JOURNAL OF SOUTHERN COMMUNICATION ENGINEERING
ISLAMIC AZAD UNIVERSITY BUSHEHR BRANCH

E-ISSN: 2980-9231
https://jce.bushehr.iau.ir

https://doi.org/10.30495/jce.2022.695339

Vol. 12/ No. 46/Winter 2023

Research Article

Design of Anomaly-Based Intrusion Detection System Using Support Vector Machine and Grasshopper Optimization Algorithm in IoT

Sepehr Sharifi, MSc Student [1] | Soulmaz Gheisari, Assistant Professor [2*]

[1]Faculty of Mechanics, Electrical Power and Compute, Science and Research Branch, Islamic Azad university, Tehran, Iran,
sepehr_sh@ymail.com

[2]Department of computer engineering, Faculty of Engineering, Islamic Azad University, Pardis Branch, Iran,
so_gheisari@pardisiau.ac.ir

Correspondence
Soulmaz Gheisari, Assistant Professor, Department of computer engineering, Faculty of Engineering, Islamic Azad University, Pardis Branch, Iran,
so_gheisari@pardisiau.ac.ir

Received: 30 August 2022
Revised: 1 October 2022
Accepted: 4 October 2022

Abstract

Computer networks play an important and practical role in communication and data exchange, and they also share resources with complete ease. Today, various types of computer networks have emerged, one of which is the Internet of Things. In the Internet of Things, network nodes can be smart objects, and in this sense, this network has many nodes and there is a lot of traffic in this network. Like any computer network, it faces its own challenges and problems, one of which is the issue of network intrusion and disruption. This dissertation focuses on detecting anomaly-based intrusion into the Internet of Things using data mining. In this study, after collecting and preparing data, the improved support vector machine with grasshopper optimization algorithm is used as a proposed method to detect anomaly-based intrusion in the Internet of Things. The bagging and k-nearest neighbor classifiers and Basic SVM are compared based on error types and standard performance criteria. The simulation results show 97.2% accuracy in the proposed method and better performance compared to other methods.

Keywords: Anomaly-Based Intrusion Detection, IoT, Support Vector Machine, Grasshopper Optimization Algorithm

Highlights

- Using the grasshopper optimization algorithm to optimize all support vector machines hyper parameters that require manual adjustment.
- Using Fisher's discriminant analysis algorithm in order to maximize the inter-class distance and minimize the intra-class distance with the aim of creating higher resolution.
- Designing an anomaly-based intrusion detection system in such a way that the parameters are optimally determined by the algorithm.

Citation: S. Sharifi and S. Gheisari, "Design of anomaly based intrusion detection system using support vector machine and grasshopper optimization algorithm in IoT," *Journal of Southern Communication Engineering*, vol. 12, no. 46, 2022, doi: 10.30495/jce.2022.695339, (in Persian).

مقاله پژوهشی

طراحی سیستم تشخیص نفوذ مبتنی بر ناهنجاری با استفاده از ماشین بردار پشتیبان و الگوریتم بهینه‌سازی ملخ در اینترنت اشیا

سپهر شریفی [۱] | سولماز قیصری *[۲]

[۱] دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، ایران.
sepehr_sh@ymail.com

[۲] استادیار گروه کامپیوتر، دانشکده فنی مهندسی ، دانشگاه آزاد اسلامی واحد پردیس ، ایران
so_gheisari@pardisiau.ac.ir

نویسنده مسئول
* سولماز قیصری، استادیار گروه کامپیوتر، دانشکده فنی مهندسی ، دانشگاه آزاد اسلامی واحد پردیس ، ایران.
so_gheisari@pardisiau.ac.ir

تاریخ دریافت: ۸ شهریور ۱۴۰۱
تاریخ بازنگری: ۹ مهر ۱۴۰۱
تاریخ پذیرش: ۱۲ مهر ۱۴۰۱

https://doi.org/10.30495/jce.2022.695339

چکیده:

امروزه شبکه‌های کامپیوتری نقش مهم و کاربردی در ارتباطات و تبادل داده‌ها دارند و در زندگی انسان‌ها انواع مختلفی از شبکه‌های کامپیوتری پا به عرصه وجود گذاشته است که یکی از آن‌ها شبکه اینترنت اشیاء است. در اینترنت اشیا گره‌های شبکه می‌توانند اشیا هوشمند باشد و از این نظر این شبکه دارای گره‌های زیادی است و ترافیک بالایی در این شبکه وجود دارد. مانند هر شبکه کامپیوتری، اینترنت اشیا با چالش‌ها و مشکلات خاص خود مواجه است که یکی از آن‌ها مسئله نفوذ به شبکه و ایجاد اختلال در آن است. در این مقاله تمرکز بر روی تشخیص نفوذ مبتنی بر ناهنجاری در شبکه اینترنت اشیا با استفاده از داده‌کاوی است. در این پژوهش پس از جمع‌آوری و آماده‌سازی داده‌ها از ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی ملخ به‌عنوان روش پیشنهادی به‌منظور تعیین بهینه پارامترهای ماشین بردار پشتیبان در جهت تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا استفاده شد و نتایج با طبقه‌بندهای بگینگ و k-نزدیک‌ترین همسایه و ماشین بردار پشتیبان پایه بر اساس انواع خطا و تحلیل آماری خطا مورد مقایسه قرار گرفت. نتایج شبیه‌سازی نشان از دقت ٪۹۷/۲در روش پیشنهادی و عملکرد بهتر در مقایسه با سایر روش‌ها دارد.

کلید واژه‌ها: تشخیص نفوذ مبتنی بر ناهنجاری، اینترنت اشیا، ماشین بردار پشتیبان، الگوریتم بهینه‌سازی ملخ.

۱-مقدمه

در سال‌های اخیر شاهد توسعه سریع و استقرار برنامه‌های IoT[1] در حوزه‌های مختلف هستیم. در حال حاضر در این سناریو، مردم آماده هستند تا از مزایای اینترنت اشیا استفاده نمایند. اینترنت اشیا به‌عنوان سومین موج در تکامل اینترنت ظاهر شده است. موج اینترنت دهه ۱۹۹۰، ۱/۲ میلیارد مشترک را به شبکه جهانی متصل کرد درحالی‌که موج موبایل در سال ۲۰۰۰، ۲/۴ میلیارد دیگر را نیز به هم متصل کرد. در واقع، طبق نظر شرکت بین‌المللی داده ، انتظار می‌رود IoT تا سال ۲۰۲۵ بیش از ۸۴ میلیارد دستگاه متصل متشکل از ۱۸۶زیتابایت داده تولید کند[۸]. با افزایش تعداد دستگاه‌های به‌هم‌پیوسته و متنوع سازی برنامه‌های آن، اینترنت اشیا به‌سرعت در حال رشد در صنایع مختلف است، درحالی‌که فناوری‌های IoT هنوز به بلوغ نرسیده‌اند و برای رفع آن چالش‌های بسیاری وجود دارد. اینترنت اشیا ترکیبی از دنیای واقعی و مجازی را در هرکجا و در هر زمان ممکن می‌سازد که بدین سبب توجه هکرها را مجذوب خود می‌کند. زیرا ترک دستگاه‌ها بدون نظارت انسان برای مدت طولانی می‌تواند منجر به سرقت آن‌ها و یا اطلاعات شود و IoT موارد بسیاری از این قبیل را در برمی‌گیرد[۲].

[1] Internet of Things

اینترنت اشیا یکی از فناوری‌های نوظهور است که توجه محققان دانشگاه و صنایع را به خود جلب کرده است. هدف اصلی اینترنت برقرار کردن اتصال بین اشیاء و انسان‌ها، با یکدیگر برای رسیدن به اهداف مشترک است. در آینده‌ای نزدیک انتظار می‌رود IoT به‌طور یکپارچه در محیط‌زیست ما ادغام شود و انسان کاملاً با این سبک زندگی آشنا و به این فناوری وابسته شود. لذا هرگونه سازوکار امنیتی در این سیستم مستقیماً بر زندگی انسان تأثیر خواهد گذاشت[۳].

اینترنت اشیا، مبتنی بر دستگاه‌های به‌هم‌پیوسته، انواع خدمات جدید و بلادرنگ را امکان‌پذیر می‌کند که در یک محیط سنتی محقق نمی‌شوند و بسیاری از این سرویس‌ها اطلاعات حساس و خصوصی را که متعلق به کاربران فردی است، برداشت می‌کنند. متأسفانه، عملکردهای امنیتی موجود برای محافظت از چنین اطلاعاتی به دلیل ظرفیت‌ها، کارکردها و نیازهای امنیتی دستگاه‌هایIoT، بسیار دشوار است. بنابراین شکست‌های امنیتیIoT می‌تواند شدید باشد، به همین دلیل مطالعه و تحقیق در مورد موضوعات امنیتی در IoT از اهمیت ویژه‌ای برخوردار است. هدف اصلی امنیت IoT حفظ حریم خصوصی، محرمانه بودن، اطمینان از امنیت کاربران، تشخیص نفوذ در زیرساخت‌ها، داده‌ها و دستگاه‌های IoT و تضمین دسترسی به خدمات ارائه‌شده توسط یک اکوسیستم IoT است. بنابراین، تحقیقات در زمینه امنیت IoT به‌تازگی با کمک ابزارهای شبیه‌سازی موجود، مدل‌سازها و سیستم عامل‌های محاسباتی و آنالیز در حال پیشرفت است[۱۲].

