Journal of New Technologies in Distributed Systems and Algorithmic Computing
Islamic Azad University of Sabzevar
E-ISSN: 3115-705X
https://sanad.iau.ir/journal/ntds

Research Article

Improving the Efficiency of an Intrusion Detection System using a Decision Tree Adjusted with an Improved Random Fractal Search Combined with the Whale Optimization Algorithm

Shahrzad Rahimi [1] | Masood Niazi Torshiz [2*] | Seyyed Abed Hosseini [3]

[1]Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, Shahrzad.rahimi00@gmail.com

[2]Department of Computer Engineering, Islamic Azad University, Mashhad, Iran, Masood.niazi@iau.ac.ir

[3]Department of Electrical Engineering, Islamic Azad University, Mashhad, Iran, Sa.hosseini@iau.ac.ir

Correspondence
*Masood Niazi Torshiz, Assistant Professor, Department of Computer Engineering, Islamic Azad University, Mashhad, Iran
Masood.niazi@iau.ac.ir

Main Subjects: Improving intrusion detection system
Received: 5 October 2025
Revised: 25 October 2025
Accepted: 26 November 2025

https://doi.org/10.82195/ntds.2025.1220215

Abstract

With the increasing use of the Internet, a large amount of information is exchanged between different communication devices. Data must be securely transmitted between communication devices, and therefore, network security is one of the dominant research areas for the current network scenario. Intrusion detection systems are widely used in conjunction with other security mechanisms such as firewalls and access control. On the other hand, the evolution in attack scenarios has been such that finding efficient and optimal intrusion detection systems with frequent updates has become a big challenge. Implementation of an intrusion detection system using machine learning techniques and updated intrusion data sets is one of the effective modeling solutions for an intrusion detection system. In this article, an improved random fractal search algorithm with chaos maps is introduced. To improve the mining ability of this algorithm, the relationships of the whale algorithm have been used in combination with it. This algorithm has been used to cluster infiltrated and normal data. Then, a decision tree classifier was used to classify the infiltrated data on the NSL-KDD dataset. To evaluate the proposed method, its results have been compared with the case without clustering. In terms of the error and sensitivity criteria for the test data, the proposed method is equal to 0.2211, 0.5816, and the decision tree without clustering is equal to 0.2477, 0.5692, respectively. In terms of specificity and accuracy criteria, the proposed method has obtained better results compared to the decision tree without clustering. Therefore, the results showed that the proposed method has better efficiency and performance compared to the decision tree method without clustering.

Keywords: Intrusion Detection System, Random Fractal Search Algorithm, Chaotic Mapping, Whale Optimization Algorithm, Clustering.

فناوری های نوین در سیستم های توزیع شده و محاسبات الگوریتمی

دانشگاه آزاد اسلامی واحد سبزوار

E-ISSN: 3115-705X
https://sanad.iau.ir/journal/ntds

پژوهشی

بهبود کارایی سیستم تشخیص نفوذ با استفاده از درخت تصمیم تنظیم‌شده با جستجوی فرکتال تصادفی بهبودیافته ترکیب‌شده با الگوریتم بهینه سازی نهنگ

شهرزاد رحیمی [۱] | مسعود نیازی ترشیز *[۲] | سیدعابد حسینی [۳]

[۱] گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران،
Shahrzad.rahimi00@gmail.com

[۲] گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران،
Masood.niazi@iau.ac.ir

[۳] گروه برق، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران،
Sa.hosseini@iau.ac.ir

نویسنده مسئول
* مسعود نیازی ترشیز ، استادیار، گروه کامپیوتر، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران،
Masood.niazi@iau.ac.ir

موضوع اصلی: بهبود سیستم تشخیص نفوذ
تاریخ دریافت: ۱۳ مهر ۱۴۰۴
تاریخ بازنگری: ۳ آبان ۱۴۰۴
تاریخ پذیرش: ۵ آذر ۱۴۰۴

https://doi.org/10.82195/ntds.2025.1220215

چکیده:

با افزایش استفاده از اینترنت، حجم زیادی از اطلاعات بین دستگاه‌های ارتباطی مختلف ردوبدل می‌شود. داده‌ها باید به‌طور ایمن بین دستگاه‌های ارتباطی منتقل شوند و بنابراین، امنیت شبکه یکی از حوزه‌های تحقیقاتی غالب برای سناریوی شبکه فعلی است. سیستم‌های تشخیص نفوذ به‌طور گسترده همراه با مکانیسم‌های امنیتی دیگر مانند فایروال و کنترل دسترسی استفاده می‌شوند. از سویی دیگر تکامل در سناریوهای حمله به‌گونه‌ای بوده است که یافتن سیستم‌های تشخیص نفوذ کارآمد و بهینه با به‌روزرسانی‌های مکرر به یک چالش بزرگ تبدیل‌شده است. پیاده‌سازی سیستم تشخیص نفوذ با استفاده از روش‌های یادگیری ماشین و مجموعه داده‌های نفوذ بروز شده یکی از راه‌حل‌های مدل‌سازی مؤثر سیستم تشخیص نفوذ است. در این مقاله یک الگوریتم جستجوی فرکتال تصادفی بهبودیافته با نگاشت‌های آشوب معرفی‌شده است. برای بهبود توانایی استخراج این الگوریتم، از روابط الگوریتم نهنگ به‌طور ترکیبی با آن استفاده‌شده است. از این الگوریتم به‌منظور خوشه‌بندی داده‌های نفوذ و عادی استفاده شده است. سپس از یک طبقه‌بندی کننده درخت تصمیم برای دسته‌بندی داده‌های نفوذ بر روی مجموعه داده NSL-KDD استفاده شده است. برای ارزیابی روش پیشنهادی نتایج آن با حالت بدون خوشه‌بندی مقایسه شده است. از نظر معیار خطا و حساسیت برای داده آزمایشی روش پیشنهادی به ترتیب برابر با ۰.۲۲۱۱، ۰.۵۸۱۶ و درخت تصمیم بدون خوشه‌بندی به ترتیب برابر با ۰.۲۴۷۷و ۰.۵۶۹۲ به‌دست‌آمده است. از نظر معیارهای خصوصیت و صحت نیز روش پیشنهادی نتایج بهتری در مقایسه با درخت تصمیم بدون خوشه‌بندی به دست آورده است؛ بنابراین نتایج نشان داد که روش پیشنهادی کارایی و عملکرد بهتری در مقایسه با روش درخت تصمیم بدون خوشه‌بندی دارد.

کلیدواژه: سیستم تشخیص نفوذ، الگوریتم جستجوی فرکتال تصادفی، نگاشت آشوبی، الگوریتم بهینه‌سازی نهنگ، خوشه‌بندی

۱- مقدمه

یکی از مهم‌ترین و قوی‌ترین ابزارهای حفظ امنیت شبکه، سیستم تشخیص نفوذ و حملات احتمالی از سوی مهاجمان به شبکه است. پس از تشخیص، راهکارهای لازم جهت جلوگیری از نفوذ انجام می‌گردد تا درصد موفقیت مهاجم کاهش یابد. هدف مهاجم، دستیابی غیرمجاز به اطلاعات، خرابکاری در سیستم‌ها و انتشار ویروس و حمله به کامپیوترهای شبکه است. برای ایجاد امنیت کامل در یک سیستم کامپیوتری، علاوه بر دیواره‌های آتش و دیگر تجهیزات جلوگیری از نفوذ، سیستم‌های دیگری به نام سیستم‌های تشخیص نفوذ[1] موردنیاز هستند تا بتوانند درصورتی‌که نفوذگر از دیواره‌ی آتش، آنتی‌ویروس و دیگر تجهیزات امنیتی عبور کرد و وارد سیستم شد، آن را تشخیص داده و چاره‌ای برای مقابله با آن بیندیشند. این فنّاوری یکی از مباحث روز فناوری در عصر جدید اطلاعات و ارتباطات است که با فراگیرشدن کاربرد آن در صنایع مختلف، مسئله‌ی امنیت و حریم خصوصی آن توجه زیادی را به سمت خود جلب نموده است و به موضوعی مهم در این حوزه تبدیل شده است [۳-۱].

سامانه‌های تشخیص نفوذ به‌صورت سامانه‌های نرم‌افزاری و سخت‌افزاری ایجادشده و هرکدام مزایا و معایب خاص خود را دارند. سرعت و دقت از مزایای سیستم‌های سخت‌افزاری است و عدم شکست امنیتی آن‌ها توسط نفوذگران، قابلیت دیگر این‌گونه سیستم‌ها است. مهاجمان به‌طور مداوم، روش‌های حمله را برای دور زدن سیستم دفاعی طراحی می‌کنند. بسیاری از حملات برای به دست آوردن اعتبار کاربر که به آنها امکان دسترسی به شبکه و داده‌ها را می‌دهد، از بدافزارها یا مهندسی اجتماعی دیگر استفاده می‌کنند. یک سیستم تشخیص نفوذ شبکه برای امنیت شبکه بسیار مهم است، زیرا این امکان را فراهم می‌کند تا ترافیک مخرب را ردیابی کرده و پاسخ مناسب به آن داده شود. [۴].

هدف اصلی یک سیستم تشخیص نفوذ اطمینان از اطلاع کارکنان فناوری اطلاعات در هنگام حمله یا نفوذ شبکه است. یک سیستم شناسایی نفوذ شبکه هم ترافیک ورودی و خروجی در شبکه را کنترل می‌کند، هم داده‌های موجود بین سیستم‌های درون شبکه را مرور می‌کند. سیستم شناسایی نفوذ شبکه نظارت بر ترافیک شبکه را انجام می‌دهد و در صورت شناسایی فعالیت مشکوک یا تهدیدهای شناخته‌شده، هشدار ایجاد می‌کند، بنابراین کارکنان بخش فناوری اطلاعات می‌توانند بادقت بیشتری بررسی کرده و اقدامات مناسب را برای مسدودکردن یا متوقف کردن حمله انجام دهند [۵].

در این تحقیق مسئله‌ی تشخیص نفوذ شبکه‌های رایانه‌ای موردتوجه قرار گرفته شده است. بدین منظور ارائه‌ی یک روش مناسب برای رسیدن به این هدف موردنیاز است؛ بنابراین یک روش جدید مبتنی بر بهینه‌سازی جستجوی فرکتال تصادفی بهبودیافته با نگاشت‌های آشوب معرفی شده است که برای بهبود توانایی استخراج آن از الگوریتم نهنگ استفاده شده است. از این الگوریتم به‌منظور خوشه‌بندی داده‌های شبکه استفاده شده است که در آن هدف خوشه‌بندی به دو دسته‌ی عادی و نفوذ یافته است. سپس به کمک یک روش طبقه‌بندی مناسب و ساده به نام درخت تصمیم داده‌های مربوط به نفوذ یافته طبقه‌بندی می‌شوند.

