می‌توان استفاده الگوریتم جستجوی فاخته[7] را نام برد[۸]. در این مقاله الگوی BIFAD[8] ارائه شده است که با استفاده از ویژگی‌های ترافیک شبکه در برابر حملهٔ سیل HTTP دفاع کرده و از الگوریتم فراابتکاری فاخته برای رده‌بندی ترافیک نرمال از غیرنرمال استفاده می‌کند که دقّت تشخیص ۹۵ درصد و نرخ هشدار کاذب ۴,۵ درصد به دست آمده است. در [۹] پژوهشگران علاوه‌بر استفاده از روش‌های فراابتکاری، نظر افراد خبره را نیز برای بهبود سامانه تشخیص نفوذ دخیل کردند. به‌منظور همگرایی

[1] Signature base IDS
[2] Anomaly base IDS
[3] Random forest
[4] Decision Tree
[5] Ensemble learning
[6] HTTP flood
[7] Cuckoo search
[8] Bio-Inspired Anomaly Based Real Time Detection

[DOI: 10.61186/jsdp.20.1.133]

نظرات خبرگان از روش رأی‌دهی اکثریت وزن‌دار (WMV[1]) استفاده‌شده است که با سه الگوریتم ترکیب و بر روی دیتاست KDD cup ارزیابی شده است. در میان روش‌های پیاده‌سازی‌شده الگوی یادگیری گروهی مبتنی بر الگوریتم PSO بهترین عملکرد را از خود نشان داده است که دقّت تشخیص به‌صورت تقریبی ۹۳ درصد به دست آمده؛ امّا از نرخ هشدار کاذب صحبتی نشده است. از دیگر روش‌های مهم اخیر برای پیشگیری از حملهٔ DDoS[2] الگوریتم فراابتکاری شیر است[۱۰]. در این پژوهش از الگوریتم شیر به‌منظور انتخاب بهترین زیرمجموعه ویژگی و از الگوریتم CNN[3] به‌منظور رده‌بندی استفاده شده است که در مقایسه با الگوریتم کلونی زنبور عسل و کلونی مورچگان عملکرد بهتری داشته است. الگوی پیشنهادی در شبکه نرم‌افزارمحور (SDN[4]) ارزیابی شده و به‌منظور آموزش الگوریتم نیز از دیتاست KDD استفاده شده، که دقّت به‌دست‌آمده در تشخیص حمله DDoS برابر با ۹۶ درصد است. در [۱۱] الگوریتم بهینه‌سازی گلهٔ فیل - بسط تیلور[5] مبتنی بر شبکهٔ باور عمیق[6] را ارائه دادند که بر روی دیتاست KDD و دو دیتاست دیگر ارزیابی شده است و با دقّت ۹۳،۸۱ درصد توانسته حملات DDoS را تشخیص دهد. بعضی از پژوهش‌های اخیر نیز از الگوریتم‌های فراابتکاری به‌منظور انتخاب ویژگی استفاده کرده‌اند. در [۱۲] الگوریتم بهینه‌سازی DDAO[7] برای روش جستجو در الگوریتم Wrapper انتخاب شده است که با استفاده از تابع‌های مختلف ارزیابی شده و با انتخاب ۱۰ ویژگی از میان ۴۱ ویژگی و با دقّت ۹۴ درصد توانسته حملات را تشخیص دهد. همچنین، در [۱۳] عامل اصلی بالا بودن نرخ هشدار کاذب را وجود افزونگی در داده‌ها می‌داند و به همین علت انتخاب ویژگی را نیز یک فاز مهم در تشخیص نفوذ می‌دانند. از این رو، به دلیل بالا بودن قدرت جستجوی الگوریتم‌های فراابتکاری، از سه الگوریتم کلونی زنبور عسل(ABC)، الگوریتم کلونی مورچگان (ACO) و ازدحام ذرات (PSO) برای انتخاب ویژگی و از الگوریتم‌های ماشین بردار پشتیبان (SVM) و نزدیک‌ترین همسایه (KNN) به‌منظور رده‌بندی استفاده می‌کند.

با توجه‌به‌این‌که سرعت، صحّت و دقّت از جمله معیارهای ارزیابی سامانه‌های تشخیص نفوذ است، یکی از مهم‌ترین مشکلاتی که در این سامانه وجود دارد، نرخ هشدارهای نادرست تولیدشده است؛ بنابراین، بهبود صحّت و دقّت سامانه و کاهش نرخ هشدارهای نادرست تولیدشده، یکی از جنبه‌های مهم پژوهشی در سامانه‌های تشخیص نفوذ است. در این پژوهش نیز، ما به بررسی بهبود تشخیص نفوذ در شبکه‌های مبتنی بر دسته‌بندی و طبقه‌بندی با کاهش ویژگی و به‌کارگیری الگوریتم فراابتکاری می‌پردازیم و سامانه‌ای را ارائه می‌دهیم که میزان دقّت و صحّت بالایی را نسبت به سایر روش‌ها داشته باشد.

۳- روش پیشنهادی

ایدهٔ کلی مقاله بررسی تأثیر انتخاب ویژگی در دقّت الگوی تشخیص نفوذ به روش جنگل تصادفی (RF)[8] و الگوریتم جستجوی گرگ خاکستری بهینه‌شده است. به‌طورکلی، هدف ایدهٔ پیشنهادی کاهش ابعاد داده‌ها جهت ساخت الگوی سریع‌تر و کاراتر است. از طرفی باتوجّه‌به‌این‌که ابعاد داده‌ها کاهش یافته است، پیچیدگی الگو نیز کاهش می‌یابد.

روش پیشنهادی ما بر اساس ترکیب الگوریتم بهینه‌سازی گرگ خاکستری (GWO)[9] با الگوریتم طبقه‌بندی RF جهت انتخاب ویژگی و ساخت الگو بهتر و بهینه‌تر است. از همین رو، نخست، مفاهیم مرتبط با انتخاب ویژگی، الگوریتم بهینه‌سازی گرگ خاکستری GWO و سپس الگوریتم جنگل تصادفی ارائه خواهد شد. سپس الگوریتم GWO به‌منظور تسریع روند و افزایش دقّت بهبود داده می‌شود و با روش سنّتی مقایسه می‌شود. در ادامه روش پیشنهادی به‌تفضیل شرح داده خواهد شد و در بخش آخر نتایج حاصل از پیاده‌سازی و موضوعات پیشنهادی آورده می‌شود.