علیرغم آنکه دستگاه‌های تشخیص نفوذ متعددی برای تشخیص نفوذ در اینترنت اشیا پیشنهادشده است اما دقت تشخیص پایین، هشدار کاذب بالا و زمان پردازش زیاد ازجمله مشکلات موجود است به همین دلیل اهمیت و ضرورت این تحقیق افزایش دقت در تشخیص نفوذ در اینترنت اشیا توسط ماشین بردار پشتیبان بهبودیافته با استفاده از الگوریتم بهینه‌سازی ملخ است. جهت افزایش امنیت و تشخیص نفوذ در شبکه‌های اینترنت اشیا، ترافیک ورودی توسط IDS[۲] آنالیز می‌شود. در این تحقیق به‌منظور تشخیص نفوذ از ماشین بردار پشتیبان استفاده‌شده است تا عملیات طبقه‌بندی و تفکیک بسته‌های مخرب بر روی‌داده‌های ورودی صورت پذیرد. جهت افزایش دقت طبقه‌بندی ماشین بردار پشتیبان، ابتدا با استفاده از الگوریتم تحلیل تفکیکی فیشر[۳] ابعاد ورودی را کاهش داده و ویژگی‌های اصلی به‌منظور دسته‌بندی حملات در ماشین بردار پشتیبان انتخاب می‌گردند. همچنین ازآنجاکه ماشین بردار پشتیبان به‌منظور عملکرد هرچه بهینه‌تر نیازمند تنظیم پارامترهای مختلف است و از چالش‌های اساسی در آن تعیین بهینه پارامترهای ضریب جریمه، تابع هسته و غیره است که توسط کاربر انجام می‌شود و ممکن است مقادیر بهینه‌ای توسط کاربر انتخاب نشود و دقت در تشخیص نفوذ کاهش یابد، بنابراین در این تحقیق سعی می‌شود پارامترهای ضریب جریمه و تابع هسته در ماشین بردار پشتیبان به‌عنوان متغیرهای تصمیم‌گیری تعریف شوند و میانگین مربعات خطا به‌عنوان تابع هدف تعریف شود و مقادیر بهینه این پارامترها به‌صورت خودکار توسط الگوریتم بهینه‌سازی ملخ تعیین شده تا تشخیص نفوذ در اینترنت اشیا با دقت بالاتری انجام شود.از نظر نوآوری این تحقیق تمام پارامترهای ماشین بردار پشتیبان را به‌صورت بهینه و بدون دخالت انسان تعیین می‌کند. همچنین استفاده از الگوریتم بهینه‌سازی جدید ملخ و همچنین نگاشت ویژگی‌ها به فضایی که بالاترین تفکیک‌پذیری را ایجاد کند از دیگر موارد این تحقیق است. شبیه‌سازی و مطالعات انجام شده نشان می‌دهد استفاده از این الگوریتم‌ها در کنار هم به‌منظور تشخیص ناهنجاری در شبکه‌های اینترنت اشیا باعث افزایش دقت و کاهش هشدار کاذب می‌گردد.

۲-مروری بر کارهای انجام شده

غفاریان و همکاران در سال ۱۳۹۸، به‌مرور روش‌های تشخیص نفوذ در اینترنت اشیا پرداختند. اینترنت اشیا الگوی جدیدی است که اینترنت و اشیای فیزیکی متعلق به زمینه‌های مختلف مانند اتوماسیون خانه، فرایند صنعتی، سلامت انسان و نظارت محیطی را یکپارچه می‌سازد. وجود دستگاه‌های متصل شده اینترنتی را در فعالیت‌های روزمره ما عمیق‌تر می‌سازد، علاوه بر داشتن مزایای زیاد، چالش‌های مربوط به مباحث امنیتی زیادی به بار می‌آورد. بیش از دو دهه است که دستگاه‌های تشخیص نفوذ، ابزار مهمی برای حفاظت از شبکه‌ها و دستگاه‌های اطلاعاتی گشته‌اند. بااین‌حال، به‌کارگیری تکنیک‌های IDS سابق برای اینترنت اشیا به دلیل ویژگی‌های خاصش مانند دستگاه‌هایی با منبع محدود، پشته پروتکل‌های ویژه و استانداردهای خاص سخت است. در این

[2] Intrusion Detection System
[3] Fisher

مقاله، مروری بر تلاش‌های پژوهشی IDS برای اینترنت اشیا ارائه شد. هدف شناسایی روندهای برجسته، مباحث آزاد و پژوهش-های آتی است IDSهای پیشنهادشده در مقالات را طبق مشخصه‌های زیر طبقه‌بندی کردند: روش تشخیص استراتژی جایگذاری IDS، تهدید امنیتی و استراتژی معتبرسازی همچنین در مورد امکانات مختلف برای هر مشخصه با شرح جزئیات تحقیقات پیشنهادشده برای طرح‌های IDS ویژه اینترنت اشیا یا طراحی استراتژی‌های تشخیص حمله برای تهدیدات اشیای تعبیه شده در IDSها بحث شد[۱۰].

خاطر و همکاران در سال ۲۰۱۹، یک پرسپترون[۴] سبک‌وزن مبتنی بر سیستم تشخیص نفوذ در محاسبات مه پرداختند. محاسبات مه، پارادایمی است که برای رفع مشکلات ذاتی ابر مانند تأخیر و عدم پشتیبانی از تحرک و آگاهی از مکان، محاسبات ابری و خدمات را تا لبه شبکه گسترش می‌دهد. مه یک سکوی غیرمتمرکز است که قادر به کار و پردازش داده‌ها به‌صورت محلی است و می‌تواند در سخت‌افزار ناهمگن نصب شود و آن را برای برنامه‌های اینترنت اشیا ایده‌آل کند. دستگاه‌های تشخیص نفوذ (IDS) برای اطمینان از کیفیت خدمات، بخش جدایی‌ناپذیر از هر سیستم امنیتی برای مه و شبکه‌های اینترنت اشیا است. با توجه به محدودیت منابع مه و دستگاه‌های اینترنت اشیا سیستم تشخیص نفوذ سبک‌وزن بسیار مطلوب است. در این مقاله، یک سیستم تشخیص نفوذ سبک‌وزن مبتنی بر بازنمایی فضای بردار با استفاده مدل پرسپترون چندلایه ارائه‌شده است. سیستم تشخیص نفوذ ارائه شده با داده‌های ADFA-LD[۵] و ADFA-WD[۶] ارزیابی شده است. شبیه‌سازی نشان می‌دهد با استفاده از یک لایه پنهان و تعداد کمی نورون دقت ۹۴٪ بر روی مجموعه داده‌ای ADFA-LD حاصل شده است و دقت ۷۴٪ در مجموعه داده‌ای ADFA-WD حاصل شده است[۳].

روپک و همکاران در سال ۲۰۲۰، یک سیستم تشخیص نفوذ برای مقابله با حملات DDOS[۷] در شبکه‌های اینترنت اشیا ارائه دادند. در این مقاله، یک سیستم تشخیص نفوذ (IDS[۸]) با استفاده از ترکیبی از تکنیک یادگیری عمیق و روش بهینه‌سازی چندهدفی برای تشخیص حملات DDoS در شبکه‌های IoT ارائه‌شده است. شبکه‌های IoT از دستگاه‌های مختلفی با تنظیمات سخت‌افزاری و نرم‌افزاری منحصربه‌فرد در ارتباط با پروتکل‌های مختلف ارتباطی تشکیل شده است، که داده‌های چندبعدی بزرگی تولید می‌کنند که باعث می‌شود شبکه‌های IoT مستعد حملات سایبری شوند. در یک شبکه، IDS ابزاری اساسی برای محافظت از آن در برابر حملات سایبری است. شناسایی تهدیدات سایبری جدید در حال ظهور برای IDS موجود دشوار می‌شود، بنابراین نیاز به شناسایی پیشرفته است. حمله DDOS از حملات سایبری است که مشکلات قابل توجهی در شبکه اینترنت اشیا به همراه داشته است.[۵].