ساختار مقاله به‌صورت زیر است. در بخش دوم مروری بر کارهای گذشته بیان شده است. در بخش سوم روش تحقیق ارائه شده است. در بخش چهارم شبیه‌سازی‌ها و نتایج ارزیابی‌ها آورده شده‌اند. درنهایت در بخش پنجم نتیجه‌گیری و پیشنهادها ارائه شده است.

۲- مروری بر کارهای گذشته

باوجود استفاده از سیستم‌های تشخیص نفوذ به‌منظور شناسایی حملات مختلف، تعداد و پیچیدگی حملات سایبری ناشناخته افزایش‌یافته است. این منجر شده است که توزیع و ناهمگنی برنامه‌ها، خدمات شبکه‌های اینترنت محلی را پیچیده و چالش‌برانگیز می‌کند [۱]. حملات سایبری مخرب مسائل جدی امنیتی را ایجاد می‌کند که نیاز به یک سیستم جدید شناسایی، نفوذپذیر و

[1] Intrusion Detection System

قابل‌اعتماد برای شناسایی نفوذ را می‌طلبد. روش سیستم¬های تشخیص نفوذ ابزاری فعال برای تشخیص نفوذ است که برای شناسایی و طبقه‌بندی به‌موقع حملات یا نقض سیاست‌های امنیتی به طور خودکار در زیرساخت‌های سطح شبکه و سطح میزبان استفاده می‌شود. بر اساس رفتارهای نفوذی، سیستم تشخیص نفوذ را می‌توان به دودسته سیستم تشخیص نفوذ مبتنی بر شبکه[1] و سیستم تشخیص نفوذ مبتنی بر میزبان[2] طبقه‌بندی نمود و معمولاً برای دست‌یابی به حداکثر کارایی و امنیت این سیستم‌هایی تشخیص نفوذ به‌صورت ترکیبی نیز مورداستفاده قرار می‌گیرند که با نام سیستم‌های تشخیص نفوذ توزیع‌شده[3] شناخته می‌شوند یک سیستم های تشخیص نفوذ که از رفتار شبکه استفاده می‌کند، سیستم تشخیص نفوذ مبتنی بر شبکه نامیده می‌شود. رفتارهای شبکه با استفاده از تجهیزات شبکه از طریق بازتاب توسط دستگاه‌های شبکه مانند سوئیچ‌ها، روترها و تَپ‌های شبکه جمع‌آوری می‌شوند و به‌منظور شناسایی حملات و تهدیدهای احتمالی پنهان در ترافیک شبکه، تجزیه‌وتحلیل می‌شوند [۱]. در تحقیقات، روش‌های مختلفی برای سیستم تشخیص نفوذ توسط محققان انجام شده است. در [۲] یک سیستم تشخیص نفوذ مبتنی بر داده مفید با استفاده از هوش مصنوعی، به‌ویژه روش‌های یادگیری ماشین توسعه‌یافته است. در این مطالعه، دو روش طبقه‌بندی مختلف برای تشخیص نفوذ، که هرکدام موارد استفاده منحصربه‌فرد خود را دارند، مقایسه شدند. مقاله [۳] یکی از کارهای اولیه در مورداستفاده از طبقه‌بندی بود. در این کار، نویسندگان از فاصله همینگ استفاده کردند تا نشان دهند تماس‌های سیستم مخرب چگونه می‌توانند با "توالی عادی موجود" متفاوت باشند. نویسندگان در [۴] یک ماشین بردار پشتیبان[4] و درخت تصمیم را پیشنهاد می‌کنند و بر اساس ادعای خود موفق‌ترین روش از نظر دقت در نظر گرفته می‌شود که معمولاً از پشتیبان‌گیری پشتیبانی می‌کند. الگوریتم‌های ماشینی بااین‌حال، این روش تعداد ایستا از خوشه‌ها را فرض می‌کند که هنوز باید تعیین شوند. در [۵]، نویسندگان یک الگوریتم مبتنی بر یک الگوریتم ژنتیک چندهدفه برای شکل‌دادن به سیستم¬های تشخیص نفوذ پیشنهاد کردند. این الگوریتم خوشه‌بندی را به‌عنوان یک تابع بهینه‌سازی تعریف می‌کند که هدف آن یافتن حداقل فاصله خوشه و حداکثر فاصله بین اهداف خوشه است. سپس یک الگوریتم ژنتیک چندهدفه را حل می‌کند. در مقاله [۶] توابع بهینه‌ساز برای تعیین بهترین خوشه‌ها تعریف می‌شود و سپس با کمک الگوریتم‌های بهینه‌سازی، تعداد خوشه‌های بهینه موردنیاز پیدا می‌گردد. نویسندگان در [۷] یک سیستم تشخیص نفوذ مبتنی بر ناهنجاری را با استفاده از روش SOM فازی پیشنهاد کردند. در [۸] یک مدل تشخیص نفوذ پیشنهاد شد که روش نیوبیز[5] و DL را با هم استفاده کرد و از الگوریتم ژنتیک نیز برای انتخاب ویژگی خوب بهره برد. در [۹] نیز یک مدل تشخیص برای ارتباطات نفوذی پیشنهاد شده است. این کار تحقیقاتی پنج روش طبقه‌بندی را آزمایش می‌کند. همچنین در [۱۰] یک رویکرد طبقه‌بندی‌کننده ترکیبی با استفاده از الگوریتم درختی برای تشخیص نفوذ شبکه ارائه گردید و مدل بر روی مجموعه داده kdd با صحت ٪۸۹/۲۴ ارزیابی شده است. مقاله [۱۱] مدلی بر اساس الگوریتم جنگل تصادفی برای انتخاب یک ویژگی مهم ارائه کرد که این مدل با استفاده از NSL-KDD مورد ارزیابی قرار گرفت. در [۱۲] یک مدل یادگیری عمیق برای تشخیص نفوذ شبکه با استفاده از الگوریتم فرا ابتکاری دوگانه بهینه سازی ازدهام ذرات توسعه داده شد.

در [۱۳] کاربرد فناوری داده‌کاوی در تشخیص نفوذ شبکه و نگهداری امنیت بررسی شده است. در این مرجع از روش خوشه‌بندی k-means بهبودیافته به‌منظور تشخیص نفوذ و نگه‌داری امنیت استفاده شد. یک الگوریتم بهینه‌سازی گرگ خاکستری لاپلاسین مبتنی بر یادگیری مقابل با ماشین یادگیری پشتیبان در سیستم تشخیص نفوذ در [۱۴] ارائه شد. کوریشف و همکاران یک دسته‌بندی کننده فازی مبتنی بر الگوریتم بهینه‌سازی نهنگ برای تشخیص نفوذ در [۱۵] ارائه نمودند. روش تشخیص خوشه-بندی ویژگی نفوذ شبکه بر اساس ماشین بردار پشتیبان و الگوریتم بلوک lCA در [۱۶] به‌کاربرده شد. در این مرجع ابتدا

[1] Network-based Intrusion Detection System
[2] Host-based Intrusion detection system
[3] Distributed-based Intrusion detection system
[4] Support Vector Machine
[5] Naive Bayes classifier

ویژگی‌های غیرمفید حذف شد تا دقت خوشه‌بندی افزایش یابد. سپس از ماشین بردار پشتیبان برای طبقه‌بندی دسته‌های مختلف استفاده شد. یک پروتکل خوشه‌بندی نامتقارن مبتنی بر کیفیت خدمات جدید با سیستم تشخیص نفوذ در شبکه‌های حسگر بیسیم در [۱۷] انجام شد. در این مقاله سیستم فازی عصبی تطبیقی برای انتخاب سر خوشه‌ها به کار برده شد. سپس از الگوریتم شکار گوزن برای یافتن سرخوشه‌های بهینه استفاده شد. در [۱۸] یک سیستم تشخیص نفوذ با روش خوشه‌بندی K-Means بهبودیافته برای شبکه‌های حسگر بیسیم مبتنی بر یادگیری متحد ارائه شد. یک سیستم تشخیص نفوذ برای اینترنت اشیا با استفاده از مدل روش‌های بهینه‌سازی گرگ خاکستری، توده ذرات و جنگل تصادفی ترکیبی در [۱۹] پیشنهاد شد. در [۲۰] از الگوریتم بهینه‌سازی ملخ برای تشخیص غیرعادی بودن شبکه برای ایجاد سیستم تشخیص نفوذ مؤثر استفاده شد. مقالات [۲۱-۲۲] هم از روش بهینه‌سازی نهنگ برای تشخیص نفوذ استفاده کرده‌اند.

با بررسی تحقیقات انجام‌شده مشخص می‌شود که آینده فناوری سیستم¬های تشخیص نفوذ به سمت یادگیری ماشین و هوش مصنوعی می‌رود و پیشرفت‌های فعلی حاصل تحقیق‌وتوسعه در این حوزه هستند. همچنین برای بهبود دقت فناوری سیستم¬های تشخیص نفوذ، کاهش خطاها، تشخیص موارد مثبت/منفی کاذب و غیره و باتوجه‌به پیچیدگی این سیستم‌ها، آن‌ها باید به‌درستی درک، توسعه و تجزیه‌وتحلیل شوند.

۳- روش تحقیق

در این بخش روش پیشنهادی تشخیص نفوذ به کمک خوشه‌بندی داده‌ها با الگوریتم فرکتال تصادفی بهبودیافته با الگوریتم نهنگ به‌منظور افزایش صحت تشخیص نفوذ معرفی می‌شود. خوشه‌بندی به‌عنوان یک رویکرد کشف الگوی بدون نظارت، فرایند تقسیم داده‌ها به گروه‌های مجزا است، به‌طوری‌که اشیای یک دسته بیشترین شباهت را به هم داشته باشند و با اشیای سایر کلاس‌ها متفاوت باشند. این مسئله که کاربردهای گسترده‌ای در حوزه‌های تحلیل پایگاه‌داده، پردازش متون و تصاویر، کاوش در شبکه‌های اجتماعی و کشف تقلب دارد، با معرفی مفهوم خوشه‌بندی K-means تاکنون محققین بسیاری را به خود جذب کرده است. توسعه و بهبود این الگوریتم در دهه‌های اخیر با ترکیب الگوریتم‌های فراابتکاری و K-means وارد فاز جدیدی شده است و باتوجه‌به مزایا و معایب این الگوریتم‌ها، تعداد زیادی از آنها تاکنون ارائه شده‌اند. در این تحقیق برای اولین‌بار الگوریتم جستجوی فرکتال اتفاقی به‌منظور بهینه‌سازی خوشه‌بندی k-means در کاربرد تشخیص نفوذ ارائه شده است. مقایسه عملکرد این الگوریتم با الگوریتم مشهور ژنتیک و بهینه‌سازی ازدحام ذرات نشان می‌دهد که الگوریتم جستجوی فرکتال موفق شده جواب بهینه سراسری را با کمترین خطا به دست آورد [۲۳].