۳-۱- انتخاب ویژگی

انتخاب ویژگی در بسیاری از زمینه‌های پژوهشی و کاربردها همانند پردازش متن، به اسناد اینترنتی، تحلیل آرایه‌ای حالت ژن‌ها که مجموعه دادهٔ آن بیش از ده‌ها، صدها یا هزاران متغیّر دارد، توجّه شده است. اهداف انتخاب ویژگی شامل سه مورد زیر است[۱۴]:

- بهبود کارایی الگوریتم‌های یادگیری ماشین
- ایجاد تخمین‌زننده‌های سریع‌تر و کاراتر

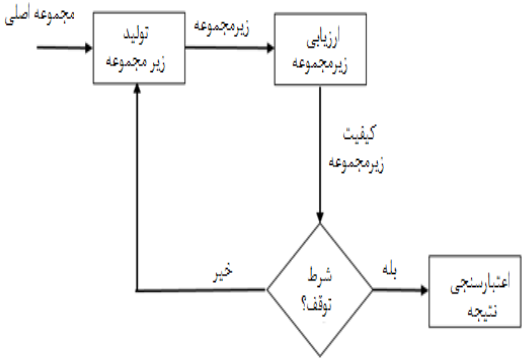
[1] Weighted majority voting
[2] Distributed Denial of Service
[3] Convolutional neural network
[4] Software defined network
[5] Taylor-elephant herd optimization
[6] Deep belief network
[7] Dynamic differential annealed optimizer
[8] Random Forest
[9] Gray Wolf Optimization

[DOI: 10.61186/jsdp.20.1.133]

- فراهم‌کردن درک بهتر از فرآیندهای صورت‌گرفته جهت تولید داده

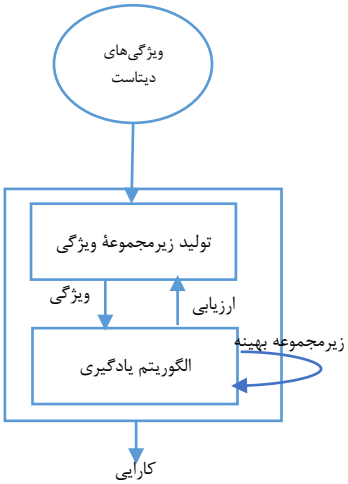
یک مسئلهٔ اساسی در یادگیری ماشین تخمین بین ورودی‌ها { x_M "," … "," x_2 "," x_1}=Xو خروجی y، بر اساس داده‌های ذخیره‌شده در حافظه (مجموعه داده)است.اغلب x_i یک بردار بوده و y_i اعداد هستند، گاهی اوقات خروجی y توسط همهٔ ویژگی‌های ورودی {x_M "," … "," x_2 "," x_1} تعریف نمی‌شود؛ به‌جای آن، دربارهٔ y توسط زیرمجموعه‌ای از ویژگی‌ها { "," x_((1)) "," x_((m)) "," …"," x_((2))} که در آن m<M است، تصمیم‌گیری می‌شود [۱۵]. روند کلی انتخاب ویژگی شامل چهار مرحله کلیدی است که در شکل (۲) نشان داده شده است و عبارت‌اند از:

- تولید زیرمجموعه‌ای از ویژگی‌ها
- ارزیابی زیرمجموعه
- بررسی شرایط توقّف
- ارزیابی نتیجه



(شکل- ۱) چهار مرحلهٔ اساسی در انتخاب ویژگی
Figure 1. four essential steps in feature selection

بهترین ویژگی‌ها از یک مجموعه با M ویژگی در زمان2^m انتخاب می‌شود که این زمان هنگامی‌که تعداد ویژگی‌ها زیاد باشد، به دلیل هزینهٔ بالای محاسباتی قبول نمی‌شود؛ بنابراین، الگوریتم‌های زیادی در این راستا پیشنهاد شدند که می‌توانند به سه دستهٔ کلی بسته‌بند (لفافه)[1] و فیلتر[2] و تعبیه‌شده[3] تقسیم‌بندی شوند و بر اساس نیاز، برای پیدا کردن بهترین زیرمجموعه ویژگی‌ها به کار می‌روند[۱۶]. در روش لفافه، برای ارزیابی زیرمجموعهٔ تولیدشده از الگوریتم یادگیری استفاده می‌شود و بسته به نتایج مراحل پیشین، ویژگی‌ها از زیرمجموعه حذف و یا به آن اضافه می‌شوند. روش فیلتر با استفاده از رتبه‌بندی و دادن امتیاز به ویژگی‌ها، تصمیم‌گیری‌های لازم را برای انتخاب زیرمجموعه ویژگی انجام می‌دهد و روش سوم هم ترکیب دو روش قبلی است. روش لفافه به‌طور معمول، کندتر از روش پالایش است، امّا کارایی بهتری دارد. در این پژوهش از روش لفافه برای انتخاب ویژگی و از الگوریتم فرابتکاری گرگ‌های خاکستری برای جستجوی فضای مسئله استفاده می‌شود. روش لفافه از تخمین‌زنندهٔ (الگوریتم یادگیری ماشین)، به‌عنوان جعبهٔ سیاه و از کارایی تخمین‌زننده به‌عنوان تابع هدف برای ارزیابی زیرمجموعهٔ ویژگی‌ها استفاده می‌کند. ازآنجاکه ارزیابی 2^M زیرمجموعه، مسئله‌ای Np-hard است، انتخاب زیرمجموعه‌ای به‌نسبت بهینه از ویژگی‌ها می‌تواند با استفاده از الگوریتم جست‌وجویی که به‌صورت مکاشفه‌ای زیرمجموعه‌ای را انتخاب می‌کند، انجام شود. روند کلی روش لفافه در شکل (۳) آمده است.