صفادین و همکاران در سال ۲۰۲۰ در مقاله‌ای با عنوان بهینه‌سازی ماشین بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی گرگ خاکستری به‌منظور شناسایی تهدیدات امنیتی در شبکه‌های حسگر بی‌سیم به ارائه سیستمی به این منظور پرداختند. نفوذ در شبکه‌های حسگر بیسیم با هدف کاهش یا حتی حذف برخی قابلیت‌های این شبکه‌ها در ارائه بهتر عملکردهای خود انجام می‌شود. در این مقاله، یک سیستم تشخیص نفوذ پیشرفته (IDS) با استفاده از بهینه‌ساز گرگ خاکستری باینری اصلاح شده با ماشین بردار پشتیبان[۹] (GWOSVM-IDS) پیشنهادشده است . هدف روش پیشنهادی افزایش دقت تشخیص نفوذ و نرخ تشخیص و کاهش زمان پردازش در محیطWSN[۱۰] از طریق کاهش نرخ هشدارهای کاذب و کاهش تعداد ویژگی‌های حاصل از IDSها در محیط WSN است. در واقع، مجموعه دادهNSL KDD'99[۱۱] برای نشان دادن عملکرد روش پیشنهادی و مقایسه آن با سایر روش‌های موجود استفاده می‌شود. روش‌های پیشنهادی از نظر دقت، تعداد ویژگی‌ها، زمان اجرا، نرخ هشدار نادرست و نرخ تشخیص ارزیابی می‌شوند. نتایج نشان داد که GWOSVM-IDS پیشنهادی با هفت گرگ بر سایر الگوریتم‌های پیشنهادی و مقایسه‌ای غلبه دارد .[۶]

---

[4] Perceptron
[5] Australian Defence Force Academy Linux Dataset
[6] Australian Defence Force Academy Windows Dataset
[7] Distributed Denial-Of-Service
[8] Intrusion Detection System
[9] Grey wolf optimizer with support vector machine
[10] Wireless sensor networks
[11] National Security Lab Knowledge Discovery and Data Mining

پس از بررسی روش‌های مشابه و تحقیقات انجام شده در این زمینه مشخص گردید در این مقالات پیچیدگی زمان آموزش به دلیل عدم کاهش ابعاد مسئله بالا است و در نتیجه زمان پردازش بیشتری نیاز دارند که این مورد در اینترنت اشیا از اهمیت بالایی برخوردار هست. همچنین عدم تنظیم پارامترهای ماشین بردار پشتیبان در روش‌های مطرح شده و تنظیم آن‌ها به‌صورت دستی باعث می‌شود دقت طبقه بند به حداکثر نرسد. علاوه بر موارد مذکور عدم مقایسه با متدهای مشابه و معمول و همچنین ماشین بردار پشتیبان پایه از دیگر نکات مهم در تحقیقات پیشین است.

۳-روش پیشنهادی

این پژوهش بر مبنای طراحی سیستم تشخیص نفوذ مبتنی بر ناهنجاری در محیط اینترنت اشیا از ۶ مرحله کلی تشکیل شده است:

۱) جمع‌آوری داده‌ها از پایگاه داده
۲) پاک‌سازی داده‌ها
۳) نرمال‌سازی داده‌ها
۴) انتخاب ویژگی
۵) تشخیص نفوذ در اینترنت اشیا با استفاده از ماشین بردار پشتیبان بهبود یافته توسط الگوریتم بهینه‌سازی ملخ
۶) مقایسه نتایج با طبقه بندهای بگینگ و k-نزدیک‌ترین همسایه و ماشین بردار پشتیبان پایه

شکل ۱ نمای کلی ۶ گام لازم برای تشخیص نفوذ در محیط اینترنت اشیا را نشان می‌دهد.

شکل ۱: سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا

۳-۱ جمع‌آوری داده‌ها

در این پژوهش از پایگاه داده NSL-KDD استفاده‌شده است. پایگاه داده شامل ۴۱ ویژگی است که در جدول ۱ نمایش داده شده است.

داده‌های حمله و نرمال به‌صورت متوازن در دیتاست توزیع شده‌اند و در شبیه‌ساز متلب به‌صورت تصادفی توسط تابع randperm درهم‌ریخته و به نسبت ۸۰ به ۲۰ به ترتیب برای داده‌های آموزشی و آزمایشی تقسیم‌بندی شدند.

جدول ۱: ویژگی‌های مجموعه داده‌ای NSL-KDD

شماره ویژگی	نام ویژگی	نوع ویژگی	شماره ویژگی	نام ویژگی	نوع ویژگی
۱	duration	Continuous	۲۲	is_guest_login	Continuous
۲	protocol_type	Symbolic	۲۳	count	Continuous
۳	service	Symbolic	۲۴	srv_count	Continuous
۴	flag	Symbolic	۲۵	serror_rate	Continuous
۵	src_bytes	Continuous	۲۶	srv_serror_rate	Continuous
۶	dst_bytes	Continuous	۲۷	rerror_rate	Continuous
۷	land	Continuous	۲۸	srv_rerror_rate	Continuous
۸	wrong_fragment	Continuous	۲۹	same_srv_rate	Continuous
۹	urgent	Continuous	۳۰	diff_srv_rate	Continuous
۱۰	hot	Continuous	۳۱	srv_diff_host_rate	Continuous
۱۱	num_failed_logins	Continuous	۳۲	dst_host_count	Continuous
۱۲	logged_in	Continuous	۳۳	dst_host_srv_count	Continuous
۱۳	num_compromised	Continuous	۳۴	dst_host_same_srv_rate	Continuous
۱۴	root_shell	Continuous	۳۵	dst_host_diff_srv_rate	Continuous
۱۵	su_attempted	Continuous	۳۶	dst_host_same_src_port_rate	Continuous
۱۶	num_root	Continuous	۳۷	dst_host_srv_diff_host_rate	Continuous
۱۷	num_file_creations	Continuous	۳۸	dst_host_serror_rate	Continuous
۱۸	num_shells	Continuous	۳۹	dst_host_srv_serror_rate	Continuous
۱۹	num_access_files	Continuous	۴۰	dst_host_rerror_rate	Continuous
۲۰	num_outbound_cmds	Continuous	۴۱	dst_host_srv_rerror_rate	Continuous
۲۱	is_host_login	Continuous			

۳-۲- پاک‌سازی داده‌ها

در بسیاری از کاربردهای دنیای واقعی کاوش داده‌ها، حتی باوجود مقدار داده‌های حجیم و فضای ذخیره‌سازی مناسب، ممکن است در نمونه‌های موجود، مقادیری از داده‌ها ازدست‌رفته (گمشده) باشند. اما مشکل ازآنجا آغاز می‌شود که برای مجموعه

داده‌های بزرگ نمی‌توان از مقادیر ازدست‌رفته چشم‌پوشی کرد. یک راه‌حل برای جایگزینی و پاک‌سازی مقادیر از دست رفته با مقادیر ثابت است در این تحقیق از مقدار میانگین برای مقادیر از دست رفته در هر ویژگی استفاده‌شده است. به‌عبارت‌دیگر براساس مقادیر موجود برای هر ویژگی میانگین محاسبه شده و در نمونه‌های فاقد مقدار جایگزین می‌شود[۴].

۳-۳- نرمال‌سازی داده‌ها

باتوجه به یکسان نبودن بازه تغییرات ویژگی‌ها و هم‌چنین واحدهای متفاوت متغیرها، مقادیر بزرگ‌تر تأثیر بیش‌تری بر توابع مورد استفاده دارند که لزوماً به معنی مهم‌تر بودن آن‌ها نیست. برای رفع این مشکل نرمال‌سازی داده‌ها انجام می‌شود. با توجه به اینکه در دیتاست مورد استفاده داده‌های نویزی و نامربوط پاک‌سازی شده‌اند و در این تحقیق با نرمال‌سازی خطی طبق رابطه ۱، داده‌ها به بازه [۱و۱-] نرمال‌سازی شده‌اند[۱۲].

$$X = 2 \times \frac{x - \min(x)}{\max(x) - \min(x)} - 1 \qquad (۱)$$

که در آن min(x) کمینه بردار ورودی x و max(x) بیشینه بردار ورودی x بوده و X نرمال شده آن است[۱۲].

۳-۴- انتخاب ویژگی

تحلیل تفکیکی فیشر یکی از انواع روش‌های کاهش ابعاد و انتخاب ویژگی است. یکی از کاربردهای مهم تحلیل تفکیکی فیشر، در طبقه‌بندی است. الگوریتم تحلیل تفکیکی فیشر داده‌ها را از فضای ورودی به فضایی جدید نگاشت می‌کند و ویژگی‌ها از بیشترین واریانس به کمترین واریانس مرتب می‌شوند و ویژگی‌هایی که واریانس بین کلاسی پایین و واریانس درون کلاسی بالایی دارند حذف می‌شوند و ویژگی‌های با واریانس بین‌کلاسی بالا و واریانس درون کلاسی پایین انتخاب می‌شوند. در این تحقیق با استفاده از الگوریتم تحلیل تفکیکی فیشر از میان ۴۱ ویژگی موجود در پایگاه داده تعداد ۵ ویژگی به علت واریانس بین‌کلاسی بالا و واریانس درون کلاسی پایین برای طبقه‌بندی دقیق‌تر انتخاب شده‌اند. ویژگی‌های منتخب توسط الگوریتم تحلیل تفکیکی فیشر ویژگی‌های ۱-نوع پروتکل ارتباطی(protocol_type)، ۲-تعداد دستورات مبادله شده (num_shell)، ۳-تعداد دستورات پروتکلFTP[۱۲] (num_outbound_cmds)، ۴-تعداد درگاه ارتباطی با سرویس تکراری در ۲ ثانیه گذشته(srv_count) و ۵ درصد خطای ارتباطی سرویس خاص(srv_serror_rate).