مراحل طرح پیشنهادی به‌صورت زیر است. ابتدا پیش‌پردازش داده‌ها انجام می‌شود. داده‌هایی که به‌صورت نمادین هستند کمی‌سازی می‌شوند. سپس دو ویژگی ابتدایی در نظر گرفته‌شده و در خوشه‌بندی استفاده می‌شود. در مرحله‌ی خوشه‌بندی داده به کمک الگوریتم جستجوی فرکتال تصادفی بهبودیافته پیشنهادی خوشه‌بندی با نمونه‌های مختلف انجام می‌شود. در حالت خوشه‌بندی داده‌ها به عادی و نفوذ یافته تقسیم می‌شوند. سپس داده‌های نفوذ که خود شامل چهار دسته هستند به همراه داده‌های عادی که از قبل تشخیص داده شده‌اند به روش دسته‌بندی درخت تصمیم ارسال می‌شوند. روش درخت تصمیم بر این داده‌ها اعمال می‌شود و نتایج مربوط به ماتریس درهم‌ریختگی شامل دقت، خصوصیت، معیار F به دست می‌آیند.

در این بخش ابتدا درخت تصمیم و الگوریتم جستجوی فرکتال تصادفی معرفی می‌شود. در ادامه روش پیشنهادی استفاده از آشوب در الگوریتم فرکتال ترکیب‌شده با نهنگ به همراه فلوچارت آن پیشنهاد می‌شود. سپس استفاده از این الگوریتم در خوشه‌بندی داده‌های مربوط به تشخیص نفوذ بیان شده است.

درخت تصمیم[1] یک نمایش سلسله‌مراتبی از تمام راه‌حل‌های ممکن برای یک مسئله‌ی تصمیم‌گیری یا طبقه‌بندی است [۲۴]. فضای مسئله داده شده را به کلاس‌های از پیش تعریف‌شده به دنبال دنباله‌ای از آزمون‌ها برای هر مقدار مشخصه طبقه‌بندی می‌کند. این آزمون‌ها توسط گره‌های داخلی درخت تصمیم انجام می‌شود. شاخه‌ها نشان‌دهنده نتایج احتمالی آزمون هستند و گره‌های خارجی نشان‌دهنده کلاس‌های فضاهای پارتیشن‌بندی شده هستند. هرچند ساخت، استفاده و تفسیر درخت تصمیم آسان است؛ اما باوجود این مزایا، درختان روش ایده‌آل برای یادگیری پیش‌بینی نیستند. درخت تصمیم دقت خوبی را با مجموعه‌داده شناخته شده ارائه می‌دهد. اما مدل ممکن است مجموعه‌داده ناشناخته را به طور دقیق طبقه‌بندی نکند. این به دلیل بیش برازش[2] است؛ بنابراین نرخ خطا در مقایسه با برخی از الگوریتم‌های دیگر بیشتر است.

۳-۱- الگوریتم جستجوی فرکتال تصادفی

الگوریتم جستجوی فرکتال از سه قانون زیر استفاده می‌کند:

۱- هر ذره یک انرژی پتانسیل الکتریکی دارد.

۲- هر ذره منتشر و باعث تولید ذرات تصادفی دیگر می‌شود و انرژی هر ذره بین ذرات تولیدشده، تقسیم می‌شود.

۳ - تنها کمی از بهترین ذرات در هر نسل بعدی باقی می‌مانند و باقی ذرات نادیده گرفته می‌شود.

مهم‌ترین چالش جستجوی فرکتال، عدم تبادل داده بین افراد است؛ لذا در سال ۲۰۱۹ این الگوریتم توسط سلیمی [۲۳] بهبود یافت. الگوریتم بهبودیافته را الگوریتم جستجوی فرکتال اتفاقی[3] می‌نامند.

دو فرایند اصلی در فرکتال اتفاقی وجود دارد:

- فرایند انتشار: هر ذره در اطراف موقعیت جاری خود برای ارضای خاصیت استخراج[4] انتشار می‌یابد. این فرایند شانس یافتن مینیمم سراسری را افزایش می‌دهد و همچنین از به دام افتادن در بهینه محلی جلوگیری می‌کند.
- فرایند به‌روزآوری: الگوریتم چگونگی به‌روزشدن یک نقطه در گروه را مبتنی بر موقعیت دیگر نقاط در گروه، شبیه‌سازی می‌کند.

۳-۲- بهبود الگوریتم فرکتال تصادفی به کمک آشوب

به‌منظور بهبود در الگوریتم جستجوی فرکتال تصادفی یک راه استفاده از نگاشت‌های [5] آشوبی به‌جای جمعیت اولیه تصادفی و یا جمعیت اولیه از طریق روش یادگیری مبتنی بر مقابل است. نگاشت‌های آشوبی خاصیت شبه‌تصادفی دارند. به علت خاصیت غیرتکراری بودن تئوری آشوب، می‌توان الگوریتم‌های بهینه‌سازی را با آشوب طوری ترکیب نمود تا تنوع جمعیت افزایش یابد. این به جستجوی سراسری با سرعت بالاتر در مقابل با جستجوهای تصادفی که وابسته به احتمالات است، کمک می‌کند. ازآنجایی‌که تنوع جمعیت افزایش می‌یابد، قابلیت همگرایی الگوریتم بیشتر می‌شود. همچنین از افتادن در بهینه محلی می‌تواند جلوگیری کند [۲۵].

الگوهای آشوبی تک‌بعدی سیستم‌های ساده‌ای هستند که قابلیت ایجاد حرکت‌های آشوبی را دارند و برای این کار مناسب‌اند. به‌عنوان نمونه الگوی آشوبی لجستیک به‌صورت معادله (۱) است [۲۵]:

$$x_{k+1} = a x_k (1 - x_k) \quad (۱)$$

[1] Decision Tree
[2] Overfitting
[3] Stochastic Fractal Search - SFS
[4] Exploitation
[5] Maps

دراین‌رابطه x_k، k امین عدد آشوبناک و k نشانه عدد تکرار است. که $x \in (0,1)$ تحت این شرایط اولیه که $x_0 \in (0,1)x$ و $x_0 \notin \{0.0, 0.25, 0.75, 0.5, 1.0\}$. بر اساس تجربه $a=4$ در نظر گرفته می‌شود.

در قسمت تولید جمعیت اولیه از الگوی آشوبی استفاده می‌کنیم. همچنین در قسمت مربوط به انتشار نیز از نگاشت آشوبی استفاده می‌کنیم. در مرحله به‌روزآوری نیز از نگاشت آشوبی استفاده می‌کنیم. انتظار می‌رود که تنوع جمعیت افزایش یابد و جستجوی سراسری الگوریتم بهبود یابد.

۳-۳- بهبود الگوریتم جستجوی فرکتال تصادفی آشوبی با الگوریتم نهنگ

برای بهبود عملکرد الگوریتم از نظر توانایی استخراج آن، از فاز استخراج در الگوریتم نهنگ استفاده می‌کنیم. توضیحات آن در ادامه آورده شده است.

جهت مدل‌سازی ریاضی رفتار حباب - تور نهنگ‌های گوژپشت، دو روش به‌صورت زیر طراحی‌شده است:

۱. مکانیسم محاصره انقباضی: این رفتار به‌وسیله‌ی افزایش مقدار **a** در معادله‌ی ۳ حاصل می‌شود. به یاد داشته باشید که محدوده‌ی نوسان A به‌وسیله a کاهش می‌یابد. به‌عبارت‌دیگر، A مقداری تصادفی در فاصله‌ی a تا a−است و **a** در طی تکرارها، از مقدار ۲ تا ۰ کاهش می‌یابد. با انتخاب مقادیر تصادفی A در فاصله‌ی ۱ تا ۱- ، می‌توان مکان جدید عامل جستجو را در هر کجای بین مکان اصلی عامل و مکان بهترین عامل کنونی، تعریف کرد.

۲. مکان در حال به‌روزرسانی مارپیچی: این روش در ابتدا فاصله بین نهنگ قرار گرفته در مختصات X,Y و طعمه موجود در X*,Y* را محاسبه می‌کند. معادله‌ای مارپیچی بین موقعیت نهنگ و طعمه ایجاد می‌شود تا حرکت حلزونی شکل نهنگ گوژپشت را تقلید کند:

$$X(t+1) = D.e^{bl}.\cos(2\pi l) + X^*(t) \quad (۲)$$

در این معادله

$$D = |X^*(t) - X(t)| \quad (۳)$$

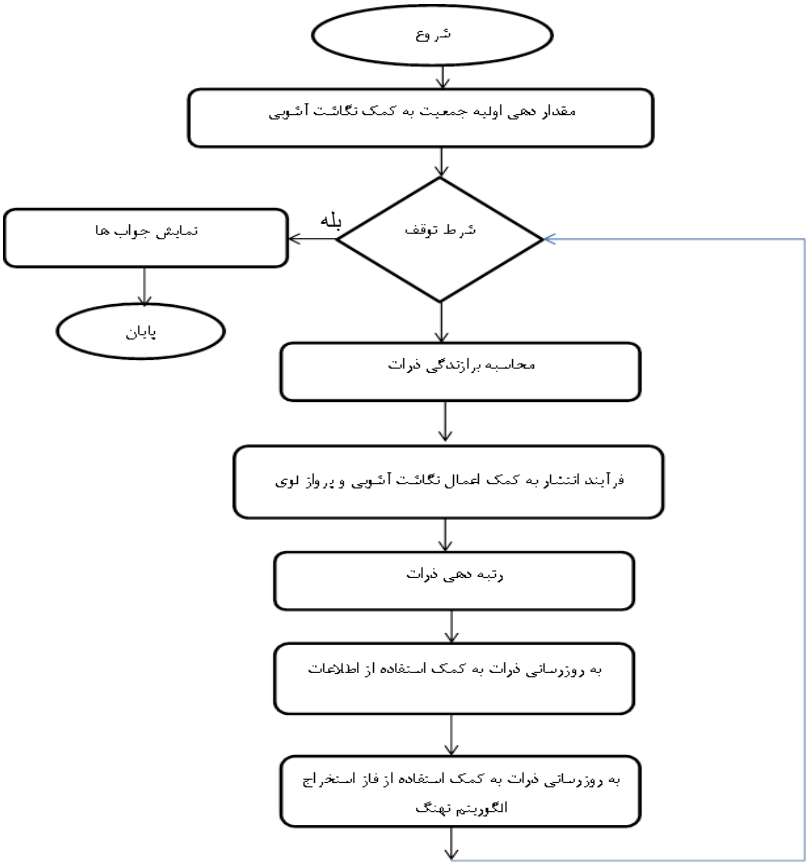
و به فاصله‌ی **i** امین نهنگ تا طعمه (بهترین راه‌حل به‌دست‌آمده تا اینجا) اشاره دارد، b ثابتی برای تعریف شکل مارپیچ لگاریتمی است، **l** عددی تصادفی بین ۱ تا ۱- و ۰ ضرب نقطه‌ای (المان در المان) است.