(شکل- ۲) روند کلی انتخاب ویژگی به روش لفافه
Figure 2. Wrapper method procedure

۳-۲- الگوریتم بهینه‌سازی گرگ خاکستری

یکی از الگوریتم‌های پرکاربرد فرابتکاری، الگوریتم گرگ خاکستری‌ست که از حملهٔ گرگ‌ها در زمان شکار الهام می‌گیرد. در زندگی اجتماعی گرگ‌های خاکستری، جفت آلفا به‌عنوان رهبر گروه شناخته می‌شوند و تصمیم‌گیری دربارهٔ شکار، مکان خواب، زمان بیدار شدن و غیره را بر عهده دارند. ردهٔ دوم در سلسله‌مراتب یک دسته، متعلّق به گرگ‌های بتا است. گرگ‌های بتا به آلفا در تصمیم‌گیری‌ها و سایر فعّالیّت‌های دسته کمک می‌کنند. گرگ‌های پایین‌ترین مقام، گرگ‌های امگا هستند. این گروه از گرگ‌ها نقش پیش‌مرگ را در دسته بازی می‌کنند. به

[1] Wrapper Method
[2] Filter Method
[3] Embedded approach

[DOI: 10.61186/jsdp.20.1.133]

گرگ‌هایی که در سلسله‌مراتب بالا ذکر نشده است، گرگ‌های دلتا گفته می‌شود. گرگ‌های دلتا تحت فرمان آلفا و بتا بوده، ولی نسبت به امگا برتری دارند[۱۷]. به‌منظور الگو کردن سلسله‌مراتب اجتماعی گرگ‌های خاکستری، بهترین پاسخ را به‌عنوان آلفا در نظر می‌گیریم. همچنین، دومین و سومین پاسخ مناسب را پس از آلفا، بتا و دلتا می‌نامیم. سایر پاسخ‌ها در گروه امگا قرار می‌گیرند. در این الگوریتم بهینه‌سازی، شکار توسط آلفا، بتا و دلتا هدایت شده و گرگ‌های امگا از این سه دسته پیروی می‌کنند. وقتی شکار توسط گرگ‌ها احاطه شد و از حرکت ایستاد، حمله به رهبری گرگ آلفا شروع می‌شود. این فرایند با استفاده از کاهش بردار a الگو می‌شود و برای پیاده‌سازی این الگو (سازوکار شکار)، از روابط زیر استفاده می‌کنیم:

$$\vec{D} = |\vec{C}.\vec{X_p}(t) - \vec{X}(t)| \quad (۱)$$

$$\vec{X}(t+1) = \vec{X_p}(t) - \vec{A}.\vec{D} \quad (۲)$$

که t بیانگر شمارهٔ تکرار است و ازآنجاکه A برداری تصادفی است، با کاهش a، بردار ضرایب A هم کاهش می‌یابد. اگر $|A|<1$ باشد، گرگ آلفا و سایر گرگ‌های گله به شکار نزدیک می‌شوند و اگر $|A|>1$، گرگ‌ها از شکار دور خواهند شد. در رابطهٔ ذکرشده بردار C به‌عنوان موانع موجود در طبیعت که نزدیک شدن گرگ‌ها را به شکار کند می‌کنند، در نظر گرفته می‌شود. بردار C به شکار وزن داده و آن را برای گرگ‌ها غیرقابل‌دستیابی‌تر می‌کند. این بردار برخلاف a به‌صورت خطی کاهش نمی‌یابد. X_p بردار مکان طعمه و X بردار مکان گرگ خاکستری‌ست. بردارهای A و C (بردارهای ضریب) به‌صورت زیر محاسبه می‌شوند:

$$\vec{A} = 2\vec{a}.\vec{r_1} - \vec{a} \quad (۳)$$

$$\vec{C} = 2.\vec{r_2} \quad (۴)$$

که a به‌صورت خطی و در طی تکرارها از مقدار دو به صفر کاهش می‌یابد و r_1 و r_2 بردارهای تصادفی در بازهٔ $[۰،۱]$ هستند. برای شبیه‌سازی نحوهٔ شکار گرگ‌های خاکستری، فرض می‌کنیم که آلفا، بتا و دلتا اطّلاعات بهتری نسبت به موقعیت شکار دارند. در نتیجه، ما سه پاسخ برتر کسب‌شده تا بدین جای کار را ذخیره کرده و سایر گرگ‌ها را وادار می‌کنیم تا موقعیت خود را با توجّه به این سه پاسخ برتر به‌روزرسانی کنند. روابط زیر برای این کار آورده شده‌اند:

$$\vec{D_\alpha} = \left|\vec{C_1}.\vec{X_\alpha} - \vec{X}\right|, \vec{D_\beta} = \left|\vec{C_2}.\vec{X_\beta} - \vec{X}\right|, \vec{D_\delta} = \left|\vec{C_3}.\vec{X_\delta} - \vec{X}\right| \quad (۵)$$

$$\vec{X_1} = \vec{X_\alpha} - \vec{A_1}.(\vec{D_\alpha}), \vec{X_2} = \vec{X_\beta} - \vec{A_2}.(\vec{D_\beta}), \vec{X_3} = \vec{X_\delta} - \vec{A_3}.(\vec{D_\delta})$$

$$\vec{X}(t+1) = \frac{\vec{X_1}+\vec{X_2}+\vec{X_3}}{3} \quad (۶)$$

در انتها ذکر چند نکته ضروری است[۱۸]:

- سلسله‌مراتب اجتماعی در این روش باعث می‌شود تا الگوریتم بهترین پاسخ‌های به‌دست‌آمده در طیّ چندین تکرار را ذخیره کند.
- سازوکار شکار اجازه می‌دهد تا موقعیّت احتمالی طعمه با پاسخ‌های برتر تعیین شود.
- جست‌وجو و استخراج با توجّه به a و A تضمین شده است.
- با کاهش A، نیمی از تکرارها به جست‌وجو ($|A|<1$) استخراج به دیگر نیمی و (1) اختصاص می‌یابد.
- این روش فقط دارای یک شاخص (a) برای تنظیم و مقداردهی است. در حقیقت، تعادل بین روند اکتشاف و استخراج با یک شاخص کنترل می‌شود، بنابراین، نقش به‌سزایی در عملکرد الگوریتم دارد.

