JIPET

Journal of Intelligent Procedures in Electrical Technology
Vol. 12/ No. 45/ Spring 2021
P-ISSN: 2322-3871, E-ISSN: 2345-5594, http://jipet.iaun.ac.ir/

20.1001.1.23223871.1400.12.1.6.2
Research Article

Improved Intrusion Detection System Based On Distributed Self-Adaptive Genetic Algorithm to Solve Support Vector Machine in Form of Multi Kernel Learning with Auto Encoder

Elahe Faghihnia[1], *Ph.D Student*, Seyed Reza Kamel Tabakh Farizani[2], *Assistant Professor*, Maryam Kheirabadi[1], *Assistant Professor*

[1]Department of Computer Engineering- Neyshabur Branch, Islamic Azad University, Neyshabur, IRAN
[2]Department of Computer Engineering- Mashhad Branch, Islamic Azad University, Mashhad, IRAN
faghihnia.elahe@gmail.com, DrKamel@mshdiau.ac.ir, maryam.abadi@gmail.com

Abstract
Intrusion into systems through network infrastructure and the Internet is one of the security challenges facing the world of information and communication technology and can lead to the destruction of systems and access to data and information. In this paper, a support vector machine model with weighted and parameters of SVM kernels are presented to detect the intrusion. Due to the high complexity of this problem, conventional optimization methods are not able to solve it. Therefore, we propose a Distributed Self Adaptive Genetic Algorithm (DSAGA). On the other hand, due to the high volume of data in such issues, Auto encoder has been used to reduce data. The proposed approach is a hybrid method based on Auto encoder, improved Support Vector Machine and Distributed Self Adaptive Genetic Algorithm (DSAGA) that it is evaluated by its execution on DARPA data set.

Keywords: big data, distributed self-adaptive genetic algorithm, intrusion detection systems, island genetic algorithm, self-adaptive genetic algorithm, support vector machine

Received: 15 October 2020
Revised: 1 December 2020
Accepted: 16 December 2020

Corresponding Author: Dr. Seyed Reza Kamel Tabakh Farizani

Citation: E. Faghihnia, S.R. Kamel-Tabakh-Farizani, M. Kheirabadi, "Improved intrusion detection system based on distributed self-adaptive genetic algorithm to solve support vector machine in form of multi kernel learning with auto encoder", Journal of Intelligent Procedures in Electrical Technology, vol. 12, no. 45, pp. 77-93, June 2021 (in Persian).

20.1001.1.23223871.1400.12.1.6.2
مقاله پژوهشی

سیستم تشخیص نفوذ بهبودیافته مبتنی بر الگوریتم ژنتیک خود تطبیق جزیره‌ای برای حل ماشین بردار پشتیبان به‌صورت یادگیری چندهسته‌ای با کد کننده‌های خودکار

الهه فقیه‌نیا[۱]، دانشجوی دکتری، سیدرضا کامل‌طباخ‌فریضنی[۲]، استادیار، مریم خیرآبادی[۱]، استادیار

۱- دانشکده مهندسی- واحد نیشابور، دانشگاه آزاد اسلامی، نیشابور، ایران
۲- دانشکده مهندسی- واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران
faghihnia.elahe@gmail.com, DrKamel@mshdiau.ac.ir, maryam.abadi@gmail.com

چکیده: نفوذ به سیستم‌ها از طریق زیرساخت شبکه و اینترنت یکی از چالش‌های امنیتی است که دنیای فناوری اطلاعات و ارتباطات را با آن روبرو کرده است و می‌تواند منجر به تخریب سیستم‌ها و دسترسی به داده‌ها و اطلاعات گردد. در این مقاله یک مدل ماشین بردار پشتیبان که هسته‌های آن وزن‌دار شده به همراه پارامترهای هسته‌های ماشین بردار پشتیبان برای سیستم تشخیص نفوذ ارائه شده است. با توجه به پیچیدگی محاسباتی این مدل، روش الگوریتم ژنتیک جزیره‌ای پویای خود تطبیقی پیشنهاد شده تا پیچیدگی محاسبات را کم نماید. در این روش از اتوانکودر نیز برای کاهش حجم داده‌ها استفاده شده است. روش پیشنهادی یک روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق است که دقت بهتری در مسائل تشخیص نفوذ را نشان می‌دهد. نتایج شبیه‌سازی بر روی مجموعه داده DARPA برای تست عملکرد مورد استفاده قرار گرفته است.