۳-۵- ارائه مدل پیشنهادی

در استفاده از ماشین بردار پشتیبان پارامترهایی مانند ضریب جریمه، تابع هسته، درجه چندجمله‌ای، شعاع پراکندگی در تابع گاوسی و الگوریتم آموزش ماشین بردار پشتیبان با استفاده از سعی خطا تعیین می‌شوند. به‌عبارت‌دیگر کاربر زمان زیادی را صرف تنظیم و تعیین مقادیر بهینه پارامترهای ذکرشده می‌نماید زیرا مقادیر بهینه آن‌ها تأثیر به‌سزایی در نتایج استفاده از ماشین بردار پشتیبان دارد که باعث پیچیدگی زمانی بالا در استفاده از ماشین بردار پشتیبان می‌شود در نتیجه ایده این تحقیق تعیین مقادیر بهینه پارامترهای ذکرشده(ضریب جریمه، تابع هسته، درجه چندجمله‌ای، شعاع پراکندگی در تابع گاوسی و الگوریتم آموزش ماشین بردار پشتیبان) با استفاده از الگوریتم بهینه‌سازی ملخ است. نتایج شبیه‌سازی‌ها نشان می‌دهد که الگوریتم ملخ باوجود سادگی زیاد قادر به ارائه نتایج برتر در مقایسه با الگوریتم‌های شناخته شده و اخیر در مسائل بهینه‌سازی است. نتایج شبیه‌سازی بر روی مسائل واقعی نیز ثابت کرد که الگوریتم ملخ قادر به حل مسائل واقعی با فضای ناشناخته است.همچنین این الگوریتم فاقد نیاز به تنظیم پارامترهای مورد استفاده توسط کاربر است[۹].

پارامترهای مذکور در ماشین بردار پشتیبان می‌توانند دارای مقادیر زیر باشند:

[12] File Transfer Protocol

۱) ضریب جریمه در ماشین بردار پشتیبان: این ضریب دارای مقادیر مختلف در بازه[۱۰۰۰ – ۰/۰۰۱] است.
۲) تابع هسته در ماشین بردار پشتیبان: توابع هسته می‌تواند به‌صورت خطی، دوجمله‌ای، چندجمله‌ای و گاوسی باشد.
۳) درجه چندجمله‌ای در ماشین بردار پشتیبان: درجه چندجمله‌ای می‌تواند دارای مقادیر ۳، ۴ و ۵ باشد.
۴) شعاع پراکندگی در تابع گاوسی در ماشین بردار پشتیبان: شعاع پراکندگی می‌تواند دارای مقادیر در بازه[۱-۰] باشد.
۵) نوع الگوریتم آموزشی در ماشین بردار پشتیبان: نوع الگوریتم می‌تواند الگوریتم‌های SMO[13]، ISDA[14] و غیره باشد.

۳-۵-۱- ساختار موقعیت ملخ در الگوریتم بهینه‌سازی ملخ

در الگوریتم بهینه‌سازی ملخ موقعیت ملخ از پنج مؤلفه تشکیل شده است که پارامترهای ماشین بردار پشتیبان است و شامل: ضریب جریمه، تابع هسته، درجه چندجمله‌ای، شعاع پراکندگی تابع گاوسی و الگوریتم آموزش ماشین بردار پشتیبان می‌شود که بایستی به‌صورت بهینه این پارامترها توسط الگوریتم بهینه‌سازی ملخ تعیین شوند و ماشین بردار پشتیبان با مقادیر بهینه این پارامترها آموزش داده شود تا تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا را با دقت بالایی انجام دهد بنابراین شکل ۲ ساختار موقعیت ملخ در الگوریتم بهینه‌سازی ملخ را نشان می‌دهد که شامل پارامترهای ماشین بردار پشتیبان است.

نوع الگوریتم آموزش	درجه چندجمله‌ای	شعاع پراکندگی	تابع هسته	ضریب جریمه

شکل ۲: ساختار موقعیت ملخ در الگوریتم بهینه‌سازی ملخ

۳-۵-۲- تابع هدف در الگوریتم بهینه‌سازی ملخ

موقعیت هر ملخ در الگوریتم بهینه‌سازی ملخ شامل موارد تعیین‌شده در شکل ۲ است، بنابراین پس از مقداردهی، تابع هدف موقعیت ملخ و پایگاه داده ورودی آموزشی را دریافت می‌کند و ماشین بردار پشتیبان با موارد تعیین‌شده در موقعیت ملخ و پایگاه داده ورودی آموزشی، آموزش داده و خروجی آموزشی تعیین می‌شود سپس میانگین مربعات خطا پایگاه داده خروجی آموزشی هدف و مدل محاسبه می‌شود و به‌عنوان مقدار تابع هدف در نظر گرفته می‌شود. در رابطه ۲ مقدار t_i مقدار خروجی هدف آموزشی و y_i مقدار ماشین بردار پشتیبان است.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(t_i - y_i)^2 \qquad (۲)$$

۳-۵-۳- مراحل الگوریتم بهینه‌سازی ملخ

مراحل استفاده از الگوریتم بهینه‌سازی ملخ شامل دو مرحله اصلی است. ۱) مرحله آماده‌سازی اولیه ۲) مرحله تکرار

مرحله آماده‌سازی اولیه

در این مرحله یک جمعیت از ملخ‌ها ایجاد می‌شود و مورد ارزیابی توسط تابع هدف قرار می‌گیرد. شکل ۳ یک نمونه از موقعیت ملخ در الگوریتم بهینه‌سازی ملخ را نشان می‌دهد. پس از مقداردهی موقعیت ملخ تابع هدف فراخوانی می‌شود و موقعیت ملخ و پایگاه داده ورودی آموزشی را دریافت می‌کند سپس براساس معیارهای مشخص شده در موقعیت ملخ ماشین بردار پشتیبان ایجاد و با پایگاه داده ورودی آموزشی، آموزش داده می‌شود و میانگین مربعات خطای پایگاه داده خروجی آموزشی هدف و مدل محاسبه شده و به‌عنوان مقدار تابع هدف در نظر گرفته می‌شود.

[13] Sequential minimal optimization
[14] Implicit semantic data augmentation

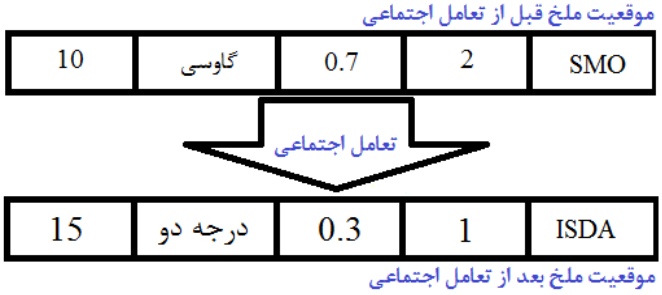
10	گاوسی	0.7	2	SMO

شکل ۳: موقعیت ملخ در الگوریتم بهینه‌سازی ملخ

مرحله تکرار

هر یک از گام‌های زیر برای هر ملخ مانند i تکرار می‌شود.

به‌روزرسانی موقعیت هر ملخ: در این گام موقعیت ملخ i بر اساس موقعیت ملخ‌های دیگر و موقعیت بهترین ملخ با استفاده از عملگر تعامل اجتماعی تعیین می‌شود. در شکل ۴ موقعیت ملخ i قبل و بعد از تعامل اجتماعی را نشان می‌دهد که نتیجه آن ایجاد مقادیر جدیدی برای ماشین بردار پشتیبان است.



شکل ۴: موقعیت جدید ملخ قبل و بعد از تعامل اجتماعی در الگوریتم بهینه‌سازی ملخ

مدل ریاضی بکارگرفته شده برای شبیه‌سازی رفتار ملخ‌ها به‌صورت رابطه زیر است.

$$X_i = S_i + G_i + A_i \quad (۳)$$

که در آن X_i موقعیت iامین ملخ، S_i تعامل اجتماعی و G_i نیروی گرانش اعمال شده به ملخ iام و A_i جهت باد را نمایش می‌دهد. به‌منظور ایجاد رفتار تصادفی می‌توان رابطه ۳ را به‌صورت رابطه ۴ بازنویسی کرد که در آن r_1، r_2 و r_3 اعدادی تصادفی در بازه [۰و۱] هستند.

$$X_i = r_1 S_i + r_2 G_i + r_3 A_i \quad (۴)$$

مقدارS_i یعنی تعامل اجتماعی برای ملخ iام با توجه به رابطه ۵ محاسبه می‌شود. در واقع تابع S قادر است تا فضای بین دو ملخ را به نواحی دافعه و جاذبه و آسایش تقسیم کند.