۳-۳- جنگل تصادفی

جنگل تصادفی[1]، یک الگوریتم یادگیری ماشین با قابلیت استفادهٔ آسان است که بیشتر اوقات نتایج بسیار خوبی را حتی بدون تنظیم فراشاخص‌های آن، فراهم می‌کند. این الگوریتم به دلیل سادگی و قابلیت استفاده، هم برای دسته‌بندی[2] و هم رگرسیون[3] یکی از پرکاربردترین الگوریتم‌های یادگیری ماشین محسوب می‌شود. این روش شامل گروهی از درخت تصمیم است و با استفاده از روش bagging شاخص‌های لازم برای ایجاد هر درخت تنظیم می‌شود. نتایج بهینه‌سازی این روش به‌طور مستقیم به همبستگی درخت‌های تصمیم بستگی دارد؛ به‌طوری‌که با افزایش همبستگی درخت‌ها خطا کاهش می‌یابد. این روش برای نتیجهٔ نهایی از روش رأی‌دهی استفاده می‌کند[۱۹]. یکی از مهمترین نقاط ضعف الگوریتم جنگل تصادفی[4] بیش‌برازش است که سبب می شود تا الگوریتم دارای دقّت آموزش بالا اما دقّت تست کم باشد. یکی از بهترین راه‌حل‌ها برای غلبه بر این مشکل استفاده از مجموعه اعتبارسنجی[5] است که در این پژوهش از اعتبارسنجی متقابل با ده دسته[6] استفاده شده است. در ابتدا داده‌ها به ده دسته تقسیم شده، نه قسمت برای آموزش و یک

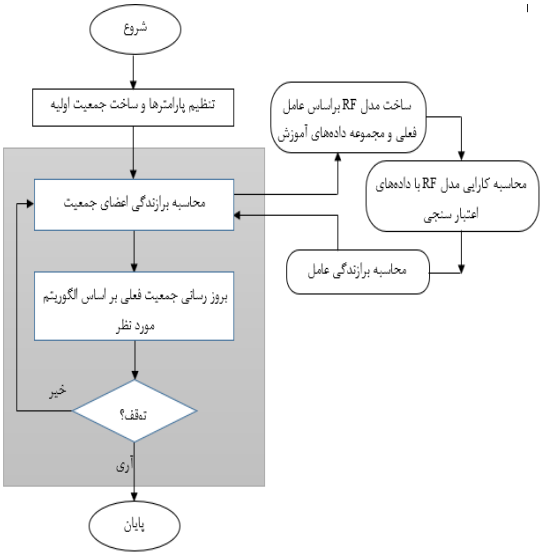
[1] Random Forest
[2] Classification
[3] Regression
[4] Overfitting
[5] Validation set
[6] 10-fold cross validation

[DOI: 10.61186/jsdp.20.1.133]

قسمت برای تست در نظر گرفته می‌شود. این کار ده بار تکرار می شود تا هر دسته یک بار به‌عنوان داده‌های تست در نظر گرفته شود.

۴-۳- الگوریتم پیشنهادی

روش پیشنهادی بر اساس انتخاب ویژگی به کمک ترکیب RF و GWO بنا نهاده شده است. این روش انتخاب ویژگی بر اساس الگوریتم‌های لفافه است. روند کلی روش پیشنهادی به‌صورت شکل (۳) نشان داده شده است:



(شکل- ۳) روند کلی روش پیشنهادی

Figure 3. proposed method flowchart

در نظر پیشنهادی، برازندگی هر گرگ را به‌صورت زیر محاسبه می‌کنیم:

در انتخاب ویژگی به‌دنبال کاهش و حذف ویژگی‌های نامربوط و تکراری و ساخت الگوی دقیق بر اساس مجموعه ویژگی‌های انتخاب شده هستیم، باتوجّه‌به‌این‌که الگوی ارائه‌شده در این پژوهش برای انتخاب ویژگی، از روش لفافه استفاده می‌کند، باید دقّت الگو در ارزیابی هر عامل استفاده شود. بنابراین، جهت محاسبهٔ برازندگی عامل‌ها در روش پیشنهادی از رابطهٔ (۸) استفاده شد.

$$Fitness(Agent_i) = \frac{Correct\ Detect\ Sample}{Total\ Sample} \qquad (۸)$$

در رابطهٔ ۸ Agent_i ام است و Correct Detect Sample، بیان‌گر تعداد نمونه‌هایی که به‌درستی تشخیص داده شدند و Total Sample تعداد کل نمونه‌هاست. همچنین، هر عامل جست‌وجو به‌صورت زیر نشان داده می‌شود. بر اساس نمایش مذکور طول هر عامل برابر تعداد ویژگی‌ها است. عامل‌های حل مسئله که در الگوریتم GWO که گرگ نام‌گذاری شدند، به‌صورت بردار شکل ۴ است.

f_1	f_2	f_3	...	f_i	...	f_N
0.5	0.2	0.3	...	0.75	...	0.6
w_1	w_2	w_3	...	w_i		w_N

شکل- ۴. اجزای گرگ خاکستری

Figure 4. Individuals in proposed GWO

با توجّه به شکل (۴) طول هر عامل برابر N و در آن N تعداد ویژگی‌های هر نمونه[1]، در مجموعه داده است. نکته‌ای که باید به آن توجّه شود، در انتخاب ویژگی هدف انتخاب (وجود-۱) و عدم انتخاب (عدم وجود-۰)، یک ویژگی است. امّا در این نمایش برای‌این‌که به یک ساختار قوی‌تر در مورد بیان اهمیّت ویژگی‌ها پرداخته شود و تنها با وجود داشتن و وجود نداشتن اهمیّت ویژگی‌ها بیان شود، از روش وزن‌دهی به ویژگی‌ها استفاده‌شده است؛ یعنی هر بخش یک عامل در قسمت انتخاب ویژگی نشان‌دهنده درجه اهمیت آن ویژگی است و هر چه این وزن $(w_i ; 1 \le i \le N)$ بیش‌تر باشد، شانس انتخاب آن ویژگی نیز بیش‌تر می‌شود. حال جهت اینکه ویژگی‌های مفید استخراج و ویژگی‌های غیر مفید حذف شوند، نمایش شکل (۶) در قسمت انتخاب ویژگی باید به‌صورت برداری دودویی (یک= انتخاب ویژگی، صفر= عدم انتخاب ویژگی) تبدیل شود. برای تبدیل قسمت انتخاب ویژگی در شکل (۵) به برداری دودویی از تابع تبدیل زیر استفاده‌شده است:

$$f(w_i) = \begin{cases} 1 & if\ w_i \ge th \\ 0 & O.w \end{cases} \qquad (۹)$$

در رابطه ۹، w_i برابر وزن ویژگی iاُم در عامل و th مقدار آستانه است. اگر وزن ویژگی در عامل بیش‌تر از مقدار آستانه باشد، آن ویژگی انتخاب می‌شود؛ درغیراین‌صورت ویژگی موردنظر انتخاب نمی‌شود. در این پژوهش مقدار آستانه برابر ۰,۵ (th=0.5) در نظر گرفته شد. بنابراین، با استفاده از رابطهٔ (۹) و مقدار th=0.5، شکل (۵) به‌صورت شکل ۵ نمایش داده می‌شود. از نمایش عامل‌ها در شکل (۵) در تابع برازندگی استفاده می‌شود.