به یاد داشته باشید که نهنگ گوژپشت، حول طعمه در امتداد یک دایره‌ی انقباضی و هم‌زمان در مسیر مارپیچی شکلی به شنا درمی‌آید. جهت مدل‌سازی این رفتار هم‌زمان، فرض کرده‌ایم که نهنگ با احتمال ۵۰ درصد از بین مکانیسم محاصره‌ی انقباضی و یا مدل مارپیچی یکی را انتخاب می‌کند تا موقعیت نهنگ‌ها در طول بهینه‌سازی به‌روزرسانی شود. مدل ریاضی بدین صورت است:

$$X(t+1) = \begin{cases} X^*(t) - A.D, & p < 0.5 \\ D.e^{bl}.\cos(2\pi l) + X^*(t), & p \geq 0.5 \end{cases} \quad (۴)$$

که در آن p عددی تصادفی بین ۰ تا ۱ است.

از این فاز استخراج پس از اتمام مراحل الگوریتم جستجوی فرکتال تصادفی استفاده می‌کنیم. با اعمال این فاز، توانایی استخراج الگوریتم اصلی جستجوی فرکتال تصادفی بهبود خواهد یافت.

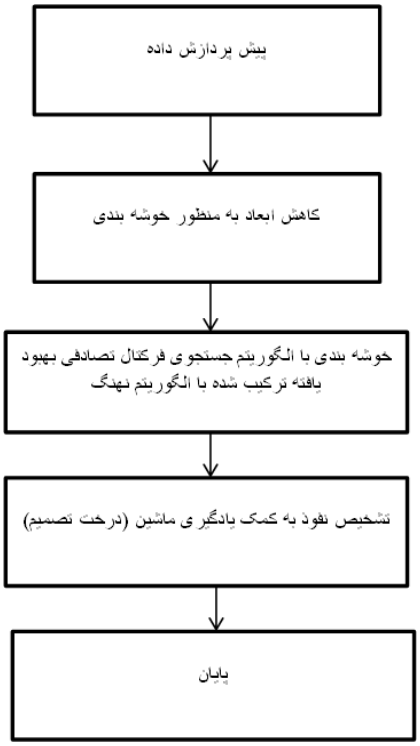


شکل ۱: الگوریتم فرکتال تصادفی بهبودیافته
Figure 1: Improved Random Fractal Algorithm

در این مقاله معیارهای مختلفی برای ارزیابی به کار برده شده است که مهم‌ترین آن‌ها در بخش بعدی توضیح داده شده‌اند.

۳-۴-روش جستجوی فرکتال تصادفی ترکیب‌شده با الگوریتم نهنگ و درخت تصمیم

طبقه‌بندی پدیده‌ها یا متغیرها از ارکان هر علمی است و تحلیل خوشه‌ای یکی از روش‌های تحلیل چند متغیره است که برای طبقه‌بندی عناصر یا متغیرها و تشخیص گروه‌های همگن به کار می‌رود. تحلیل خوشه‌ای طبقه‌بندی عناصر یا متغیرها به گروه‌های همگن است به‌گونه‌ای که عناصر (یا متغیرهای) هر گروه دارای بیشترین شباهت باهم و کمترین شباهت با عناصر (یا متغیرهای) گروه‌های دیگر باشند. خوشه‌بندی به‌عنوان یک رویکرد کشف الگوی بدون نظارت، فرآیند تقسیم داده‌ها به گروه‌های مجزا است، به‌طوری‌که اشیای یک کلاس بیشترین شباهت را به هم داشته باشند و با اشیای سایر کلاس‌ها متفاوت باشند. این مسئله که کاربردهای گسترده‌ای در حوزه‌های تحلیل پایگاه داده، پردازش متون و تصاویر، کاوش در شبکه‌های اجتماعی و کشف تقلب دارد. در این مقاله با توجه به ویژگی‌های روش خوشه‌بندی k-means از این روش به‌منظور خوشه‌بندی داده‌های عادی و نفوذ یافته استفاده می‌شود. فلوچارت مربوط به الگوریتم تشخیص نفوذ مبتنی بر خوشه‌بندی با الگوریتم جستجوی فرکتال تصادفی بهبودیافته جهت تشخیص حملات شبکه در شکل ۲ ارائه شده است.



شکل ۲: فلوچارت سیستم تشخیص نفوذ پیشنهادی
Figure 2: Flowchart of the Proposed Intrusion Detection System

باتوجه‌به شکل ۲ در مرحله اول پیش‌پردازش داده انجام می‌شود. در این مرحله داده‌ها باید طوری آماده شوند که در کاربرد خوشه‌بندی و طبقه‌بندی بتوان استفاده نمود. در مرحله پیش‌پردازش مراحل زیر انجام می‌شود:

- نوع پروتکل داده‌ها شامل TCP, UDP, ICMP است که به ترتیب برابر با ۱، ۲ و ۳ تعیین می‌کنیم.
- پایگاه‌داده شامل پنج دسته است که این دسته‌ها شامل Prob, R2L, U2R, DOS, Normal هستند.

در مرحله خوشه‌بندی، کل داده‌ها به دودسته حمله (نفوذ) و نرمال تقسیم‌بندی می‌شوند. این عمل به کمک الگوریتم پیشنهادی جستجوی فرکتال تصادفی بهبودیافته ترکیب‌شده با الگوریتم نهنگ انجام می‌شود.

پس ازآنکه مرحله‌ی خوشه‌بندی به اتمام رسید، داده‌ها به عادی و حمله تقسیم‌بندی می‌شوند. این مرحله طوری انجام می‌شود که به طور کامل داده‌های عادی تشخیص داده می‌شوند. اما برای دسته‌بندی داده‌های حملات باید از طبقه‌بندی‌کننده استفاده کرد. در این تحقیق از طبقه‌بندی‌کننده‌ی درخت تصمیم به دلیل ساختار ساده و دقت مناسب آن استفاده می‌شود. در مرحله‌ی طبقه‌بندی داده‌ها به پنج دسته طبقه‌بندی می‌شوند. پیش‌ازاین مرحله داده‌های دسته پنجم (عادی) تشخیص‌داده‌شده بودند که در شبیه‌سازی و ارزیابی عملکرد طبقه‌بندی‌کننده لحاظ می‌شوند. داده کل شامل داده‌های آموزشی و آزمایشی است، دو مثال مختلف برای ارزیابی روش پیشنهادی با تعداد نمونه‌های متفاوت در نظر گرفته شده‌اند. برای نشان‌دادن عملکرد روش پیشنهادی نتایج آن باحالت بدون خوشه‌بندی مقایسه شده است.

۴- ارزیابی روش پیشنهادی

به‌منظور ارزیابی عملکرد الگوریتم پیشنهادی از مجموعه‌داده‌ی NSL-KDD استفاده می‌شود که در آن داده‌ها به پنج دسته‌ی نرمال و نفوذهای نوع Prob, R2L, U2R, DOS تقسیم شده‌اند. این مجموعه‌داده شامل ۴۱ ویژگی است. این مجموعه‌داده متشکل از ۱۲۵۹۷۳ نمونه است.

۴-۱- انواع معیارهای ارزیابی

در این بخش معیارهای ارزیابی برای دسته‌بندی بررسی می‌شود. برای ارزیابی باید برچسبی که دسته‌بند به حمله نسبت داده با برچسب مرجع مقایسه شود. برای این کار از ماتریس سردرگمی استفاده می‌شود [۲۶].

معیارهای عملکرد سیستم‌های تشخیص نفوذ ماتریس سردرگمی برای به‌تصویرکشیدن دسته‌های واقعی و پیش‌بینی‌شده در حملات امنیت سایبری استفاده می‌شود. ماتریس سردرگمی با عبارات زیر نشان داده می‌شود.

مثبت واقعی (TP) : یک نمونه حمله به‌درستی به‌عنوان یک حمله پیش‌بینی می‌شود.

منفی واقعی (TN) : یک نمونه عادی به‌درستی به‌عنوان یک نمونه غیر حمله یا عادی پیش‌بینی‌شده است.

مثبت کاذب (FP) : یک نمونه عادی به‌اشتباه به‌عنوان حمله پیش‌بینی می‌شود.

منفی کاذب (FN) : یک حمله واقعی به‌اشتباه به‌عنوان غیر حمله یا نمونه عادی پیش‌بینی می‌شود.

موارد مثبت کاذب در مواردی که یک فعالیت عادی شبکه به‌عنوان حمله طبقه‌بندی می‌شود، می‌تواند زمان ارزشمند مدیران امنیتی را تلف کند. منفی‌های کاذب بدترین تأثیر را بر سازمان‌ها دارند؛ زیرا حمله به‌هیچ‌وجه شناسایی نمی‌شود.

باتوجه‌به پارامترهای مطرح‌شده معیارهای ارزیابی مختلفی ارائه‌شده است که ازجمله مهم‌ترین آن‌ها می‌توان به معیارهای خصوصیت[۱]، حساسیت[۲] و صحت[۳] اشاره کرد.