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} s(d_{ij})\widehat{d_{ij}} \quad (۵)$$

که در آن d_{ij} فاصله بین ملخ iام با ملخ jام را نشان می‌دهد و به‌صورت رابطه ۶ محاسبه می‌شود.

$$d_{ij} = |x_j - x_i| \quad (۶)$$

و بردار d_{ij} یک بردار واحد از iامین ملخ به jامین ملخ است طبق رابطه ۷ محاسبه می‌شود.

$$\widehat{d_{ij}} = \frac{x_j - x_i}{d_{ij}} \quad (۷)$$

تابع $S(r_{ij})$ نیروی اجتماعی را تعریف می‌کند و توسط رابطه ۸ محاسبه می‌شود. فاصله بین ملخ‌ها در بازه [۴و۰] در نظر گرفته می‌شود. در بازه (۲/۰۷۵و۰] نیروی اجتماعی بین ملخ‌ها به‌صورت دافعه و در بازه [۴ و ۲/۰۷۵) نیروی اجتماعی بین ملخ‌ها به‌صورت جاذبه خواهد بود. فاصله ۲/۰۷۵ نیز نقطه آسایش محسوب می‌شود.

$$s(rij) = f e^{\frac{-r}{l}} - e^{-r} \qquad (۸)$$

که در آن f نشان دهنده شدت جاذبه و l نشان‌دهنده طول مقیاس جاذبه است. مقدار l=۱/۵ و مقدار f=۰/۵ است. r_{ij} نیز از رابطه ۹ به دست می‌آید.[۹].

$$r_{ij}=2+rem(d_{ij},2) \qquad (۹)$$

حال برای به دست آوردن موقعیت بعدی هر ملخ از رابطه ۱۰ استفاده می‌شود.

$$X_i^d(t+1) = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\frac{ub_d - lb_d}{2} * \frac{s(r_{ij})}{d_{ij}}\left(\left|x_j^d - x_i^d\right|\right)\right) + \hat{T}^d \qquad \forall j \qquad (۱۰)$$

که در آن ub_d حد بالا در بعد dام، lb_d حد پایین در بعد dام و بردار T^d مقدار بهترین راه‌حل در بعد dام است و c یک ثابت کاهش برای کم کردن منطقه آسایش، دافعه و جاذبه است و از رابطه ۱۱ به دست می‌آید. به‌منظور حفظ تعادل بین اکتشاف و بهره‌برداری، این پارامتر(c) نیاز است .که با افزایش دفعات تکرار و در حین الگوریتم کاهش می‌یابد. این مکانیسم اکتشاف را با افزایش دفعات تکرار تقویت می‌کند. ضریب منطقه آسایش متناسب با تعداد تکرارها را کاهش می‌دهد و به‌صورت رابطه ۱۱ محاسبه می‌شود که در آن c_{max} بیشترین مقدار، c_{min} کمترین مقدار، l شمارنده تکرار جاری و L حداکثر تعداد دفعات تکرار الگوریتم است. در این تحقیق مقدار c_{max} برابر ۱ و c_{min} برابر ۰/۰۰۰۰۱ در نظر گرفته شده است.

$$c = c_{max} - l\frac{c_{max} - c_{min}}{L} \qquad (۱۱)$$

رابطه ۱۰ نشان می‌دهد که موقعیت بعدی یک ملخ بر اساس موقعیت فعلی آن، موقعیت بهترین ملخ و موقعیت همه ملخ‌های دیگر تعریف می‌شود. همچنین شایان‌ذکر است که پارامتر c دو مرتبه در رابطه ۱۰ به دلایل زیر استفاده‌شده است:
اولین c از چپ خیلی شبیه به وزن داخلی w در الگوریتم PSO[۱۵] است و جابجایی‌های ملخ در اطراف هدف را کاهش می‌دهد. به‌عبارت‌دیگر این پارامتر تعادل بین اکتشاف و بهره‌برداری در اطراف هدف را ایجاد می‌کند.
پارامتر c داخلی در هر بار تکرار منجر به کاهش نیروی جاذبه، دافعه بین ملخ‌ها می‌شود تا به نقطه آسایش برسند و الگوریتم همگرا شود.

ارزیابی موقعیت جدید ملخ: در این گام موقعیت جدید ملخ پس از تولید مورد ارزیابی توسط تابع هدف قرار می‌گیرد. یعنی موقعیت جدید ملخ به همراه پایگاه داده ورودی آموزشی به تابع هدف ارسال می‌شود و ماشین بردار پشتیبان بر اساس آن ایجاد و آموزش داده می‌شود و خطای میانگین مربعات پایگاه داده خروجی آموزشی هدف و مدل محاسبه می‌شود و به‌عنوان مقدار تابع هدف برگشت داده می‌شود.

به‌روزرسانی ضریب کاهشی: در این گام پارامتر ضریب کاهشی C به‌صورت خطی به‌روزرسانی می‌شود.

تکرار عملیات از مرحله تکرار تا برآورده شدن شرایط خاتمه تکرار می‌شوند. خروجی الگوریتم بهینه‌سازی ملخ عضوی از جمعیت است که خطای میانگین مربعات از بقیه اعضای جمعیت کمتر است و مؤلفه‌های آن بهترین معیار ضریب جریمه، تابع هسته، درجه چندجمله‌ای، شعاع پراکندگی تابع گاوسی و الگوریتم آموزش در ماشین بردار پشتیبان را نشان می‌دهد.

[15] Particle swarm optimization

۳-۶- جمع‌بندی

در این بخش کلیه مراحل تحقیق بیان شد و در مرحله پیش‌پردازش ابتدا پاک‌سازی داده‌ها با استفاده از آماره میانگین انجام شد و با استفاده از نرمال‌سازی خطی داده‌ها به بازه [۱و۱-] منتقل شدند و با استفاده از الگوریتم تحلیل تفکیکی فیشر عمل انتخاب ویژگی انجام شد و تعداد ۵ ویژگی انتخاب و در مرحله پس‌پردازش بهبود ماشین بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی ملخ انجام شد.

۴-نتایج شبیه‌سازی

در این بخش به نتایج حاصل از سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا توسط ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی ملخ به ازای داده‌های آموزشی و آزمایشی و کل داده‌ها می‌پردازیم. در الگوریتم بهینه‌سازی ملخ، اندازه جمعیت برابر ۱۵ و حداکثر دفعات تکرار الگوریتم برابر ۲۰ در نظر گرفته شده است. نتایج توسط معیارهای کارایی استاندارد باهم مقایسه شدند که مؤلفه‌ها به‌صورت زیر است.

مثبت صحیح[۱۶] (TP): تعداد نمونه‌هایی است که در طبقه حمله قرار دارد و در طبقه‌بندی به‌درستی حمله تشخیص داده شده است.

منفی ناصحیح[۱۷] (FN): تعداد نمونه‌هایی است که در طبقه حمله قرار دارد اما به نادرستی توسط طبقه‌بند غیر حمله تشخیص داده شده است.

منفی صحیح[۱۸] (TN): تعداد نمونه‌هایی است که در طبقه غیر حمله قرار دارد و در طبقه‌بندی به‌درستی غیر حمله تشخیص داده شده است.

مثبت ناصحیح[۱۹] (FP): تعداد نمونه‌هایی است که در طبقه غیر حمله قرار دارد اما به نادرستی توسط طبقه‌بند حمله تشخیص داده شده است.

حساسیت[۲۰](S_n): درصد نمونه‌های حمله را که به‌درستی به‌عنوان نمونه‌های حمله طبقه‌بندی شده‌اند را نشان می‌دهد. در واقع معیار حساسیت نشان‌دهنده درصد دقت تشخیص صحیح در نمونه‌های حمله است.

ویژگی[۲۱](S_p): درصد نمونه‌های غیر حمله را که به‌درستی به‌عنوان نمونه غیر حمله طبقه‌بندی شده‌اند را نشان می‌دهد. در واقع معیار ویژگی نشان‌دهنده درصد دقت تشخیص صحیح در نمونه‌های غیر حمله است.

مقدار پیش‌گویی مثبت[۲۲](PPV): درصد نمونه‌هایی که حمله طبقه‌بندی شده‌اند به کل نمونه‌های حمله طبقه‌بندی شده توسط طبقه‌بند را نشان می‌دهد.

مقدار پیش‌گویی منفی[۲۳](NPV): درصد نمونه‌هایی که غیر حمله طبقه‌بندی شده‌اند به کل نمونه‌های غیر حمله طبقه‌بندی شده توسط طبقه‌بند را نشان می‌دهد.

دقت[۲۴](P): درصد نمونه‌هایی که به‌درستی طبقه‌بندی شده‌اند به کل نمونه‌هایی که درست یا نادرست طبقه‌بندی شده را نشان می‌دهد[۷].