کلمات کلیدی: الگوریتم ژنتیک جزیره‌ای، الگوریتم ژنتیک خود تطبیق، سیستم‌های تشخیص نفوذ، کلان داده‌ها، ماشین بردار پشتیبان

تاریخ ارسال مقاله: ۱۳۹۹/۷/۲۴
تاریخ بازنگری مقاله: ۱۳۹۹/۹/۱۱
تاریخ پذیرش مقاله: ۱۳۹۹/۹/۲۶

نام نویسنده‌ی مسئول: دکتر سیدرضا کامل‌طباخ‌فریضنی
نشانی نویسنده‌ی مسئول: مشهد- قاسم آباد (شهرک غرب)- تقاطع استاد یوسفی- دانشگاه آزاد اسلامی مشهد- دانشکده مهندسی- گروه کامپیوتر

۱- مقدمه

ارائه و طراحی سیستم‌های محافظتی از سیستم‌های کامپیوتری در شبکه با تشخیص حملات موردتوجه جوامع دانشگاهی و صنعتی واقع شده است. این حملات می‌توانند عملکرد سیستم‌های امنیتی را با خطر مواجه سازند. تشخیص نفوذ یک راه مؤثر به‌منظور مقابله با این حملات بوده و شامل تعیین و تشخیص مجموعه‌ای از عملیات مغرضانه است که یکپارچگی، رازداری یا حالت محرمانه و دسترسی به منابع اطلاعاتی سیستم را تهدید می‌کند [۱]. چالش اصلی تمامی تحقیقات بهبود در دقت سیستم تشخیص نفوذ است. به‌طوریکه کمترین نرخ حملات اشتباه را داشته باشد. بنابراین پیاده سازی روشی که قادر باشد کمترین نرخ حملات اشتباه و دقت بالایی در تشخیص نفوذ داشته باشد گام مهمی در آخرین خط دفاعی امنیت سیستم‌های کامپیوتری می‌تواند داشته باشد. با توجه به ناکارآمدی روش‌های مبتنی بر امضا در تشخیص حملات جدید، روش‌های مبتنی بر نابهنجاری مورد توجه بیشتر واقع شده‌اند و تشخیص درست و سریع نفوذ در سیستم‌ها از اهمیت بالایی برخوردار است. روش‌های مبتنی بر نابهنجاری در سیستم‌های تشخیص نفوذ بر مسئله طبقه‌بندی داده‌ها استوار هستند. طبقه‌بندی داده‌ها یکی از فعالیت‌های مهم در فرآیند داده‌کاوی است و تحقیقات زیادی مبتنی بر داده کاوی در زمینه تشخیص نفوذ صورت گرفته است. روش‌های مبتنی بر ماشین بردار پشتیبان همراه مورد توجه محققان بوده و اخیرا نیز از آن استفاده کرده‌اند. در ادامه به تحلیل هر یک از دسته‌ها اشاره می‌گردد.