معیار صحت: مهم‌ترین معیار برای تعیین کارایی یک الگوریتم دسته‌بندی است و صحت کل یک دسته‌بندی را محاسبه می‌کند. این معیار نشان‌دهنده این موضوع است که چند درصد از کل مجموعه‌داده به‌درستی دسته‌بندی‌شده است.

$$\text{Accuracy} = \frac{T_p + T_N}{T_p + F_p + T_N + F_N} * 100 \quad (۵)$$

معیار حساسیت: تعیین می‌کند که نمونه‌های حمله تا چه میزان در دسته مثبت قرار داده شده‌اند، یا به‌عبارت‌دیگر چه تعدادی از حملات به‌درستی تشخیص‌داده‌شده‌اند.

$$\text{Sensitivity} = \frac{T_p}{T_p + F_N} * 100 \quad (۶)$$

معیار خصوصیت: تعیین می‌کند که چه تعدادی از نمونه‌های نرمال به‌درستی تشخیص‌داده‌شده‌اند.

$$\text{Specificity} = \frac{T_p}{F_p + T_N} * 100 \quad (۷)$$

معیار دقت: یک متریک عمومی جهت اندازه‌گیری مفیدبودن الگوریتم پیشنهادی است که به‌صورت رابطه (۴) محاسبه می‌شود.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (۸)$$

معیار فراخوانی: یکی دیگر از متریک‌های ارزیابی در سیستم‌های تشخیص نفوذ است که به‌صورت رابطه (۲-۵) قابل‌محاسبه است.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (۹)$$

معیار F: از ترکیب دو پارامتر دقت و فراخوانی محاسبه می‌شود.

$$\text{F} - \text{score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision} + \text{Recall}} \quad (۱۰)$$

۴-۲- نتایج شبیه‌سازی

[1] Specificity
[2] Sensitivity
[3] Accuracy

الف- حالت اول تعداد ۲۰ هزار نمونه

ابتدا ۲۰ هزار نمونه برای داده آموزش و ۲۰ هزار نمونه برای داده آزمایش در نظر می‌گیریم. الگوریتم جستجوی فرکتال تصادفی بهبودیافته ترکیب‌شده با الگوریتم نهنگ پیشنهادی را اجرا می‌کنیم. پس از اجرای الگوریتم نتیجه‌ی به‌دست‌آمده به این صورت است که تمام ۱۰۶۹۳ نمونه مربوط به داده‌های عادی به یک خوشه و مابقی به خوشه دیگر تعلق می‌یابند. در مرحله‌ی طبقه‌بندی ازآنجایی‌که خوشه عادی درست تشخیص‌داده‌شده است، داده‌های مربوط به کلاس آن تخصیص می‌یابند، اما طبقه‌بندی برای دسته‌بندی داده‌های حملات انجام می‌شود. در ادامه نتایج مربوط به داده آموزشی در حالت ۲۰ هزار نمونه به کمک روش پیشنهادی آورده شده‌اند.

در جدول ۱ نتایج مربوط به پیش‌بینی تعداد داده‌ها در هر دسته به کمک روش پیشنهادی نشان داده شده است.

جدول ۱: نتیجه داده آموزشی با روش پیشنهادی

Table 1: Result of the Training Data Using the Proposed Method

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۷۳۳۹	۰	۰	۵	۲
دسته واقعی ۲	۰	۱۰	۰	۱	۰
دسته واقعی ۳	۰	۵	۱۴۸	۰	۶
دسته واقعی ۴	۱	۰	۰	۱۷۸۷	۳
دسته واقعی ۵	۰	۰	۰	۰	۱۰۶۹۳

باتوجه‌به جدول ۱ مشاهده می‌شود که تعداد ۷۳۳۹ نمونه به دسته ۱ اختصاص‌یافته‌اند که درست است، اما به تعداد ۵ نمونه و ۲ نمونه به ترتیب به دسته‌های ۴ و ۵ اختصاص‌یافته‌اند که نادرست است. از سطر سوم این جدول مشاهده می‌شود که تعداد ۱۰ نمونه به دسته دوم اختصاص‌یافته‌اند که درست بوده ولی به تعداد یک نمونه به دسته ۴ تخصیص‌یافته که نادرست است. از سطر چهار این جدول مشاهده می‌شود که به تعداد ۱۴۸ نمونه به دسته سوم به‌طور درست تخصیص‌یافته‌اند. اما ۵ نمونه و ۶ نمونه به ترتیب به دسته‌های نادرست ۲ و ۵ اختصاص‌یافته‌اند. از سطر پنجم مشاهده می‌شود که به تعداد ۱۷۸۷ نمونه به دسته چهارم تخصیص‌یافته‌اند که درست بوده اما به تعداد ۱ و ۳ نمونه به دسته‌های ۱ و ۵ تخصیص‌یافته‌اند که نادرست است. از سطر آخر این جدول مشاهده می‌شود که به تعداد ۱۰۶۹۳ نمونه به دسته ۵ اختصاص‌یافته‌اند که درست است. هیچ نمونه‌ای به دسته‌های نادرست در این سطر اختصاص نیافته است. در جدول ۲ نتیجه‌ی مربوط به ماتریس درهم‌ریختگی چند دسته‌ای به کمک روش پیشنهادی نشان داده شده است.

جدول ۲: نتایج ماتریس درهم‌ریختگی مربوط به معیارهای TP,FP,FN,TN داده آموزش روش پیشنهادی

Table 2: Confusion Matrix Results for the TP, FP, FN, TN Metrics of the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۷۳۳۹	۱	۷	۱۲۶۵۳
دسته واقعی ۲	۱۰	۵	۱	۱۹۹۸۴
دسته واقعی ۳	۱۴۸	۰	۱۱	۱۹۸۴۱
دسته واقعی ۴	۱۷۸۷	۶	۴	۱۸۲۰۳
دسته واقعی ۵	۱۰۶۹۳	۱۱	۰	۹۲۹۶

در جدول ۲ مشاهده می‌شود که معیار مثبت درست برای دسته‌های ۱ الی ۵ به ترتیب برابر با ۷۳۳۹، ۱۰، ۱۴ ، ۱۷۸۷ ، ۱۰۶۹۳ شده‌اند. تعداد مثبت‌های نادرست برای دسته‌های ۱ الی ۵ به ترتیب برابر با ۱،۵، ۰ ، ۶ و ۱۱ شده‌اند.

در جدول ۳ نتایج مقایسه‌ای مربوط به معیارهای دقت، خطا، حساسیت، خصوصیت، صحت، امتیاز F1، ضریب همبستگی میتوز آورده شده‌اند.

جدول ۳: نتایج مقایسه میان روش پیشنهادی و طبقه‌بندی کننده درخت تصمیم بدون گام خوشه‌بندی

Table 3: Comparison Results Between the Proposed Method and the Decision Tree Classifier Without the Clustering Step

معیارها	روش پیشنهادی		درخت تصمیم بدون خوشه‌بندی	
نتایج	داده آموزش	داده آزمایش	داده آموزش	داده آزمایش
دقت	0.9989	0.7789	0.9982	0.7523
خطا	0.0011	0.2211	0.0018	0.2477
حساسیت	0.9673	0.5816	0.9671	0.5692
خصوصیت	0.9996	0.9248	0.9995	0.9186
صحت	0.9324	0.8616	0.9121	0.7676
نرخ مثبت نادرست	3.6812e-04	0.0752	5.1588e-04	0.0814
امتیاز F1	0.9459	0.6082	0.9326	0.5773
ضریب همبستگی متیوز	0.9476	0.5942	0.9335	0.5442

در جدول ۳ مشاهده می‌شود که روش پیشنهادی برای معیار دقت در داده آموزشی مقدار ۰/۹۹۸۹۸ بدست آورده است در حالیکه درخت تصمیم بدون خوشه بندی برابر با ۰/۹۹۸۲ بدست آورده است. معیار دقت بدست آمده برای داده آزمایشی به کمک روش پیشنهادی برابر با ۰/۷۷۸۹ و روش درخت تصمیم بدون خوشه بندی برابر با ۰/۷۵۲۳ شده است. معیار حساسیت برای داده آموزشی بدست آمده از روش پیشنهادی و درخت تصمیم بدون خوشه بندی به ترتیب برابر با ۰/۹۶۷۳ و ۰/۹۶۷۱ شده است. معیار حساسیت برای داده آزمایشی بدست آمده از روش پیشنهادی برابر با ۰/۵۸۱۶ و روش درخت تصمیم بدون خوشه‌بندی برابر با ۰/۵۶۹۲ شده است

نتیجه‌ی مربوط به معیار صحت برای داده آموزشی و آزمایشی از روش پیشنهادی به ترتیب برابر با ۰/۹۳۲۴ و ۰/۸۶۱۶ شده است. درحالی‌که این نتیجه برای روش درخت تصمیم بدون خوشه‌بندی به ترتیب برابر با ۰/۹۱۲۱ و ۰/۷۶۷۶ شده است. در مجموع از جدول ۳ نتیجه می‌شود که روش پیشنهادی برای تمام معیارها موفق‌تر عمل نموده است.

در جدول ۴ نتایج مربوط به پیش‌بینی دسته‌های ۱ الی ۵ برای داده آزمایشی به کمک روش پیشنهادی نشان داده شده است.

جدول ۴: نتایج داده تست روش پیشنهادی

Table 4: Test Data Results of the Proposed Method

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۵۴۷۵	۰	۰	۶۲	۱۰۸۴
دسته واقعی ۲	۰	۲۴	۰	۰	۳۵
دسته واقعی ۳	۲۰۱	۵	۸۳	۱۶۴	۲۰۹۴
دسته واقعی ۴	۱۴۳	۰	۱	۱۳۹۲	۶۳۳
دسته واقعی ۵	۰	۰	۰	۰	۸۶۰۴

در جدول ۴ مشاهده می‌شود که در سطر دوم پیش بینی برای دسته اول برابر با ۵۴۷۵ نمونه شده است. اما به تعداد ۶۲ و ۱۰۸۴ نمونه به ترتیب برای دسته ۴ و ۵ پیش‌بینی شده‌اند که نادرست است. در سطر سوم جدول مشاهده می‌شود که به تعداد ۲۴ نمونه برای دسته دوم پیش‌بینی شده اند. اما به تعداد ۳۵ نمونه به اشتباه به دسته ۵ اختصاص یافته‌اند. از سطر چهارم مشاهده می‌شود که به تعداد ۲۰۱ نمونه به دسته اول، ۵ نمونه به دسته دوم، ۸۳ نمونه به دسته سوم، ۱۶۴ نمونه به دسته چهارم تخصیص یافته‌اند که نادرست است. در سطر آخر به تعداد ۸۶۰۴ نمونه به دسته پنجم اختصاص یافته است که کاملا درست تشخیص داده شده‌اند.

در جدول ۵ نتایج ماتریس در هم ریختگی چند دسته‌ای داده آزمون بدست آمده از روش پیشنهادی نشان داده‌اند.

جدول ۵: نتایج ماتریس در هم ریختگی چند دسته‌ای داده آزمون روش پیشنهادی

Table 5: Multi-class Confusion Matrix Results for the Test Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۴۷۵	۳۴۴	۱۱۴۶	۱۳۰۳۵
دسته واقعی ۲	۲۴	۵	۳۵	۱۹۹۳۶
دسته واقعی ۳	۸۳	۱	۲۴۶۴	۱۷۴۵۲
دسته واقعی ۴	۱۳۹۲	۲۲۶	۷۷۷	۱۷۶۰۵
دسته واقعی ۵	۸۶۰۵	۳۸۴۶	۰	۷۵۵۰

از جدول ۵ مشاهده می‌شود که برای دسته واقعی ۱ به تعداد ۵۴۷۵ مثبت درست، ۳۴۴ مثبت نادرست، ۱۱۴۶ منفی نادرست، ۱۳۰۳۵ منفی درست تشخیص داده شده‌اند.