f_1	f_2	f_3	...	f_i	...	f_N
1	0	0	...	1	...	1
w_1	w_2	w_3	...	w_i		w_N

(شکل- ۵) نمایش عامل به‌صورت دودویی جهت مشخص کردن ویژگی‌های انتخاب‌شده

Figure 5. binary representation of feature (Binary GWO)

[1] Instance

[DOI: 10.61186/jsdp.20.1.133]

یکی از مشکلات اصلی الگوریتم فراابتکاری گرگ‌های خاکستری استفاده از یک شاخص برای ایجاد توازن بین اکتشاف و استخراج است که در این پژوهش از دو روش برای بهبود این الگوریتم استفاده‌شده است. در روش سنّتی، شاخص a یک رفتار خطی دارد به‌صورتی‌که در آغاز اجرای الگوریتم میل به اکتشاف و در تکرارهای آخر میل به استخراج دارد. با تبدیل رفتار خطی این متغیر از خطی به غیرخطی می‌توان توازن قابل قبولی بین اکتشاف و استخراج ایجاد کرد[۲۰]. ازاین‌رو، فرمول زیر برای به‌روزرسانی این متغیر در هر تکرار استفاده می‌شود.

$$a = 2 - 2(\frac{iteration}{Max\_Iteration})^2 \qquad (۱۰)$$

در گام بعدی از موقعیت گرگ‌های بتا در تصمیم‌گیری حرکت به سمت هدف استفادهٔ بیشتری می‌شود. به‌این‌ترتیب که با ایده گرفتن از الگوریتم ژنتیک، گرگ‌های آلفا و بتا به‌عنوان والد در نظر گرفته می‌شوند و با استفاده از روش تقاطع[1] دو فرزند تولید می‌شود[۲۱] که پس از بررسی برازندگی آن‌ها یا به جمعیّت اضافه می‌شوند و باعث حذف گرگ‌های دلتا می‌شوند؛ یا در تصمیم‌گیری تأثیری نمی‌گذارند. این روش باعث تسریع همگرایی و جستجوی بهتر فضای جستجو می‌شود. به‌منظور ازدیاد ویژگی در روش‌های انتخاب ویژگی در این مقاله از تقاطع یکنواخت برای تولید فرزند استفاده‌شده است. به‌صورت تجربی قابل‌درک است که این روش بیشتر از تقاطع تک نقطه‌ای و چند نقطه‌ای می‌تواند تأثیرگذار باشد. در تقاطع یکنواخت هر بیت ویژگی با احتمال یکسان بین دو والد (گرگ دلتا و بتا) آمیزش می‌شود. در این مقاله برای بررسی کارایی روش پیشنهادی از معیارهای متناسب در جهت سنجش کارایی طبقه‌بندی استفاده شد. بر همین اساس، برای محاسبهٔ معیارهای ارزیابی که در ادامه بیان می‌شود، از ماتریس اغتشاش استفاده شده است. جدول (۱) ماتریس اغتشاش را نشان می‌دهد.

بر اساس ماتریس اغتشاش معیارهای زیر برای محاسبهٔ کارایی الگوهای پیشنهادی استفاده می‌شود.

$$Accuracy = (TP + TN)/N \qquad (۱۰)$$

$$Recall = TP/(TP + FN) \qquad (۱۱)$$

$$F\_score = (2 * TP)/(2 * TP + FP + FN) \qquad (۱۲)$$

$$Precision = TP/(TP + FP) \qquad (۱۳)$$

[1] Crossover

این معیارها برای سنجش کارایی الگوریتم‌های طبقه‌بندی به کار می‌روند. هر چه مقدار این معیارها بیشتر باشد، الگوریتم کارایی بهتری داشته است. در بخش بعدی نتایج حاصل از روش پیشنهادی آورده می‌شود.

(جدول- ۱): ماتریس اغتشاش

Table 1. Confusion matrix

		رده واقعی نمونه	
	کل نمونه‌ها	Positive	Negative
خروجی طبقه‌بندی‌ها (رده تخمین زده‌شده)	Positive	TP	FP
	Negative	FN	TN

۴- نتایج روش پیشنهادی

برای پیاده‌سازی روش پیشنهادی از نرم‌افزار متلب و برای آزمایش الگو، از مجموعه داده NSL-KDD استفاده شده است که با وجود قدیمی بودن این مجموعه داده، همچنان در زمینه‌های تشخیص نفوذ پژوهشگران از آن استفاده می‌کنند. [۲۲-۲۴]. برای ارزیابی روش پیشنهادی نیز از ماتریس اغتشاش استفاده شده است، که بر همین اساس ماتریس اغتشاش برای روش پیشنهادی روی داده‌های آزمون به‌صورت جدول (۲) بوده است. همان‌طورکه در جدول مشاهده می‌شود، الگوی پیشنهادی به‌خوبی توانسته است با دقّت و صحّت بالا و نرخ هشدار کاذب کم نفوذ را تشخیص دهد. در ادامه، روند پژوهش به‌منظور بهبود سرعت همگرایی و افزایش کارایی الگوی پیشنهادی، الگوریتم گرگ خاکستری بهبود داده شده که نتایج حاصل از آن در جدول (۳) نمایش داده شده است. در مقایسه با روش پیشنهادی، روش بهبودیافته عملکرد بهتری دارد.