با توجه به سادگی روش‌های ماشین‌های بردار پشتیبان در عین حفظ دقت لازم این روش نسبت به روش‌های جدیدتر مانند ماشین‌های یادگیری توسعه یافته کماکان کاربرد بیشتری دارند. کارایی ماشین‌های بردار پشتیبان در تشخیص نفوذ اثبات شده است، زیرا ماشین‌های بردار پشتیبان، ماشین‌های یادگیری هستند که نقاط یادگیری را در یک فضا با ابعاد بالا نگاشت کرده و سپس هر بردار را به کلاسش برچسب‌گذاری می‌کنند و آن‌ها به تعداد نقاط داده حساس نیستند. بنابراین پیچیدگی طبقه‌بندی به ابعاد فضای مشخصه وابسته نیست به‌گونه‌ای که آن‌ها یادگیری مجموعه‌ای از الگوهای بزرگ با دقت بالا را دارند و می‌توانند بهتر از شبکه‌های عصبی عمل کنند؛ بنابراین، در بسیاری از تحقیقات صورت گرفته در زمینه تشخیص نفوذ، ماشین‌های بردار پشتیبان، مورد استفاده قرار گرفته است [۲]. در مرجع [۳] برای تشخیص نفوذ از ماشین بردار پشتیبان و الگوریتم پروانه‌ای استفاده کرده که در این روش با کاهش ابعاد داده و از بین بردن ویژگی‌های نامربوط ، عملکرد تشخیص نفوذ را بهبود می‌بخشد و زمان طبقه‌بندی را کاهش می‌دهد. در این روش دقت میزان هشدار کاذب کمی بهبود داده شده است. در مرجع [۴] برای به حداکثر رساندن میزان تشخیص و به حداقل رساندن میزان زنگ هشدار کاذب از ماشین بردار پشتیبان مبتنی بر الگوریتم هایپرکلیک بهبود یافته (HC-IBGSA SVM) و از خوشه‌بندی به‌منظور کم‌کردن محتوای پیش‌پردازش داده‌ها با حفظ کیفیت پیشنهاد شده است. این مقاله از یک تابع هدف وزنی و تعداد مطلوب ویژگی‌ها استفاده کرده و توانسته است کمی دقت تشخیص را افزایش دهد ولی پیچیدگی محاسباتی زیادی دارد. در مرجع [۵] برای تشخیص نفوذ از ماشین بردار پشتیبان و با بهینه‌سازی کلونی مورچه و برای کاهش ابعاد از pca[1] استفاده کرده است. محققان این مقاله با استفاده از روش پیشنهادی پارامترهای ماشین بردار پشتیبان را بهینه کرده و کمی کیفیت تشخیص نفوذ را بهبود داده‌اند. از نقاط ضعف این روش حساس بودن ماشین بردار پشتیبان به انتخاب هسته‌ها است.

روش مدل مخفی مارکوف به‌عنوان ابزاری برای تشخیص نفوذ در مقاله [۶] و استفاده از قوانین وابستگی که یک ابزار داده‌کاوی است در تشخیص نفوذ در مقاله [۷] پیشنهاد شده است.

بررسی عملکرد روش‌های خوشه‌بندی مختلف برای یک مجموعه داده بزرگ و تجزیه و تحلیل عوامل اصلی، خوشه‌بندی‌های k-means و الگوریتم‌های خوشه‌بندی فیلتر شده در مرجع [۸] انجام شده است. محققان در مرجع [۹] الگوریتم‌های رگرسیون خطی[2] و خوشه‌بندی k-means را مورد بررسی قرار داده و یک تجزیه و تحلیل مقایسه‌ای از این روش‌ها برای تشخیص تهاجم نیز انجام داده‌اند. نتایج نشان می‌دهد که روش رگرسیون خطی و خوشه‌بندی کارایی بسیار بالایی در تشخیص حملات شبکه با دقت بالا دارد. از نقاط ضعف آن سرعت پایین با افزایش کلان داده‌ها است. در مرجع [۱۰] از روش‌های داده‌کاوی در سیستم‌های تشخیص نفوذ استفاده شده که بر اساس خوشه‌بندی و درخت تصمیم و الگوریتم ژنتیک است. نتایج شبیه‌سازی نشان می‌دهد که روش پیشنهادی نرخ تشخیص و نرخ آلارم غلط را نسبت به سایر روش‌های خوشه‌بندی بهبود داده‌اند. در

مرجع [۱۱] یک سیستم تشخیص نفوذ کارا ارائه شده که از الگوریتم k-means و ژنتیک بهبودیافته[۳]به‌منظور تشخیص نوع حمله استفاده می‌کند. نتایج نشان می‌دهد که روش پیشنهادی در تشخیص حمله در مقایسه با الگوریتم k-mans++ دقیق‌تر است. محققان در مرجع [۱۲] به ارزیابی الگوریتم‌های یادگیری ماشین از قبیل k-means و C-means و منطق فازی اشاره نموده‌اند که نتایج نشان می‌دهد بیشتر حملات توسط الگوریتم خوشه‌بندی C-means و منطق فازی تشخیص داده می‌شوند و روش پیشنهادی توانایی دارد که بهبودی در تشخیص حمله داشته باشد. از معایب این روش عدم تحلیل توابع عضویت فازی است. مرجع [۱۳] از روش یادگیری ماشین توسعه یافته در سیستم های تشخیص نفوذ استفاده کرده که از الگوریتم ژنتیک برای انتخاب ویژگی استفاده شده است. نتایج شبیه‌سازی‌های آن دقت خوبی را نشان داده ولی پیاده سازی روش مذکور نسبت به روش بردار پشتیبان بسیار پیچیده‌تر است. در مرجع [۱۴] یک سیستم تشخیص نفوذ[۴]با اصلاح الگوریتم ژنتیک فازی و کاهش مشخصه‌ها ارائه شده که کاهش مشخصه‌ها باعث می‌شود زمان آموزش به‌طور قابل ملاحظه‌ای کاهش یابد. نتایج آزمایش‌ها نشان می‌دهد نرخ تشخیص افزایش‌یافته و نرخ مثبت غلط کاهش‌یافته است و از معایب آن سرعت پایین و حجم محاسبات بالا است. در مرجع [۱۵] از یک رویکرد فازی-ژنتیک[۵]که با قوانین فازی داده‌های حمله شبکه را مرتب می‌کنند به‌منظور تشخیص نفوذ در شبکه استفاده شده است. نتایج نشان می‌دهد که در روش پیشنهادی نرخ دقت تشخیص نفوذ برای حملات بهتر شده لیکن از معایب آن حساسیت به تابع عضویت فازی و حجم محاسبات بالا است.