---

[16] True Positive=TP
[17] False Negative=FN
[18] True Negative=TN
[19] False Positive=FP
[20] Sensitivity=Sn
[21] Specificity=Sp
[22] Positive Predictive Value=PPV
[23] Negative Predictive Value=NPV
[24] Precision=P

در جدول ۲ تحلیل آماری خطا در سیستم تشخیص نفوذ در اینترنت اشیا توسط ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی ملخ به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را نشان می‌دهد. مقدار دقت در داده‌های آموزشی برابر ۹۷٪، در داده‌های آزمایشی برابر ۹۸٪ و در کل داده‌ها برابر ۹۷/۲٪ است.

جدول ۲: تحلیل آماری خطا در SVM بهبودیافته با الگوریتم GOA در تشخیص نفوذ در اینترنت اشیا

Data	Sn	Sp	PPV	NPV	P
Train Data	۹۵/۸	۹۸/۲	۹۸/۲	۹۵/۸	۹۷
Test Data	۹۸	۹۸	۹۸	۹۸	۹۸
All Data	۹۶/۳	۹۸/۲	۹۸/۲	۹۶/۲	۹۷/۲

شکل ۵ نمودار رگرسیون به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را در ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی ملخ را نشان می‌دهد. ضریب رگرسیون به ازای داده‌های آموزشی برابر ۰/۹۴، به ازای داده‌های آزمایشی برابر ۰/۹۶و به ازای کل داده‌ها برابر ۰/۹۴۴ است. هرچه مقدار ضریب رگرسیون به یک نزدیک‌تر باشد همبستگی بین خروجی‌های هدف و خروجی‌های مدل بیشتر و خطا در تشخیص کم‌تر است.

شکل ۵: نمودار رگرسیون به ازای انواع داده‌ها در SVM بهبودیافته با الگوریتم بهینه‌سازی ملخ

۴-۱- استفاده از طبقه‌بند بگینگ در تشخیص نفوذ در اینترنت اشیا

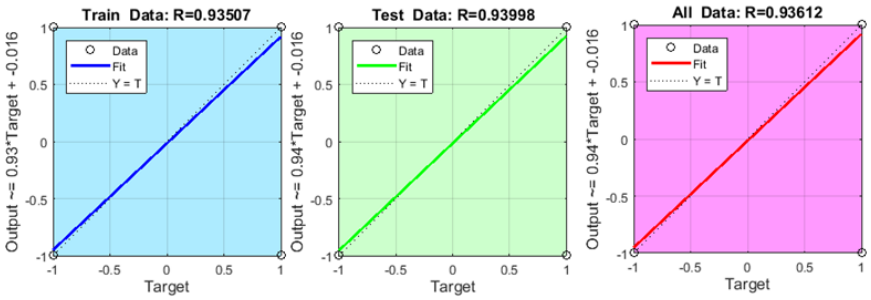
در استفاده از طبقه‌بند بگینگ تعداد طبقه‌بندها برابر ۵ و از نوع درخت تصمیم درنظرگرفته شده است تا سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا را انجام دهیم. پس از انجام شبیه‌سازی‌ها در محیط متلب نتایج به‌صورت زیر مشخص گردید.

در جدول ۳ تحلیل آماری خطا در سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا توسط طبقه‌بند بگینگ به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را نشان می‌دهد.

جدول ۳: تحلیل آماری خطا در طبقه‌بند بگینگ در تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا

Data	Sn	Sp	PPV	NPV	P
Train Data	۹۶/۱	۹۷/۴	۹۷/۵	۹۵/۹	۹۶/۸
Test Data	۹۵/۸	۹۸/۱	۹۷/۸	۹۶/۳	۹۷
All Data	۹۶/۱	۹۷/۶	۹۷/۶	۹۶	۹۶/۸

شکل ۶ نمودار رگرسیون به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را در طبقه‌بند بگینگ را نشان می‌دهد. ضریب رگرسیون به ازای داده‌های آموزشی برابر ۰/۹۲۵، به ازای داده‌های آزمایشی برابر ۰/۸۷۷ و به ازای کل داده‌ها برابر ۰/۹۱۵ است. هرچه مقدار ضریب رگرسیون به یک نزدیک‌تر باشد همبستگی بین خروجی‌های هدف و خروجی‌های مدل بیشتر و خطا در تشخیص کم‌تر است.



شکل ۶: نمودار رگرسیون به ازای انواع داده‌ها در طبقه‌بند بگینگ در تشخیص نفوذ در اینترنت اشیا

۴-۲- استفاده از طبقه‌بند k-نزدیک‌ترین همسایه در تشخیص نفوذ در اینترنت اشیا

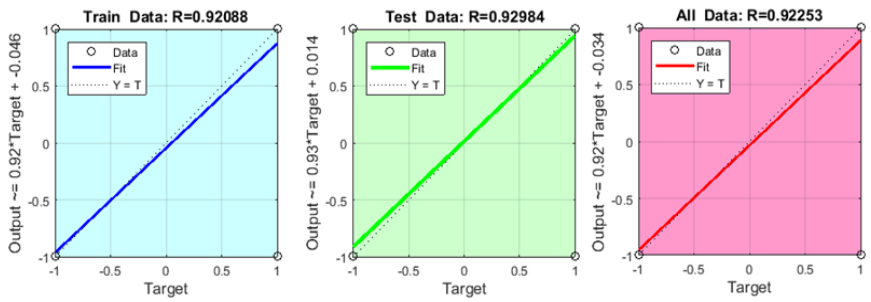
در استفاده از طبقه‌بند k-نزدیک‌ترین همسایه تعداد طبقه‌ها برابر دو درنظرگرفته شده است تا سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا را انجام دهد.

در جدول ۴ تحلیل آماری خطا در سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا توسط طبقه‌بند k-نزدیک‌ترین همسایه به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را نشان می‌دهد.

جدول ۴: تحلیل آماری خطا در طبقه‌بند k-نزدیک‌ترین همسایه در تشخیص نفوذ در اینترنت اشیا

Data	Sn	Sp	PPV	NPV	P
Train Data	۹۴/۱	۹۸/۱	۹۸/۳	۹۳/۷	۹۶
Test Data	۹۶/۸	۹۶/۲	۹۵/۸	۹۷/۱	۹۶/۵
All Data	۹۴/۶	۹۷/۷	۹۷/۸	۹۴/۴	۹۶/۱

شکل ۷ نمودار رگرسیون به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را در طبقه‌بند k-نزدیک‌ترین همسایه را نشان می‌دهد. ضریب رگرسیون به ازای داده‌های آموزشی برابر ۰/۸۴۲، به ازای داده‌های آزمایشی برابر ۰/۸۶ و به ازای کل داده‌ها برابر ۰/۸۴۶ است. هرچه مقدار ضریب رگرسیون به یک نزدیک‌تر باشد همبستگی بین خروجی‌های هدف و خروجی‌های مدل بیشتر و خطا در تشخیص کم‌تر است.



شکل ۷: نمودار رگرسیون به ازای انواع داده‌ها در طبقه‌بند k-نزدیک‌ترین همسایه در تشخیص نفوذ در اینترنت اشیا

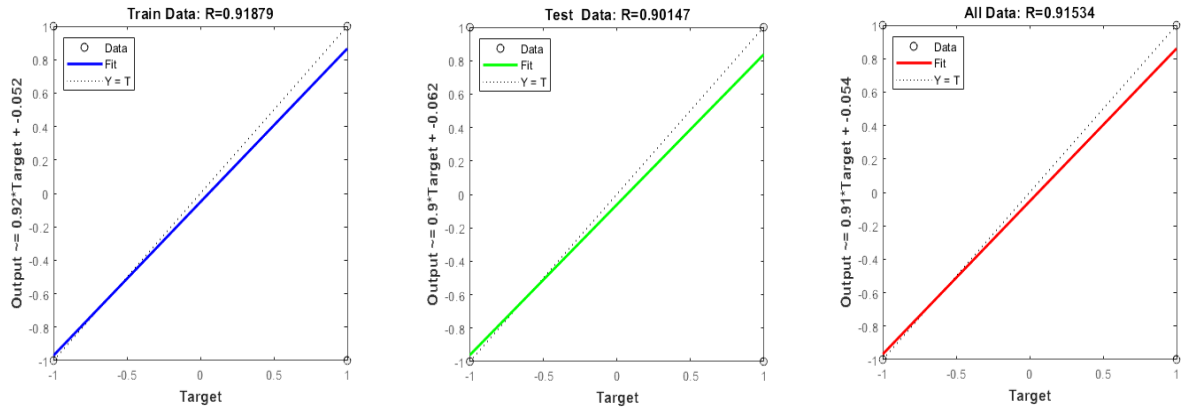
۴-۳- استفاده از طبقه‌بند ماشین بردار پشتیبان پایه در تشخیص نفوذ در اینترنت اشیا

در جدول ۵ تحلیل آماری خطا در سیستم تشخیص نفوذ مبتنی بر ناهنجاری در اینترنت اشیا توسط طبقه‌بند ماشین بردار پشتیبان پایه به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را نشان می‌دهد.