(جدول- ۲) داده‌های آزمون ماتریس اغتشاش برای الگو پیشنهادی

Table 2. data of confusion matrix for proposed model

		برچسب واقعی	
		ردهٔ مثبت	ردهٔ منفی
پیش‌بینی‌شده	رده مثبت	۹۶۰۹(TP)	۳۴۰(FP)
	رده منفی	۱۰۱(FN)	۵۳۸۹ (TN)

(جدول-۳) نتایج روش پیشنهادی و بهبودیافته

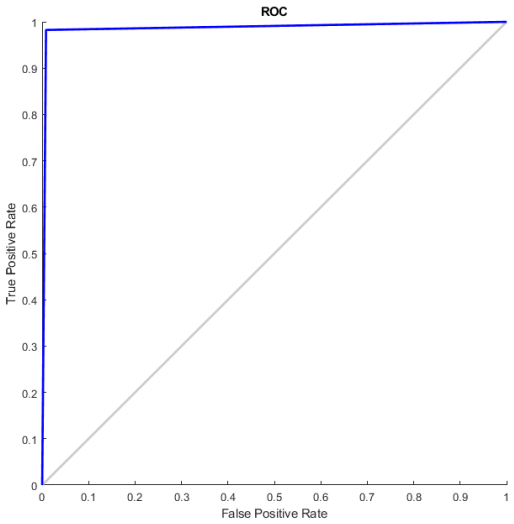
Table 3. Results of conventional GWO and Improved GWO

معیار/ روش	دقّت	صحّت	F-score	FPR	Recall
ایده پیشنهادی	۰,۹۷۱۴	۰,۹۶۵۸	۰,۹۷۷۶	۰,۰۵۹۳	**۰,۹۸۹۶**

[DOI: 10.61186/jsdp.20.1.133]

بهبودیافته	۰,۹۸۹۷	۰,۹۸۸۳	۰,۹۹۰۷	۰,۰۱۷۵	۰,۹۹۱۸

نمودار ROC روش پیشنهادی به‌صورت شکل (۶) بوده است:



(شکل-۶) نمودار ROC روش بهبودیافته

Figure 6. Improved GWO roc curve

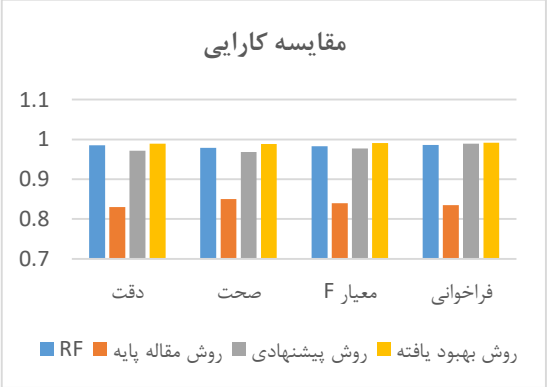
در این قسمت نتایج به‌دست‌آمده از روش پیشنهادی را با مقاله‌های مشابه ارزیابی می‌کنیم. در همین راستا بر اساس گزارشی که در مقالهٔ پایه [۲۵] آمده است، نتایج به‌صورت جدول (۴) نشان داده شده است. در مقالهٔ ذکرشده همانند این پژوهش از دیتاست NSL-KDD استفاده شده که این به‌دلیل جامعیت و یکپارچگی در بسیاری از پژوهش‌ها به آن توجه شده است. الگوهای پیاده‌سازی‌شده در مقالهٔ پایه شامل جنگل تصادفی، SVM، رگرسیون لاجستیک و بیز ساده است و نتایج به‌دست‌آمده در جدول زیر دیده می‌شود. یکی از مهمترین دلایل انتخاب این مرجع پیاده‌سازی روش‌های گوناگون و مقایسهٔ مناسب روش‌های پیاده‌سازی‌شده است. با این حال دارای نقاط ضعفی مانند دقّت پایین هستند که در این مقاله جهت بهبود نتایج الگو ترکیبی پیشنهادی ارایه شده است. نتایج نشان می‌دهد روش پیشنهادی و بهبود توانسته است نسبت به سایر پژوهش‌های بررسی‌شده کارایی بهتری داشته باشد.

با مقایسه الگوی پیشنهادی با دیگر پژوهش‌ها باصراحت می‌توان مدعی شد که الگوی پیشنهادی با ترکیب الگوریتم فراابتکاری و یادگیری ماشین کارایی بهتری داشته است. از طرفی، باید دقّت داشت نتیجه روش مقالهٔ پایه بر اساس اینکه وابسته به نتیجه سایر طبقه‌بندی‌هاست، کارایی کمتری داشته است. برای مقایسهٔ بهتر، شکل (۷) مقایسهٔ چهار روش را به‌صورت گرافیکی نمایش داده است.

(جدول- ۴) مقایسه نتایج ایدهٔ پیشنهادی

Table 4. Comparison analyses

معیار / روش	دقّت	صحّت	F-score	(TPR) Recall
جنگل تصادفی	۰,۸۸۵	۰,۸۷۹	۰,۸۸۳	۰,۸۶
SVM	۰,۷۳۶	۰,۷۴۲	۰,۷۹۶	۰,۷۸
رگرسیون لاجستیک	۰,۸۰	۰,۸۶	۰,۸۵	۰,۸۳
بیز ساده	۰,۷۸	۰,۷۳	۰,۷۵	۰,۷۷۲
روش مقالهٔ پایه	۰,۸۳	۰,۸۵	۰,۸۴	۰,۸۳۵
ایدهٔ پیشنهادی	۰,۹۷۱۴	۰,۹۶۵۸	۰,۹۷۷۶	۰,۹۸۹۶
روش بهبودیافته	۰,۹۸۹۷	۰,۹۸۸۳	۰,۹۹۰۷	۰,۹۹۱۸



(شکل- ۷) مقایسهٔ روش‌های پیشنهادی و روش ارائه‌شده در [۲۵]