در مرجع [۱۶] از درخت رگرسیون و منطق فازی استفاده شده و با پیش‌پردازش و پاک‌سازی داده‌ها نتایج دقت و اعتبارسنجی با استفاده از ماتریس اغتشاش[۶]تعیین شده است. بر اساس نتایج تجزیه و تحلیل به‌دست آمده می‌توان نتیجه گرفت که برای یک سیستم تشخیص نفوذ، الگوریتم CART پیشنهادی و منطق فازی با تعداد داده‌های مختلف اعتبارسنجی، میانگین دقت و میانگین زمان اعتبارسنجی بهتری را نشان می‌دهد اما از معایب آن حساسیت به تابع عضویت فازی است. مرجع [۱۷] از یک مجموعه داده شبکه برچسب‌گذاری شده Kyoto 2006+ استفاده شده است. در این مجموعه داده هر نمونه به عنوان نرمال (حمله نبودن)، حمله (حمله شناخته شده) و حمله ناشناس برچسب‌گذاری شده است. آن‌ها از الگوریتم درخت تصمیم (J48) به‌منظور طبقه‌بندی بسته‌های شبکه برای سیستم‌های تشخیص نفوذ استفاده کرده‌اند. روش پیشنهادی دقت بالایی در تشخیص ارتباط یعنی حمله نبودن، حمله شناخته شده و حمله ناشناس دارد اما ناتوانی در کار با حجم بالای داده را دارد. در مرجع [۱۸] یک ساختار تشخیص حمله کارا را با استفاده از یک روش بهینه سازی دقیق، مقاوم و تطبیقی به نام بهینه سازی ذرات گروهی آشوب با زمان متغیر[۷](TVCPSO) به‌منظور تنظیم پارامتر و انتخاب مشخصه برای برنامه‌ریزی خطی معیار چندگانه[۸]و ماشین بردار پشتیبان[۹]ارائه شده است. در روش پیشنهادی، یک تابع هدف وزنی ارائه شده که تبادل بین ماکزیمم سازی نرخ تشخیص و کمینه سازی نرخ الارم اشتباه در نظر گرفته می شود که تعدادی از مشخصه‌ها را در نظر می‌گیرد. همچنین برای سریع‌تر کردن الگوریتم بهینه‌سازی اجتماع ذرات در جستجوی بهینه و اجتناب از افتادن در بهینه محلی، مفهوم آشوب[۱۰]در بهینه‌سازی اجتماع ذرات تطبیق داده شده و معیار وزن با زمان تغییر کرده و ضریب تغییر زمان معرفی شده است. عملکرد روش‌های پیشنهادی با استفاده از آزمایش‌های صورت گرفته نشان می‌دهد که روش پیشنهادی عملکرد بهتری از نظر تشخیص حمله با دقت بالا و نرخ آلارم غلط در مقایسه با نتایج به‌دست آمده از همه مشخصه‌ها دارد و از معایب این روش وابستگی به مقداردهی پارامترهای اجتماع ذرات و سرعت پایین یادگیری است. کارلوس کاتانیا و همکاران یک رویکرد برای برچسب‌گذاری خودکار ترافیک نرمال برای موقعیت‌هایی که توزیع کلاس تعادل مورد نیاز را ندارد در مقاله [۱۹] ارائه می‌دهند. در این حالت، فرایند برچسب‌گذاری خودکار با استفاده از SNORT، برای یک سیستم تشخیص نفوذ مبتنی بر بدرفتاری[۱۱]انجام می‌گیرد. نتایج نشان می‌دهد که رویکرد پیشنهادی نه‌تنها عملکرد بهتری نسبت به ماشین بردار پشتیبان موجود دارد بلکه تحت برخی از توزیع‌های حمله، بهبود نسبت به خود SNORT ایجاد می‌کند. در مرجع [۲۰] با الگوریتم رأی‌گیری[۱۲]با اطلاعات به‌دست آمده[۱۳]توزیع احتمالی این آموزش‌دهنده‌ها را به‌منظور انتخاب مشخصه‌های مهم که تأثیر مثبت بر دقت مدل پیشنهادی دارند، ارائه شده و سپس الگوریتم ترکیبی روش بیزین و درخت رگرسیون و درخت تصادفی و درخت تصمیم را ارائه کرده‌اند، بر اساس نتایج به‌دست آمده با استفاده از مدل پیشنهادی، شاهد بهبود در دقت، نرخ منفی غلط بالا هستیم ولی در کلان داده‌ها نا کارآمد است. در مرجع [۲۱] از ترکیب درخت تصمیم[۱۴]و ماشین بردار پشتیبان به‌عنوان یک روش ترکیبی در تشخیص نفوذ