در ادامه نتایج مربوط به روش درخت تصمیم بدون خوشه‌بندی در جداول ۶ الی ۸ آورده شده است.

جدول ۶: نتیجه داده آموزشی با روش درخت تصمیم بدون خوشه بندی

Table 6: Training Data Results Using the Decision Tree Method Without Clustering

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۷۳۳۹	۰	۰	۵	۲
دسته واقعی ۲	۰	۱۰	۰	۱	۰
دسته واقعی ۳	۰	۵	۱۴۸	۰	۶
دسته واقعی ۴	۱	۰	۰	۱۷۸۷	۳
دسته واقعی ۵	۲	۲	۳	۶	۱۰۶۸۰

در جدول ۶ مشاهده می‌شود که به تعداد ۷۳۳۹ نمونه مربوط به دسته ۱ پیش‌بینی شده است که درست است. اما برای دسته‌های دوم و سوم به تعداد صفر نمونه پیش‌بینی شده است. اما به تعداد ۵ و ۲ نمونه برای دسته‌های ۴ و ۵ پیش‌بینی شده‌اند. در جدول ۷ نتایج معیارهای مربوط به ماتریس درهم‌ریختگی نشان داده شده است.

جدول ۷: نتایج معیارهای TP,FP,FN,TN داده آموزش روش درخت تصمیم بدون خوشه بندی

Table 7: TP, FP, FN, TN Metrics Results for the Training Data Using the Decision Tree Method Without Clustering

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۷۳۳۹	۳	۷	۱۲۶۵۱
دسته واقعی ۲	۱۰	۷	۱	۱۹۹۸۲
دسته واقعی ۳	۱۴۸	۳	۱۱	۱۹۸۳۸
دسته واقعی ۴	۱۷۸۷	۱۲	۴	۱۸۱۹۷
دسته واقعی ۵	۱۰۶۸۰	۱۱	۱۳	۹۲۹۶

در جدول ۷ مشاهده می‌شود که معیار مثبت درست بدست آمده از دسته اول برابر با ۷۳۳۹، مثبت نادرست برابر با ۳، منفی نادرست برابر با ۷ و منفی درست برابر با ۱۲۶۵۱ شده است. معیار مثبت درست، مثبت نادرست ، منفی نادرست و منفی درست برای دسته دوم به ترتیب برابر با ۱۰، ۷، ۱ و ۱۹۹۸۲ شده است. این معیارها برای دسته سوم به ترتیب برابر با ۱۴۸، ۳، ۱۱، ۱۹۸۳۸ شده است. این معیارها برای دسته چهارم برابر با ۱۷۸۷، ۱۲ ، ۴ و ۱۸۱۹۷ شده است. این معیارها برای دسته پنجم برابر با ۱۰۶۸۰، ۱۱، ۱۳ و ۹۲۹۶ شده است. در جدول ۸ نتایج داده آزمایشی روش درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۸: نتایج داده آزمایشی روش درخت تصمیم بدون خوشه‌بندی
Table 8: Test Data Results of the Decision Tree Method

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	5475	0	0	62	1084
دسته واقعی ۲	0	24	0	0	35
دسته واقعی ۳	201	5	83	164	2094
دسته واقعی ۴	143	0	1	1392	8072
دسته واقعی ۵	55	7	9	461	8072

در جدول ۸ مشاهده می‌شود که از سطر دوم به تعداد ۵۴۷۵ نمونه به درستی برای دسته اول اختصاص یافته‌اند. نمونه‌ای برای دسته ۲ و ۳ پیش‌بینی نشده است. برای دسته چهارم ۶۲ نمونه و دسته پنجم ۱۰۸۴ نمونه پیش‌بینی شده است. از سطر سوم مشاهده می‌شود که برای دسته دوم، به تعداد صفر نمونه به دسته اول، ۲۴ نمونه به درستی به دسته دوم، صفر نمونه به دسته‌های سوم و چهارم و ۳۵ نمونه به دسته پنجم تخصیص یافته است. از سطر چهارم مشاهده می‌شود که به تعداد ۲۰۱ نمونه به دسته اول، ۵ نمونه به دسته دوم، ۸۳ نمونه به درستی به سوم، ۱۶۴ نمونه به دسته چهارم، ۲۰۹۴ نمونه به دسته پنجم اختصاص یافته است. از سطر پنجم مشاهده می‌شود که به تعداد ۱۴۳ نمونه به دسته اول، صفر نمونه به دسته دوم، ۱ نمونه به دسته سوم و ۱۳۹۲ نمونه به دسته چهارم و ۸۰۷۲ نمونه به دسته پنجم پیش‌بینی شده است. از سطر ششم مشاهده می‌شود که ۵۵ نمونه به دسته اول، ۷ نمونه به دسته دوم، ۹ نمونه به دسته سوم، ۴۶۱ نمونه به دسته چهارم و ۸۰۷۲ نمونه برای دسته پنجم پیش‌بینی شده است.

در جدول ۹ نتایج معیارهای TP,FP,FN,TN به‌دست‌آمده از روش درخت تصمیم بدون خوشه‌بندی آورده شده است.

جدول ۹: نتایج معیارهای TP,FP,FN,TN داده آزمایش روش درخت تصمیم بدون خوشه‌بندی
Table 9: TP, FP, FN, TN Metrics Results for the Test Data Using the Decision Tree Method Without Clustering

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۴۷۵	۳۹۹	۱۱۴۶	۱۲۹۸۰
دسته واقعی ۲	۲۴	۱۲	۳۵	۱۹۹۲۹
دسته واقعی ۳	۸۳	۱۰	۲۴۶۴	۱۷۴۴۳
دسته واقعی ۴	۱۳۹۲	۶۸۷	۷۷۷	۱۷۱۴۴
دسته واقعی ۵	۸۰۷۲	۳۸۴۶	۵۳۲	۷۵۵۰

در جدول ۹ مشاهده می‌شود که از نظر معیار مثبت درست برای تمام دسته‌ها اعداد بالایی به‌دست‌آمده است. اما مثبت نادرست و منفی نادرست برای دسته‌های ۱، ۳، ۴ و به خصوص پنجم زیاد است. در حالی‌که روش پیشنهادی مقادیر کمتری بدست آورده است.

ب- حالت دوم نتایج مربوط به داده ۱۰۰ هزار نمونه‌ای

در این حالت تعداد ۱۰۰ هزار نمونه برای داده آموزشی در نظر گرفته شده است. برای داده تست تعداد ۲۰ هزار نمونه در نظر گرفته شده است. ابتدا الگوریتم جستجوی فرکتال تصادفی بهبودیافته ترکیب شده با الگوریتم نهنگ را اجرا می‌کنیم. هدف آن خوشه‌بندی داده‌ها به دو خوشه عادی و حمله است. پس از اجرا نتایج به‌دست‌آمده مربوط به اندیس ماتریس خوشه‌های به‌دست‌آمده را ذخیره می‌کنیم. از نتایج به‌دست‌آمده مشاهده شد که به تعداد ۵۳۴۴۵ نمونه به دسته واقعی ۵ اختصاص یافته که مربوط به داده‌های عادی است و تماما به درستی شناخته شده‌اند. اما مابقی نمونه‌ها به منظور طبقه‌بندی درست نیاز به یک دسته‌بندی کننده است که در اینجا از درخت تصمیم استفاده شده است. نتایج بدست آمده از این خوشه‌بندی به مرحله‌ی بعدی طبقه‌بندی داده می‌شوند.

در جدول ۱۰ نتایج مربوط به داده آموزشی با روش پیشنهادی نشان داده شده است. در جدول ۱۱ نتایج مربوط به معیارهای TP,FP,FN,TN آورده شده است.

جدول ۱۰: نتیجه مربوط به پیش بینی دسته‌های داده آموزشی با روش پیشنهادی

Table 10: Results of Training Data Category Prediction Using the Proposed Method

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۳۶۴۴۶	۰	۰	۵	۹
دسته واقعی ۲	۰	۳۳	۲	۱	۷
دسته واقعی ۳	۰	۱	۷۷۰	۲	۱۵
دسته واقعی ۴	۰	۰	۱	۹۲۴۱	۲۲
دسته واقعی ۵	۰	۰	۰	۰	۵۳۴۴۵

در جدول ۱۰ مشاهده می‌شود که برای دسته واقعی ۱ به تعداد ۳۶۴۴۶ پیش‌بینی شده است. به تعداد صفر نمونه به دسته‌های دوم و سوم اختصاص یافته است. به تعداد ۵ و ۹ نمونه برای دسته‌های چهارم و پنجم پیش‌بینی شده است. برای دسته واقعی دوم ، به درستی ۳۳ نمونه پیش‌بینی شده است. از سطر چهارم جدول۱۰ مشاهده می‌شود که به تعداد صفر نمونه مربوط به دسته اول پیش بینی شده است. برای دسته چهارم مشاهده می‌شود که به تعداد ۹۲۴۱ نمونه بدرستی تشخیص داده شده‌اند. اما به تعداد ۱ نمونه به دسته سوم و ۲۲ نمونه به دسته پنجم به اشتباه تخصیص یافته است. از سطر آخر مشاهده می‌شود که به درستی ۵۳۴۴۵ نمونه به درستی به دسته واقعی ۵ اختصاص یافته‌اند.

در جدول ۱۱ معیارهای TP,FP,FN,TN مربوط به داده آموزشی بدست آمده از روش پیشنهادی نشان داده شده است.

جدول ۱۱: نتایج معیارهای TP,FP,FN,TN داده آموزش روش پیشنهادی

Table 11: TP, FP, FN, TN Metrics Results for the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۳۶۴۴۶	۰	۱۴	۶۳۵۴۰
دسته واقعی ۲	۳۳	۱	۱۰	۹۹۹۵۶
دسته واقعی ۳	۷۷۰	۳	۱۸	۹۹۲۰۹
دسته واقعی ۴	۹۲۴۱	۸	۲۳	۹۰۷۲۸
دسته واقعی ۵	۵۳۴۴۵	۵۳	۰	۴۶۵۰۲

در جدول ۱۲ نتایج پیش بینی دسته بندی مربوط به داده آزمایشی با روش پیشنهادی که تعداد آن ۲۰ هزار نمونه است، نشان داده شده است.