جدول ۵: تحلیل آماری خطا در طبقه‌بند ماشین بردار پشتیبان پایه در تشخیص نفوذ در اینترنت اشیا

Data	Sn	Sp	PPV	NPV	P
Train Data	۹۳/۶	۹۸/۴	۹۸/۵	۹۳/۳	۹۵/۹
Test Data	۹۲/۶	۹۷/۸	۹۸	۹۱/۸	۹۵
All Data	۹۳/۴	۹۸/۳	۹۸/۴	۹۳	۹۵/۷

شکل ۸ نمودار رگرسیون به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها را در طبقه‌بند ماشین بردار پشتیبان پایه را نشان می‌دهد. ضریب رگرسیون به ازای داده‌های آموزشی برابر ۰/۹۱۸، به ازای داده‌های آزمایشی برابر ۰/۹۰۱ و به ازای کل داده‌ها برابر ۰/۹۱ است. هرچه مقدار ضریب رگرسیون به یک نزدیک‌تر باشد همبستگی بین خروجی‌های هدف و خروجی‌های مدل بیشتر و خطا در تشخیص کم‌تر است.

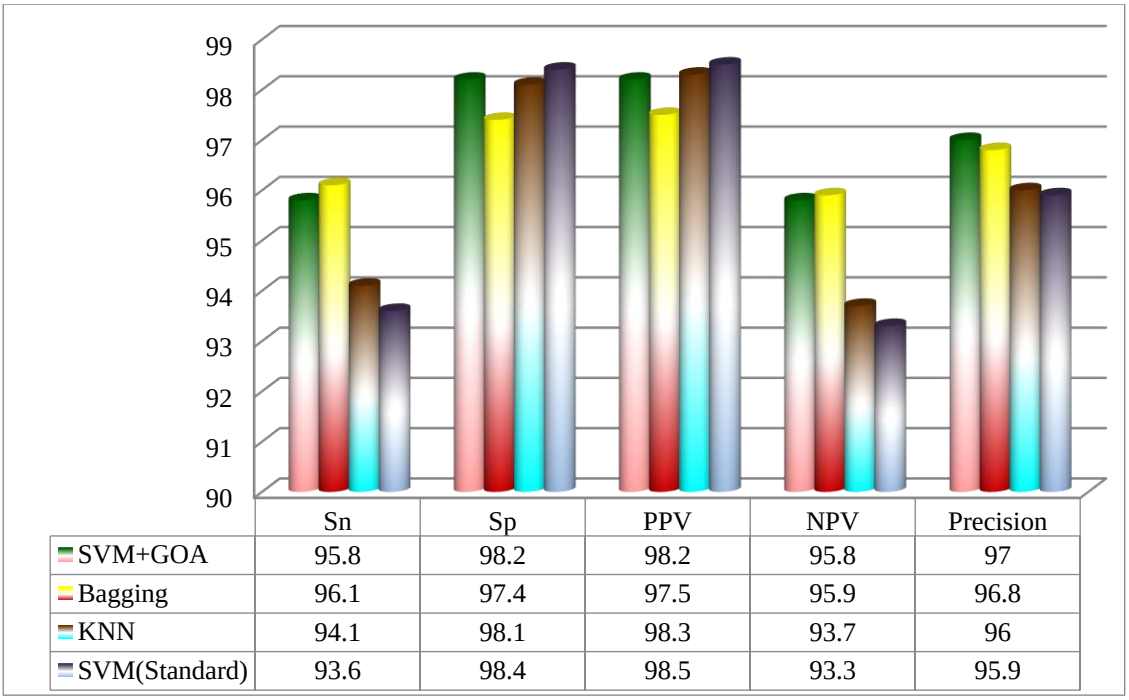


شکل ۸: نمودار رگرسیون به ازای انواع داده‌ها در طبقه‌بند ماشین بردار پشتیبان پایه در تشخیص نفوذ در اینترنت اشیا

۵-ارزیابی و نتیجه‌گیری

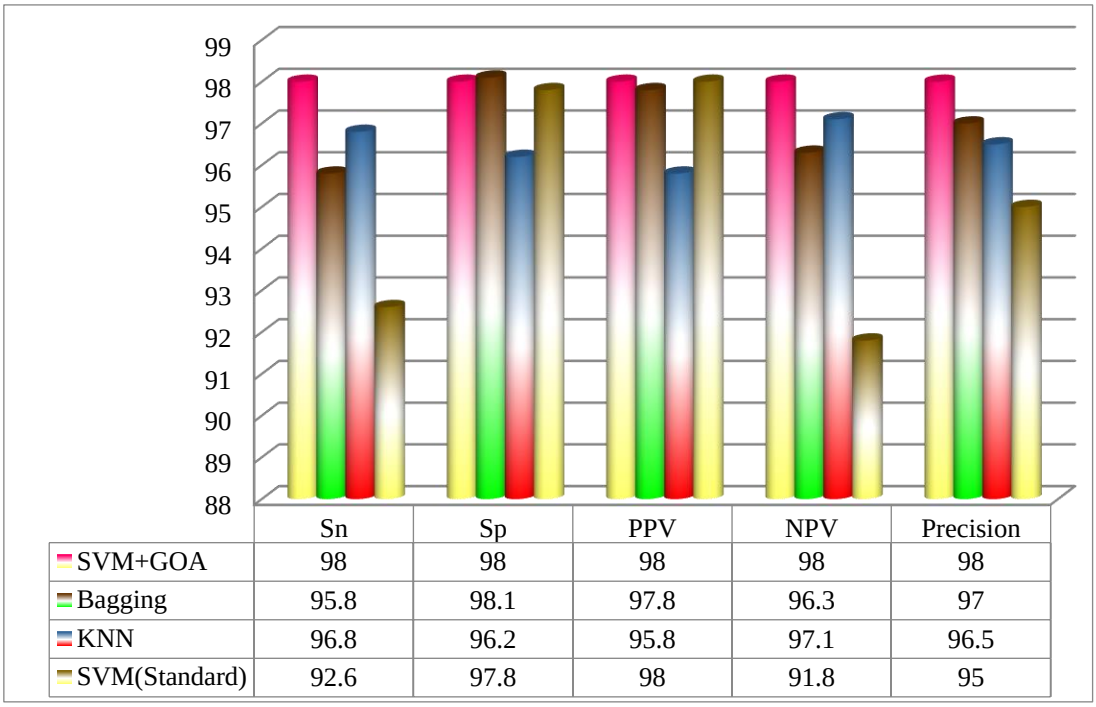
در این پژوهش ابتدا به جمع‌آوری داده‌های سیستم تشخیص نفوذ در اینترنت اشیا پرداخته و پس از پاک‌سازی داده‌ها با استفاده از آماره مرکزی میانگین، نرمال‌سازی خطی داده‌ها انجام و با الگوریتم تحلیل تفکیکی فیشر انتخاب ویژگی انجام شد و تعداد ۵ ویژگی از ۴۱ ویژگی انتخاب شد. سپس ماشین بردار پشتیبان با استفاده از الگوریتم بهینه‌سازی ملخ بهبود یافت و نتایج با استفاده از طبقه‌بند بگینگ و طبقه‌بند k-نزدیک‌ترین همسایه مقایسه و نتایج زیر حاصل شده است:

باتوجه به شکل‌های ۹، ۱۰ و ۱۱ ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی ملخ به ترتیب به ازای داده‌های آموزشی، آزمایشی و کل داده‌ها عملکرد بهتری از لحاظ انواع تحلیل آماری خطا نسبت به طبقه‌بندهای بگینگ و k-نزدیک‌ترین همسایه دارد. طبقه‌بند بگینگ عملکرد بهتری در مقایسه با طبقه‌بند k-نزدیک‌ترین همسایه دارد.