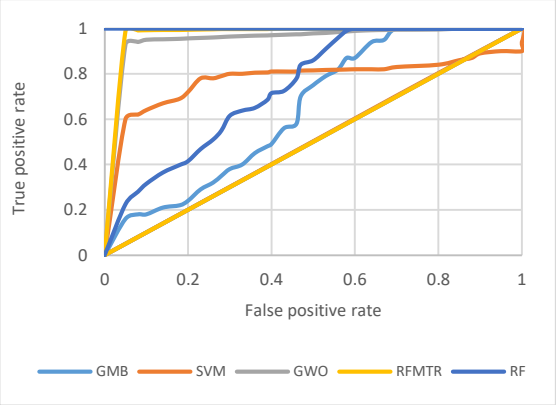
Figure 7. comparison of proposed method and methods in [25]

بر اساس موارد بیان‌شده، در شکل(۷) می‌توان گفت: الگوریتم جنگل تصادفی بدون انتخاب ویژگی در سه معیار دقّت، صحّت و معیار F نسبت‌به سایر روش‌ها کارایی بهتری داشته است. از طرفی این اختلاف نسبت‌به روش پیشنهادی ما فاحش نبوده است. همچنین، باید توجّه کرد در روش پیشنهادی ما الگو با ۱۸ ویژگی ساخته‌شده است، ولی در حالت RF الگو با همهٔ ویژگی‌های مجموعه داده (۴۱ ویژگی) ساخته شده است که این امر بیانگر کارایی خوب روش پیشنهادی است. همچنین، در شکل ۹ زیر نمودار ROC روش مقاله مقایسه‌شده، در کنار روش پیشنهادی نشان داده شده است و با مقایسهٔ آن با شکل ۶

[DOI: 10.61186/jsdp.20.1.133]

کاملاً مشهود است که الگوی بهبوددادهشده بسیار کارا و مقاوم است.

در تحلیل نمودارهای ROC باید توجّه کرد که هرچه سطح زیر نمودار ROC بیشتر باشد، کارایی الگو بهتر بوده است، با مقایسهٔ شکل (۶) با شکل (۸) می‌توان گفت کارایی الگوی بهبوددادهشده قابل‌قبول است. به‌منظور ارزیابی دقیق‌تر، در جدول (۵) روش پیشنهادی با چند پژوهش انجام‌شده در زمینهٔ تشخیص نفوذ مقایسه شده است.



(شکل- ۸) مقایسه نمودارهای ROC روش مقاله[25] و روش پیشنهادی

Figure 8. comparison of roc curve for proposed method and [25]

(جدول- ۵) ارزیابی الگوی پیشنهادی با سایر پژوهش‌ها

Table 5. Evaluation of the proposed model with other researches

معیار	روش	
دقّت (۹۷,۰۰)	ترکیب گرگ خاکستری و الگوریتم SVM	[۲۶]
دقّت (۹۴,۷۰)	Dynamic differential annealed optimizer	[۱۲]
دقّت (۹۲,۸۰)	درخت j.48 به‌عنوان بهترین الگوریتم و الگوریتم فراابتکاری ترکیبی	[۲۷]
دقّت (۹۸,۹۷)	گرگ خاکستری بهبود یافته و جنگل تصادفی	روش پیشنهادی

امروزه امنیّت شبکه از اهمیت بالایی برخوردار است. یکی از مسائل و چالش‌های مهم در امنیّت شبکه، تشخیص نفوذ است. روش‌های زیادی برای بررسی تشخیص نفوذ ارائه شده است. در این گزارش یک روش پیشنهادی و بهبود روش پیشنهادی بر اساس انتخاب ویژگی به روش لفافه جهت تشخیص نفوذ ارائه‌شده و از قدرت جستجوی الگوریتم‌های فراابتکاری جهت انتخاب زیرمجموعه ویژگی‌ها و از الگوریتم جنگل تصادفی به‌منظور رده‌بندی استفاده شده است. نتایج بیانگر این مسئله بودند که انتخاب ویژگی باعث ساخت الگوهای با کارایی قابل‌قبول در مقایسه با روش‌های معمولی هستند. همچنین، شایان‌ذکر است که در روش ارائه‌شده کمتر از ۵۰ درصد ویژگی‌ها انتخاب و بقیهٔ ویژگی‌ها حذف شدند. ازاین‌رو، برای داده‌های با ابعاد بالا، انتخاب ویژگی می‌تواند بسیار کمک‌کننده و مؤثر باشد. همچنین، با بهبود الگوریتم فراابتکاری گرگ خاکستری مدت‌زمان ساخت الگو کاهش می‌یابد و در نهایت، پیچیدگی الگو نیز کمتر می‌شود. دقّت و صحّت و نرخ هشدار کاذب به‌دست‌آمده در الگوریتم پیشنهادی بیانگر کارایی الگوی پیشنهادی است و بهبود این الگو به سرعت عملکرد و کاهش پیچیدگی کمک شایانی می‌کند. برای کارهای آینده نیز می‌توان گفت بهتر است از دیگر الگوریتم‌های فراابتکاری، روش‌های آماری یا تلفیقی در جهت کاهش ابعاد داده‌ها و از معماری‌های چندلایه به‌وسیلهٔ الگوریتم‌های فراابتکاری جهت انتخاب ویژگی‌ها استفاده شود.

۵- مراجع 5- Refrence

[1] I. Manzoor and N. Kumar, "A feature reduced intrusion detection system using ANN classifier," Expert Systems with Applications, vol. 88, pp. 249-257, 2017.

[2] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, pp. 1-22, 2019.

[3] T. A. Alamiedy, M. Anbar, Z. N. Alqattan, and Q. M. Alzubi, "Anomaly-based intrusion detection system using multi-objective grey wolf optimisation algorithm," Journal of Ambient Intelligence and Humanized Computing, pp. 1-22, 2019.

[4] E.-G. Talbi, "Machine learning into metaheuristics: A survey and taxonomy of data-driven metaheuristics," 2020.

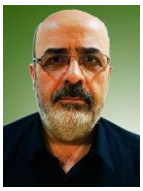
[5] D. Molina, J. Poyatos, J. Del Ser, S. García, A. Hussain, and F. Herrera, "Comprehensive Taxonomies of Nature-and Bio-inspired Optimization: Inspiration Versus Algorithmic Behavior, Critical Analysis Recommendations," Cognitive Computation, vol. 12, no. 5, pp. 897-939, 2020.

[6] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7, pp. 82512-82521, 2019.