جدول ۱۲: نتیجه داده آزمایشی با روش پیشنهادی

Table 12: Test Data Results Using the Proposed Method

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۵۳۶۶	۰	۰	۶۳	۱۱۹۲
دسته واقعی ۲	۰	۲۱	۷	۰	۳۱
دسته واقعی ۳	۱	۵	۱۳۵	۲۲۲	۲۱۸۴
دسته واقعی ۴	۱۶۹	۰	۰	۱۳۷۰	۶۳۰
دسته واقعی ۵	۰	۰	۰	۰	۸۶۰۴

در جدول ۱۳ نتایج مقایسه‌ای میان روش پیشنهادی و درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۱۳: مقایسه روش پیشنهادی و درخت تصمیم بدون خوشه‌بندی

Table 13: Comparison Between the Proposed Method and the Decision Tree Without Clustering

نتایج	روش پیشنهادی		درخت تصمیم بدون خوشه‌بندی	
	داده آموزش	داده آزمایش	داده آموزش	داده آزمایش
دقت	0.9993	0.7748	0.9991	0.7615
خطا	6.5000e-04	0.2252	9.0000e-04	0.2384
حساسیت	0.9483	0.5702	0.9483	0.5640
خصوصیت	0.9997	0.9233	0.9997	0.9201
صحت	0.9930	0.8472	0.9757	0.8118
نرخ مثبت نادرست	2.5337e-04	0.0767	3.1541e-04	0.0799
امتیازF1	0.9683	0.6008	0.9610	0.5875
ضریب همبستگی متیوز	0.9693	0.5882	0.9612	0.5658

در جدول ۱۳ مشاهده می‌شود که مقدار معیار دقت به‌دست‌آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۹۹۳ و داده آموزشی برابر با ۰/۹۹۹۱ است. معیار دقت برای داده‌ی آزمایشی به‌دست‌آمده از روش پیشنهادی برابر با ۰/۷۷۴۸ و ۰/۷۶۱۵ است. معیار خطای به‌دست‌آمده از روش پیشنهادی برای داده آموزش برابر با 6.5000e-04 ، داده آزمایش برابر با ۰/۲۲۵۲ ، این معیار برای روش درخت تصمیم بدون خوشه‌بندی برابر با 9.0000e-04 داده آزمایش برابر با ۰/۲۳۸۴ است. معیار حساسیت به‌دست‌آمده از روش پیشنهادی برای داده آموزش برابر با ۰/۹۴۸۳ ، داده آزمایش برابر با ۰/۵۷۰۲ ، این معیار حساسیت به‌دست‌آمده به کمک درخت تصمیم بدون خوشه‌بندی برای داده آموزش برابر با ۰/۹۴۸۳ و ۰/۵۶۴ شده است. معیار خصوصیت به‌دست‌آمده به کمک روش پیشنهادی برای داده آموزش برابر با ۰/۹۹۹۷ ، داده آزمایش برابر با ۰/۹۲۳۳ شده است. در این معیار به کمک روش دیگر برابر برای داده آموزشی برابر با ۰/۹۹۹۷ شده ولی برای داده آزمایشی برابر با ۰/۹۲۰۱ شده است. معیار صحت به‌دست‌آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۹۳۰ و برای داده آزمایشی برابر با ۰/۸۴۷۲ شده است. این معیار به‌دست‌آمده از روش دیگر برای داده آموزشی برابر با ۰/۹۷۵۷، ۰/۸۱۱۸ است. نرخ مثبت نادرست به‌دست‌آمده از روش پیشنهادی برای داده آموزشی برابر با 2.5337e-04 و برای داده آزمایشی برابر با ۰/۰۷۶۷ است. این معیار برای داده آموزشی برابر با 3.1541e-04 و برای داده آزمایشی برابر با ۰/۰۷۹۹ شده است. معیار امتیاز F1 به‌دست‌آمده از روش پیشنهادی برای داده آموزشی برابر با ۰/۹۶۸۳ و برای داده آزمایشی برابر با ۰/۶۰۰۸ شده است. این معیار برای داده آموزشی برابر با ۰/۹۶۱ و داده آزمایشی برابر با ۰/۵۸۷۵ شده است. در ادامه نتایج مربوط به درخت تصمیم بدون خوشه‌بندی ارائه شده است. در جدول ۱۴ نتایج مربوط به پیش‌بینی دسته‌ها از درخت تصمیم بدون خوشه‌بندی نشان داده شده است.

جدول ۱۴: نتایج مربوط به داده آموزش روش درخت تصمیم بدون خوشه‌بندی

Table 14: Training Data Results Using the Decision Tree Method Without Clustering

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۳۶۴۴۶	۰	۰	۵	۹
دسته واقعی ۲	۰	۳۳	۲	۱	۷
دسته واقعی ۳	۰	۱	۷۷۰	۲	۱۵
دسته واقعی ۴	۰	۰	۱	۹۲۴۱	۲۲
دسته واقعی ۵	۹	۳	۵	۸	۵۳۴۲۰

در جدول ۱۴ مشاهده می‌شود که برای دسته شماره ۱، به تعداد ۳۶۴۴۶ نمونه درست پیش‌بینی شده است. درحالی‌که به تعداد ۵ نمونه به دسته چهار و نمونه به دسته پنج به‌طور نادرست اختصاص‌یافته است. در جدول ۱۵ نتایج معیارهای TP,FP,FN,TN مربوط به داده آموزش با روش پیشنهادی نشان داده شده است.

جدول ۱۵: نتایج معیارهای TP,FP,FN,TN مربوط به داده آموزش با روش پیشنهادی

Table 15: TP, FP, FN, TN Metrics Results for the Training Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۳۶۴۴۶	۹	۱۴	۶۳۵۳۱
دسته واقعی ۲	۳۳	۴	۱۰	۹۹۹۵۳
دسته واقعی ۳	۷۷۰	۸	۱۸	۹۹۲۰۴
دسته واقعی ۴	۹۲۴۱	۱۶	۲۳	۹۰۷۲۰
دسته واقعی ۵	۵۳۴۲۰	۵۳	۲۵	۴۶۵۰۲

در جدول ۱۵ مشاهده می‌شود که معیار TP به‌دست‌آمده از دسته‌های ۱ الی ۵ به ترتیب برابر با ۳۶۴۴۶، ۳۳، ۷۷۰، ۹۲۴۱، ۵۳۴۲۰ می‌باشند. از این جدول مشاهده می‌شود که مقادیر مثبت نادرست و منفی نادرست در مقایسه با روش پیشنهادی بیشتر شده‌اند. معیارهای مثبت درست، منفی درست به‌دست‌آمده از درخت تصمیم بدون خوشه‌بندی در مجموع کمتر از روش پیشنهادی شده است. این نشان می‌دهد که روش پیشنهادی از نظر معیارهای TP,FP,FN,TN به‌طورکلی موفق‌تر عمل نموده است. در جدول ۱۶ نتایج مربوط به داده آزمون به کمک روش درخت تصمیم بدون خوشه‌بندی نشان داده است.

جدول ۱۶: نتایج دسته‌بندی مربوط به داده آزمایش روش درخت تصمیم بدون خوشه‌بندی

Table 16: Classification Results for the Test Data Using the Decision Tree Method Without Clustering

دسته	پیش‌بینی کلاس ۱	پیش‌بینی کلاس ۲	پیش‌بینی کلاس ۳	پیش‌بینی کلاس ۴	پیش‌بینی کلاس ۵
دسته واقعی ۱	۵۳۶۶	۰	۰	۶۳	۱۱۹۲
دسته واقعی ۲	۰	۲۱	۷	۰	۳۱
دسته واقعی ۳	۱	۵	۱۳۵	۲۲۲	۲۱۸۴
دسته واقعی ۴	۱۶۹	۰	۰	۱۳۷۰	۶۳۰
دسته واقعی ۵	۶۳	۲	۲	۱۹۸	۸۳۳۹

در جدول ۱۶ مشاهده می‌شود که پیش‌بینی‌ها برای داده آزمایشی در حد مناسبی انجام شده است اما روش پیشنهادی در مقایسه با درخت تصمیم بدون خوشه‌بندی موفق‌تر عمل نموده است.

در جدول ۱۷ نتایج مربوط به TP,FP,FN,TN داده آزمایشی به کمک روش درخت تصمیم بدون خوشه‌بندی نشان داده است.

جدول ۱۷: نتایج معیارهای TP,FP,FN,TN داده آزمایش روش پیشنهادی

Table 17: TP, FP, FN, TN Metrics Results for the Test Data Using the Proposed Method

دسته	مثبت درست	مثبت نادرست	منفی نادرست	منفی درست
دسته واقعی ۱	۵۳۶۶	۱۷۰	۱۲۵۵	۱۳۲۰۹
دسته واقعی ۲	۲۱	۵	۳۸	۱۹۹۳۶
دسته واقعی ۳	۱۳۵	۷	۲۴۱۲	۱۷۴۴۶
دسته واقعی ۴	۱۳۷۰	۲۸۵	۷۹۹	۱۷۵۴۶
دسته واقعی ۵	۸۶۰۴	۴۰۳۷	۰	۷۳۵۹

در جدول ۱۷ مشاهده می‌شود که برای دسته واقعی ۱، معیار مثبت درست ۵۳۶۶ ، مثبت نادرست ۱۷۰، منفی نادرست ۱۲۵۵، منفی درست برابر با ۱۳۲۰۹ است. از این جدول مشاهده می‌شود که معیار مثبت درست در مقایسه با مثبت نادرست خیلی بیشتر است.

باتوجه‌به جداول ۱۰ الی ۱۷ مشاهده می‌شود که برای مجموعه داده ۱۰۰ هزارتایی نیز در مجموع نتایج به‌دست‌آمده از روش پیشنهادی خوشه‌بندی با الگوریتم جستجوی فرکتال تصادفی بهبودیافته ترکیب‌شده با الگوریتم نهنگ بهتر از حالت بدون خوشه‌بندی است.