	Sn	Sp	PPV	NPV	Precision
SVM+GOA	95.8	98.2	98.2	95.8	97
Bagging	96.1	97.4	97.5	95.9	96.8
KNN	94.1	98.1	98.3	93.7	96
SVM(Standard)	93.6	98.4	98.5	93.3	95.9

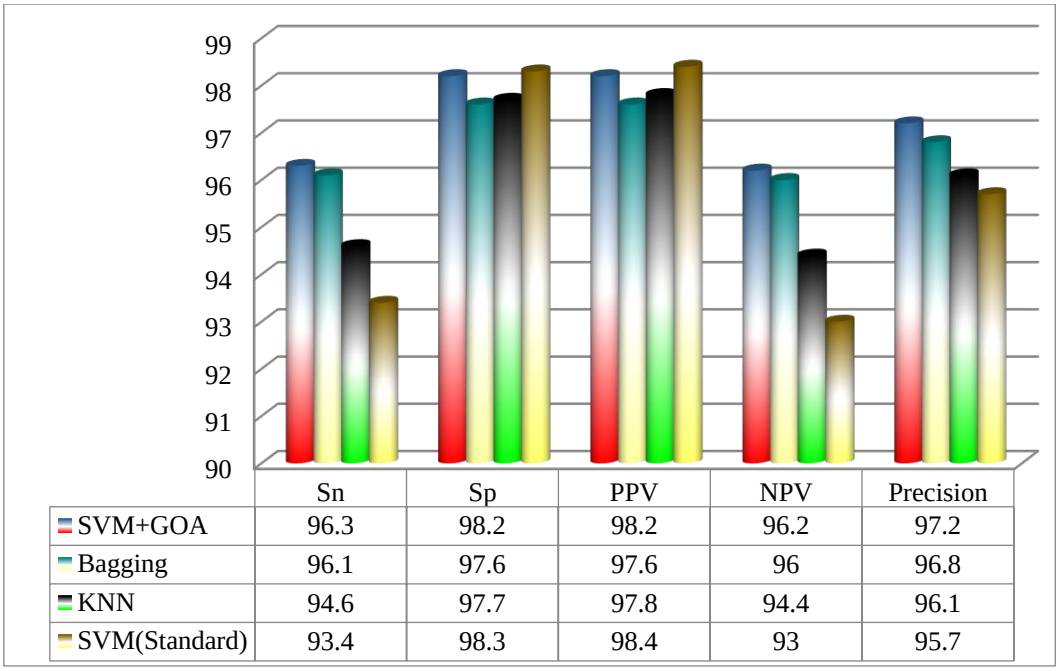
شکل ۹: عملکرد الگوریتم‌های مختلف براساس تحلیل آماری خطا به ازای داده‌های آموزشی



	Sn	Sp	PPV	NPV	Precision
SVM+GOA	98	98	98	98	98
Bagging	95.8	98.1	97.8	96.3	97
KNN	96.8	96.2	95.8	97.1	96.5
SVM(Standard)	92.6	97.8	98	91.8	95

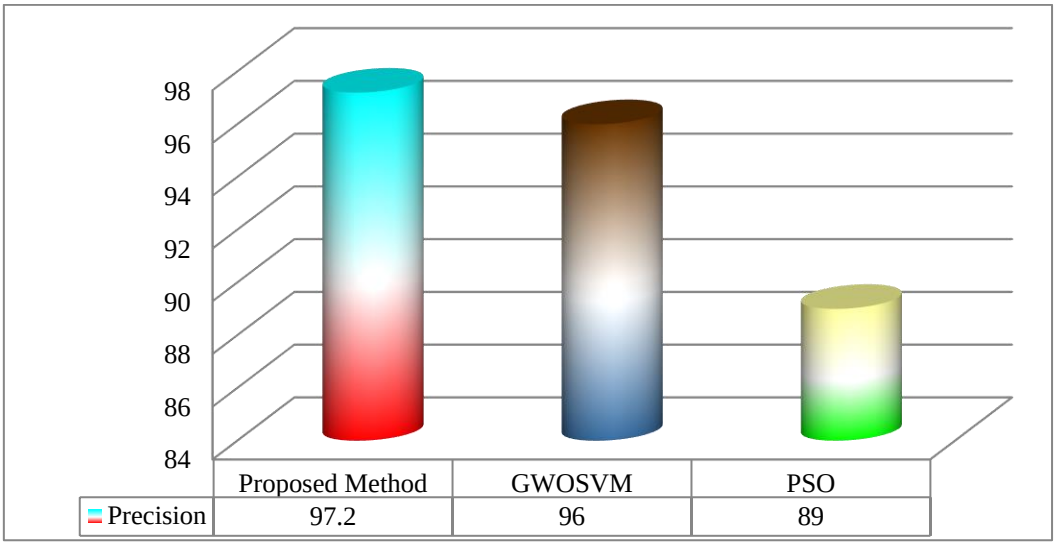
شکل ۱۰: عملکرد الگوریتم‌های مختلف براساس تحلیل آماری خطا به ازای داده‌های آزمایشی

با توجه به شکل ۱۲ روش پیشنهادی بر اساس دقت با ماشین بردار پشتیبان بهبودیافته با الگوریتم بهینه‌سازی گرگ خاکستری و بهینه‌سازی ازدحام ذرات مقایسه شده است. نتایج حاکی از عملکرد بهتر روش پیشنهادی دارد[۶].



	Sn	Sp	PPV	NPV	Precision
SVM+GOA	96.3	98.2	98.2	96.2	97.2
Bagging	96.1	97.6	97.6	96	96.8
KNN	94.6	97.7	97.8	94.4	96.1
SVM(Standard)	93.4	98.3	98.4	93	95.7

شکل ۱۱: عملکرد الگوریتم‌های مختلف براساس تحلیل آماری خطا به ازای کل داده‌ها



	Proposed Method	GWOSVM	PSO
Precision	97.2	96	89

شکل ۱۲: مقایسه روش پیشنهادی بر اساس دقت با پژوهش های مشابه

پس از پایان شبیه‌سازی و مقایسه الگوریتم‌های پیشنهادی با روش‌های مشابه و مقالات دیگر مشخص گردید روش پیشنهادی باعث بهبود در دقت تشخیص و همچنین کاهش هشدار کاذب و افزایش بهره‌وری الگوریتم در شناسایی تهدیدات امنیتی به‌صورت تشخیص ناهنجاری بر روی دیتاست NSL-KDD شده است.همچنین به دلیل استفاده از الگوریتم‌های فاقد پیچیدگی زیاد و مصرف منابع پایین و سرعت بیشتر روش مناسبی جهت تشخیص حملات در شبکه‌های IOT است. همچنین می‌توان در پژوهش های آتی از موضوعات زیر نیز استفاده کرد.

- استفاده از ماشین بردار پشتیبان در سیستم تشخیص نفوذ به‌منظور رتبه‌بندی حملات و درجه‌بندی خطرات
- استفاده از الگوریتم بهینه‌سازی ملخ در کاهش ابعاد مسئله در سیستم تشخیص نفوذ در اینترنت اشیا

- استفاده از الگوریتم‌های بهینه‌سازی هوشمند دیگر مانند الگوریتم بهینه‌سازی کفتار، بهینه‌سازی نهنگ، بهینه‌سازی کلونی پنگوئن و غیره برای بهبود ماشین بردار پشتیبان و بررسی دقت تشخیص نفوذ ماشین بردار پشتیبان و مقادیری که توسط الگوریتم‌ها برای پارامترها تعیین می‌شوند.
- استفاده از شبکه‌های عصبی دیگر نظیر شبکه عصبی روش گروهی مدل‌سازی داده برای تشخیص نفوذ در اینترنت اشیا و بررسی پیچیدگی زمان آموزش
- تشخیص نوع حمله و دسته‌بندی بر اساس ماهیت آن پس از تشخیص ناهنجاری

مراجع

[1] A. J. Siddiqui and A. Boukerche, "TempoCode-IoT: temporal codebook-based encoding of flow features for intrusion detection in Internet of Things," *Cluster Computing*, vol. 24, no. 1, pp. 17-35, 2021, doi: 10.1007/s10586-020-03153-8.

[2] A. Khraisat and A. Alazab, "A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges," *Cybersecurity*, vol. 4, no. 1, pp. 1-27, 2021, doi: 10.1186/s42400-021-00077-7.

[3] B. S. Khater, A. Wahab, A. W. Bin, M. Y. I. B. Idris, M. A. Hussain, and A. A. Ibrahim, "A lightweight perceptron-based intrusion detection system for fog computing," *Applied Sciences*, vol. 9, no. 1, p. 178, 2019, doi: 10.3390/app9010178.

[4] J. Bard, “What Is Data Mining?” PowerKids Press, 2018.

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

[6] M. Safaldin, M. Otair, and L. Abualigah, "Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks," *Journal of ambient intelligence and humanized computing*, vol. 12, no. 2, pp. 1559-1576, 2021, doi: 10.1007/s12652-020-02228-z.

[7] N. Huber, S. R. Kalidindi, B. Klusemann, and C. J. Cyron, ”Machine Learning and Data Mining in Materials Science,” Frontiers Media SA, 2020.

[8] N. Islam *et al.*, "Towards machine learning based intrusion detection in IoT networks," *Comput. Mater. Contin*, vol. 69, pp. 1801-1821, 2021, doi: 10.32604/cmc.2021.018466.

[9] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Advances in Engineering Software*, vol. 105, pp. 30-47, 2017, doi: 10.1016/j.advengsoft.2017.01.004.

[10] Sh. Ghafarian and K. Rezaei and A. Kafash, ” A survey on intrusion detection approaches in IOT”, Third National Conference on Applied Research in Electrical, Computer and Medical Engineering,2019

[11] V. Kumar, A. K. Das, and D. Sinha, "UIDS: a unified intrusion detection system for IoT environment," *Evolutionary Intelligence*, vol. 14, no. 1, pp. 47-59, 2021, doi: 10.1007/s12065-019-00291-w.

[12] X. S. Yang, “Introduction to Algorithms for Data Mining and Machine Learning”. Elsevier Science & Technology, 2019.