[7] J. M. Fossaceca, T. A. Mazzuchi, and S. Sarkani, "MARK-ELM: Application of a novel Multiple Kernel Learning framework for

[DOI: 10.61186/jsdp.20.1.133]

improving the robustness of Network Intrusion Detection," Expert Systems with Applications, vol. 42, no. 8, pp. 4062-4080, 2015.
[8] K. M. Prasad, A. R. M. Reddy, and K. V. Rao, "BIFAD: Bio-inspired anomaly based HTTP-flood attack detection," Wireless Personal Communications, vol. 97, no. 1, pp. 281-308, 2017.
[9] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," Applied Soft Computing, vol. 38, pp. 360-372, 2016.
[10] D. Arivudainambi, V. K. KA, and S. S. Chakkaravarthy, "LION IDS: A meta-heuristics approach to detect DDoS attacks against Software-Defined Networks," Neural Computing and Applications, vol. 31, no. 5, pp. 1491-1501, 2019.
[11] S. Velliangiri and H. M. Pandey, "Fuzzy-Taylor-elephant herd optimization inspired Deep Belief Network for DDoS attack detection and comparison with state-of-the-arts algorithms," Future Generation Computer Systems, vol. 110, pp. 80-90, 2020.
[12] A. J. Wilson and S. Giriprasad, "A Feature Selection Algorithm for Intrusion Detection System Based On New Meta-Heuristic Optimization," Journal of Soft Computing and Engineering Applications, vol. 1, no. 1, 2020.
[13] T. Khorram and N. A. Baykan, "Feature selection in network intrusion detection using metaheuristic algorithms," International Journal of Advanced Research, Ideas and Innovations in Technology, vol. 4, no. 4, 2018.
[14] Q. Al-Tashi, S. J. Abdulkadir, H. M. Rais, S. Mirjalili, and H. Alhussian, "Approaches to multi-objective feature selection: A systematic literature review," IEEE Access, vol. 8, pp. 125076-125096, 2020.
[15] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," Neurocomputing, vol. 300, pp. 70-79, 2018.
[16] M. Di Mauro, G. Galatro, G. Fortino, and A. Liotta, "Supervised feature selection techniques in network intrusion detection: A critical review," Engineering Applications of Artificial Intelligence, vol. 101, p. 104216, 2021.
[17] R. Purushothaman, S. Rajagopalan, and G. Dhandapani, "Hybridizing Gray Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for text feature selection and clustering," Applied Soft Computing, vol. 96, p. 106651, 2020.
[18] E. Emary, H. M. Zawbaa, and C. Grosan, "Experienced gray wolf optimization through reinforcement learning and neural networks," IEEE transactions on neural networks and learning systems, vol. 29, no. 3, pp. 681-694, 2017.
[19] A. Thakkar and R. Lohiya, "Attack classification using feature selection techniques: a comparative study," Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 1, pp. 1249-1266, 2021.
[20] R. Ahmadi, G. Ekbatanifard, and P. Bayat, "A Modified Grey Wolf Optimizer Based Data Clustering Algorithm," Applied Artificial Intelligence, vol. 35, no. 1, pp. 63-79, 2021.
[21] A. N. Singh, J. Mrudula, R. Pandey, and S. Das, "A Comparative Study of Four Genetic Algorithm-Based Crossover Operators for Solving Travelling Salesman Problem," in Intelligent Algorithms for Analysis and Control of Dynamical Systems: Springer, 2021, pp. 33-40.
[22] G. S. Kushwah and V. Ranga, "Optimized extreme learning machine for detecting DDoS attacks in cloud computing," Computers & Security, p. 102260, 2021.
[23] K. Singh, L. Kaur, and R. Maini, "Comparison of Principle Component Analysis and Stacked Autoencoder on NSL-KDD Dataset," in Computational Methods and Data Engineering: Springer, 2021, pp. 223-241.
[24] S. Gavel, A. S. Raghuvanshi, and S. Tiwari, "Distributed intrusion detection scheme using dual-axis dimensionality reduction for Internet of things (IoT)," The Journal of Supercomputing, pp. 1-24, 2021.
[25] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117-123, 2016.
[26] S. Shakya, "Modified Gray Wolf Feature Selection and Machine Learning Classification for Wireless Sensor Network Intrusion Detection," 2021.
[27] O. Almomani, "A Hybrid Model Using Bio-Inspired Metaheuristic Algorithms for Network Intrusion Detection System," CMC-COMPUTERS MATERIALS & CONTINUA, vol. 68, no. 1, pp. 409-429, 2021.



شهریار محمدی مدرک دکترای خود را در رشتهٔ مهندسی رایانه از دانشگاه سالفورد انگلستان دریافت کرده است و هم‌اکنون عضو هیئت علمی دانشگاه خواجه نصیرالدین طوسی در گرایش فناوری اطّلاعات، دانشکده صنایع است. مرتبهٔ علمی ایشان دانشیاری و زمینه‌های پژوهشی ایشان نیز امنیّت در تجارت الکترونیک، رمزنگاری و رمزگشایی، پرداخت الکترونیک است

نشانی رایانامهٔ ایشان:

Email: Mohammadi@kntu.ac.ir

[DOI: 10.61186/jsdp.20.1.133]

احمد خلعتبری دانش‌آموختهٔ رشته کارشناسی ارشد دانشگاه خواجه نصیرالدین طوسی در رشتهٔ مهندسی فناوری اطّلاعات، گرایش تجارت الکترونیک است. زمینهٔ پژوهشی ایشان الگوی پذیرش تکنولوژی و امنیّت اطّلاعات است.

نشانی رایانامهٔ ایشان:

Email: Khalatbary@gmail.com

مهدی باباگلی دانشجوی دکترای دانشگاه خواجه نصیرالدین طوسی است و مدرک کارشناسی ارشد خود را از دانشگاه صنعتی ارومیه دریافت کرده است. ایشان برای مراحل تحصیلات تکمیلی خود از سهمیه استعداد درخشان استفاده کرده‌است. زمینهٔ پژوهشی ایشان نیز تشخیص حملات اینترنتی، یادگیری ماشین، الگوریتم‌های فراابتکاری و داده‌کاوی است.

نشانی رایانامهٔ ایشان:

Email: Mehdi.babagoli@email.kntu.ac.ir

[DOI: 10.61186/jsdp.20.1.133]