استفاده شده است. ایده اصلی این تحقیق استفاده از یک روش ترکیبی است که دقت روش ماشین بردار پشتیبان به شکل به‌کاررفته در مقاله‌های [۲۲] و [۷] و روش ترکیبی درخت تصمیم و ماشین بردار پشتیبان ارائه شده توسط محققان در مرجع [۲۱] را بهبود دهد. محققان در مرجع [۲۳] به شناسایی مشخصه‌های کاهش‌یافته مهم با انتخاب آن‌ها توسط فیلتر کمی مشخصه‌ها[۵]پرداختند. تمرکز آن‌ها تشخیص کلاس‌های مختلف حملات با استفاده از روش بیزین و درخت تصمیم و RBF است. طبقه‌بند بیزین عملکرد بهتری نسبت به طبقه‌بندهای J48 و RBF ازنظر دقت و خطای طبقه‌بندی دارد. نتایج آزمایش‌ها نشان می‌دهد که مشخصه‌های انتخابی عملکرد بهتری برای طراحی سیستم تشخیص نفوذ کارا می‌دهند. یرونگ یک رویکرد ترکیبی مبتنی بر ماشین بردار پشتیبان و الگوریتم ژنتیک به‌منظور تشخیص نفوذ ارائه کرده است [۲۴]. در این رویکرد از الگوریتم ژنتیک به‌منظور پیدا کردن پارامترهای بهینه ماشین بردار پشتیبان استفاده شده و آن‌ها اثبات کردند بهره‌گیری از پارامترهای بهینه حاصل از الگوریتم ژنتیک در ماشین بردار پشتیبان تأثیر بالایی بر تشخیص نفوذ دارد. کویتن و همکاران از یک معماری سلسله مراتبی به‌منظور تشخیص نفوذ استفاده کردند [۲۵]. آن‌ها رویکردی مبتنی بر ماشین بردار پشتیبان برای این منظور ارائه کردند. طرح سیستم تشخیص نفوذ سلسله مراتبی پیشنهادی از مزایای سیستم تشخیص نفوذ مبتنی بر جریان و سیستم تشخیص نفوذ مبتنی بر بسته به‌منظور تشخیص بهتر نتایج استفاده می‌کند، اما از نقاط ضعف این روش نرخ مثبت غلط بالا و عدم تحلیل پارامترهای تصمیم گیری و استفاده از یک هسته است. ویجایاباند و همکارانش [۲۶] از یک رویکرد ترکیبی از الگوریتم ژنتیک و ماشین بردار پشتیبان به‌منظور تشخیص نفوذ استفاده کردند. مسأله مورد بررسی آنها انتخاب مشخصه‌های بااهمیت و کاهش ابعاد داده‌ها با کمترین از دست رفتن اطلاعات و دقت حاصل از ماشین بردار پشتیبان بود که از الگوریتم ژنتیک برای این کار استفاده کردند. گاتاما رامان و همکارانش [۲۷] از ماشین بردار پشتیبان و الگوریتم ژنتیک به‌منظور انتخاب مشخصه و تنظیم بهینه پارامترهای ماشین بردار پشتیبان با هسته تابع اساس شعاعی در مسئله تشخیص نفوذ استفاده کردند. آنها به‌منظور تعیین جمعیت اولیه از مفهوم کلیک[۶]بهره گرفتند. مطالب ذکر شده به صورت خلاصه در جدول (۱) بیان شده است. با توجه به جداول دسته‌بندی شده مشاهده می شود که روش‌های ماشین بردار پشتیبان برای مسئله تشخیص نفوذ داری دقت بالایی است ولی حساسیت آن به انتخاب هسته یکی از محدودیت‌های این روش است. همچنین با توجه به سادگی روش‌های ماشین‌های بردار پشتیبان در عین حفظ دقت لازم این روش نسبت به روش‌های جدیدتر مانند ماشین‌های یادگیری توسعه یافته کماکان کاربرد بیشتری دارند و در کنار روش‌های دیگر از آن استفاده کردند.