۵- نتیجه‌گیری

در این مقاله یک الگوریتم بهینه‌سازی جستجوی فرکتال تصادفی بهبودیافته با نگاشت‌های آشوبی ترکیب‌شده با الگوریتم بهینه‌سازی نهنگ معرفی شد. از این الگوریتم برای خوشه‌بندی داده‌ی NSL-KDD به دو خوشه حمله و عادی استفاده‌شده است. در این مرحله الگوریتم پیشنهادی به داده اعمال شد و نتایج آن نشان داد که به طور کامل داده‌های عادی از حملات جداشده‌اند و هرکدام در خوشه خود جای گرفته‌اند. دو مثال برای نشان‌دادن این مورد انجام شد که مثال اول با ۲۰ هزار نمونه برای داده آموزش، ۲۰ هزار نمونه برای داده آزمایش و مثال دوم شامل ۱۰۰ هزار نمونه برای داده آموزش و ۲۰ هزار نمونه برای داده آزمایش انجام شد. نتایج نشان داد که در هر دو مثال مذکور خوشه‌های مربوط به عادی و نفوذ یافته به‌درستی و به طور کامل از هم تفکیک‌شده‌اند. این خوشه‌ها ذخیره‌شده و به مرحله‌ی بعدی طبقه‌بندی اعمال‌شده‌اند. سپس از روش درخت تصمیم برای طبقه‌بندی داده‌های حملات استفاده شده است. در این مرحله دسته مربوط به داده‌های نرمال از پیش از آن درست تشخیص‌داده‌شده و تنها فقط دسته‌بندی برای حملات باید انجام شود. برای ارزیابی روش پیشنهادی نتایج آن باحالت درخت تصمیم بدون خوشه‌بندی مقایسه شده است. هر ۵ دسته باید ارزیابی شوند تا مقایسه عادلانه با روش مشابه درخت تصمیم بدون خوشه‌بندی انجام شود. نتایج نشان داده‌اند که از نظر معیارهای TP,FP,FN,TN روش پیشنهادی برای تعداد نمونه‌های ۲۰ هزارتایی داده آموزش، ۲۰ هزارتایی داده آزمایش، ۱۰۰ هزارتایی داده آموزش، ۲۰ هزارتایی داده آزمایش موفق‌تر از روش درخت تصمیم بدون خوشه‌بندی عمل نموده است. نتایج از نظر معیارهای دقت، خاصیت، حساسیت در هر دو مثال برای داده‌های آموزشی و آزمایشی انجام شد. نتایج برای مثال اول نشان داد که معیار دقت به‌دست‌آمده از روش پیشنهادی برابر با ۰/۹۹۸۹ درحالی‌که روش دیگر برابر با ۰/۹۹۸۲ به دست آورده است. معیار حساسیت به‌دست‌آمده از روش پیشنهادی برابر با ۰/۹۶۷۳ شده است درحالی‌که این معیار از روش دیگر برابر با ۰/۹۶۷۱ شده است. ازنظر خصوصیت روش پیشنهادی برابر با ۰/۹۹۹۶ و روش درخت تصمیم بدون خوشه‌بندی برابر با ۰/۹۹۹۵ شده است. نتایج به‌دست‌آمده برای داده آزمایشی برای معیار دقت از روش پیشنهادی برابر با ۰/۷۷۸۹، درحالی‌که برای روش دیگر برابر با ۰/۷۵۲۳ شده است. از نظر معیار حساسیت با روش پیشنهادی برابر با ۰/۵۸۱۶ و روش دیگر برابر با ۰/۵۶۹۲ شده است. از نظر معیار خصوصیت روش پیشنهادی مقدار ۰/۹۲۴۸ و روش دیگر برابر با ۰/۹۱۸۶ به دست آورده است.

نتایج به‌دست‌آمده برای مثال دوم حالت ۱۰۰ هزار نمونه داده آموزشی و ۲۰ هزار نمونه داده آزمایشی به این صورت شده‌اند. برای داده آموزشی از نظر معیار دقت روش پیشنهادی توانسته است که مقدار ۰/۹۹۹۳ و روش درخت تصمیم بدون خوشه‌بندی برابر با ۰/۹۹۹۱ به دست آورد. ازنظر معیار حساسیت روش پیشنهادی توانسته است که مقدار ۰/۹۴۸۳ و روش دیگر برابر با ۰/۹۴۸۳ به دست آورد. از نظر معیار خاصیت روش پیشنهادی مقدار ۰/۹۹۹۷ و روش دیگر ۰/۹۹۹۷ بدست آورده است. اما برای داده‌های آزمایشی روش پیشنهادی برای معیار دقت برابر با ۰/۷۷۴۸، درحالی‌که روش درخت تصمیم بدون خوشه‌بندی برابر با ۰/۷۶۱۵ به دست آورده است. از نظر معیار حساسیت ۰/۵۷۰۲ ، ۰/۵۶۴۰ شده است. برای معیار خاصیت روش پیشنهادی توانسته که مقدار ۰/۹۲۳۳ و روش درخت تصمیم بدون خوشه‌بندی مقدار ۰/۹۲۰۱ به دست آورد. در مجموع از نتایج گرفته شده نتیجه می‌شود که روش پیشنهادی خوشه‌بندی با الگوریتم جستجوی فرکتال تصادفی بهبودیافته با نگاشت‌های آشوب ترکیب‌شده با الگوریتم نهنگ موفق‌تر از حالت بدون خوشه‌بندی عمل نموده است.

به‌منظور پیشنهاد برای تحقیقات آینده، می‌توان از این الگوریتم از نوع باینری آن برای انتخاب ویژگی برای افزایش دقت طبقه-بندی استفاده نمود. سپس نتایج آن را برای داده NSL-KDD و داده‌های مشابه در تحقیقات استفاده نمود. یک پیشنهاد دیگر استفاده از روش‌های طبقه کننده پیچیده‌تر مانند یادگیری عمیق بهبود با الگوریتم پیشنهادی جستجوی فرکتال تصادفی بهبودیافته ترکیب‌شده با الگوریتم نهنگ است که می‌تواند با روش پیشنهادی برای افزایش دقت طبقه‌بندی استفاده شود.

مراجع

[1] Diro, A. A., & Chilamkurti, N., Distributed attack detection scheme using deep learning approach for Internet of Things. Future Generation Computer Systems, 82, 761-768, 2018

[2] Roseline Oluwaseun Ogundokun, Joseph Bamidele Awotunde, Peter Sadiku, Emmanuel Abidemi Adeniyi, Moses Abiodun, Oladipo Idowu Dauda, An Enhanced Intrusion Detection System using Particle Swarm Optimization Feature Extraction Technique,Procedia Computer Science,Volume 193, 2021, Pages 504-512, ISSN 1877-0509.

[3] S. Hofmeyr, S. Forrest, and A. Sornayaji, “Lightweight intrusion detection for networked operating systems,” Journal of Computer Security, vol. 5, no. 2, 1997.

[4] A. Puri and N. Sharma, “A novel technique for intrusion detection system for network security using hybrid svm-cart,” IJEDR, vol. 5, no. 2, pp. 155–161, 2017.

[5] P. Kabiri and A. A. Ghorbani, “Research on intrusion detection and response: A survey.” IJ Network Security, vol. 1, no. 2, pp. 84–102, 2005.

[6] M. Shojafar, R. Taheri, Z. Pooranian, R. Javidan, A. Miri and Y. Jararweh, "Automatic Clustering of Attacks in Intrusion Detection Systems," 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), 2019, pp. 1-8, doi: 10.1109/AICCSA47632.2019.9035238.

[7] A. Karami, “An anomaly-based intrusion detection system in presence of benign outliers with visualization capabilities, “Expert Systems with Applications, vol. 108, pp. 36–60, 2018.

[8] M. Tabash, M. Abd Allah, and B. Tawfik, “Intrusion detection model using naive bayes and deep learning technique,” 3e International Arab Journal of Information Technology, vol. 17,no. 2, 2020.

[9] A. Ghazali, W. Nuaimy, A. Al-Atabi, and I. Jamaludin, “Comparison of classification models for Nsl-Kdd dataset for network anomaly detection,” Academic Journal of Science, vol. 4, no. 1, pp. 199–206, 2015.

[10] J. Kevric, S. Jukic, and A. Subasi, “An effective combining classifier approach using tree algorithms for network intrusion detection,” Neural Computing & Applications, vol. 28,no. S1, pp. 1051–1058, 2017.

[11] A. Hadi, “Performance analysis of big data intrusion detection system over random forest algorithm,” International Journal of Applied Engineering Research, vol. 13, no. 2, pp. 1520–1527,2018.

[12] W. Elmasry, A. Akbulut, and A. H. Zaim, “Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic,” Computer Networks, vol. 168,Article ID 107042, 2020

[13] Zhu, Y., Gaba, G. S., Almansour, F. M., Alroobaea, R., & Masud, M. (2021). Application of data mining technology in detecting network intrusion and security maintenance. Journal of Intelligent Systems, 30(1), 664-676.

[14] Anitha, P., & Kaarthick, B. (2021). Oppositional based Laplacian grey wolf optimization algorithm with SVM for data mining in intrusion detection system. Journal of Ambient Intelligence and Humanized Computing, 12(3), 3589-3600.

[15] Koryshev, N., Hodashinsky, I., & Shelupanov, A. (2021). Building a fuzzy classifier based on whale

optimization algorithm to detect network intrusions. Symmetry, 13(7), 1211.

[16] Zhang, J., Sun, J., & He, H. (2021). Clustering Detection Method of Network Intrusion Feature Based on Support Vector Machine and LCA Block Algorithm. Wireless Personal Communications, 1-15.

[17] Maheswari, M., & Karthika, R. A. (2021). A novel QoS based secure unequal clustering protocol with intrusion detection system in wireless sensor networks. Wireless Personal Communications, 118(2), 1535-1557.

[18] Xie, B., Dong, X., & Wang, C. (2021). An Improved-Means Clustering Intrusion Detection Algorithm for Wireless Networks Based on Federated Learning. Wireless Communications and Mobile Computing, 2021.

[19] Keserwani, P. K., Govil, M. C., Pilli, E. S., & Govil, P. (2021). A smart anomaly-based intrusion detection system for the Internet of Things (IoT) network using GWO–PSO–RF model. Journal of Reliable Intelligent Environments, 7(1), 3-21.

[20] Dwivedi, S., Vardhan, M., & Tripathi, S. (2021). Building an efficient intrusion detection system using grasshopper optimization algorithm for anomaly detection. Cluster Computing, 24(3), 1881-1900.

[21] Haider AL-Husseini, et al. Whale Optimization Algorithm-Enhanced Long Short-Term Memory Classifier with Novel Wrapped Feature Selection for Intrusion Detection, Journal of Sensor and Actuator Networks, Nov 2024, 13, 73, https://doi.org/10.3390/jsan13060073

[22] Rajashekar Kandakatla, et al. Whale-Optimized Probabilistic Selection For Enhanced Intrusion Detection In Cloud Environments, Journal of Theoretical and Applied Information Technology, June 2024. Vol. 102. No. 12

[23] Salimi, H., Stochastic fractal search: a powerful metaheuristic algorithm. Knowledge-Based Systems, 2015. 75: p. 1-18

[24] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. IEEE Trans Syst Man Cybern. 1991;21(3):660–674.

[25] Misaghi, M., & Yaghoobi, M. (2019). Improved invasive weed optimization algorithm (IWO) based on chaos theory for optimal design of PID controller. Journal of Computational Design and Engineering, 6(3), 284-295.

[26] Kumar Ahuja, Dr. Gulshan. (2015). Evaluation Metrics for Intrusion Detection Systems-A Study. International Journal of Computer Science and Mobile Applications. 11.