در این مقاله پس از بیان مساله و مروری بر روش‌های سایر مراجع در بخش اول، در بخش دوم، روش پیشنهادی که حساسیت هسته در آن بهبود یافته و نسبت به سایر روش‌ها برتری نسبی دارد ارایه خواهد شد. در بخش سوم ضمن ارایه مفروضات مساله به تحلیل و ارزیابی روش و روش پیشنهادی را با سایر مراجع مقایسه خواهیم کرد. در بخش چهارم نتیجه‌گیری ارایه خواهد شد.

۲- روش پیشنهادی

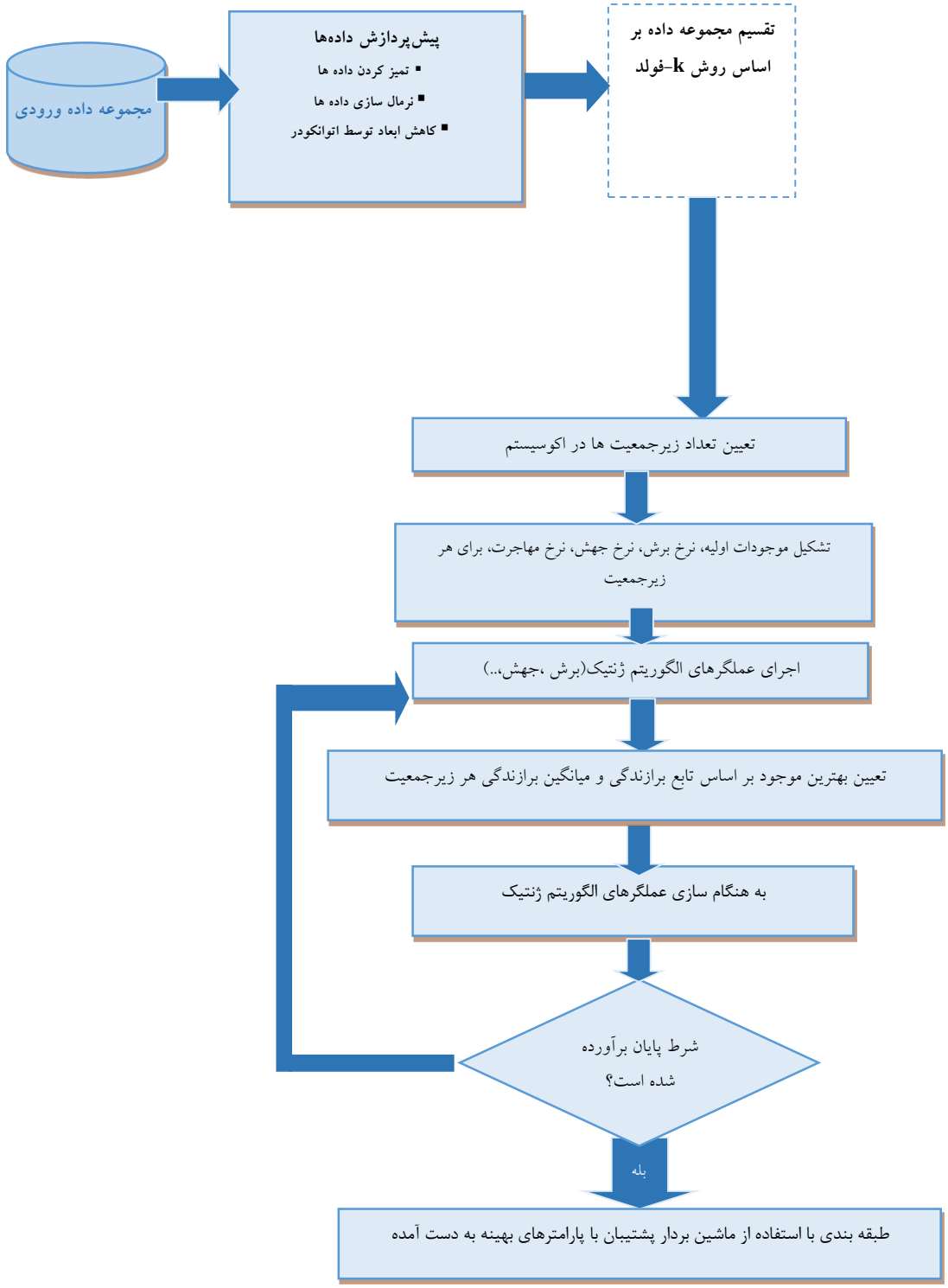
همان‌طور که اشاره شد برآنیم تا دقت تشخیص نفوذ را افزایش داده و محدودیت‌های روش‌های طبقه‌بندی را بهبود دهیم. برای این مهم از الگوریتم ژنتیک خود تطبیق جزیره‌ای برای حل ماشین بردار پشتیبان به‌صورت یادگیری چندهسته‌ای با کد کننده‌های خودکار ارائه‌شده است. گرچه روش ماشین بردار پشتیبان یک روش کارا و متداول برای تشخیص نفوذ است اما حساسیت به انتخاب هسته و تاثیر یکسان هر یک از هسته‌ها و پیدا کردن ضرایب مناسب برای هسته‌ها از محدودیت‌های آن است. در روش پیشنهادی برای رفع این محدودیت راه‌کارهای زیر را ارائه می‌شود:

- بهره‌گیری از هسته‌های مختلف در طبقه‌بند مبتنی بر روش ماشین بردار پشتیبان
- ارائه یک رویکرد کارا به‌منظور ترکیب هسته‌ها
- تعیین پارامترهای بهینه هر یک از هسته‌ها به همراه تأثیر هر هسته بر طبقه‌بندی
- نحوه عملکرد الگوریتم ژنتیک پویای خود تطبیق و بهینه‌سازی پارامترهای ماشین بردار پشتیبان و استفاده از اتوانکودر در شکل (۱) نشان داده شده است.

Table (1): Comparison and analysis of different methods in the field of intrusion detection

جدول(۱) مقایسه و تحلیل روشهای مختلف در زمینه تشخیص نفوذ

ردیف	رفرنس	سال	روش	نقاط قوت	نقاط ضعف
۱	[۳]	۲۰۱۹	ماشین بردار پشتیبان- الگوریتم پروانه ای	افزایش دقت هشدار کاذب- انتخاب کارا مشخصه‌ها	عدم بهبود در دقت تشخیص نفوذ
۲	[۴]	۲۰۲۰	ماشین بردارپشتیبان- هایپر گراف کلیک	دقت بالا	پیچیدگی محاسباتی
۳	[۸]	۲۰۱۴	خوشه‌بندی	عدم نیاز به فیلتر	حساسیت به تعداد هسته و انتخاب مراکز هسته اولیه
۴	[۹]	۲۰۱۶	خوشه‌بندی- درخت رگرسیون	کاهش محاسبات- دقت بالا	سرعت پایین- عدم بررسی تاثیر پارامترهای تصمیم
۵	[۱۱]	۲۰۱۷	خوشه‌بندی- الگوریتم ژنتیک	قابلیت کار با داده‌های نویزدار و پرت	حساسیت به مراکز هسته و تعداد هسته
۶	[۱۲]	۲۰۱۷	خوشه‌بندی- منطق فازی	انعطاف‌پذیری بالا در عضویت توانایی کار با داده‌های نویزدار	عدم تحلیل توابع عضویت فازی
۷	[۱۳]	۲۰۲۰	الگوریتم ژنتیک- یادگیری ماشین توسعه یافته-ELM	دقت بالا	پیچیدگی محاسباتی
۸	[۱۴]	۲۰۱۴	انتخاب مشخصه- منطق فازی- الگوریتم ژنتیک	انتخاب مشخصه کارا- نرخ مثبت غلط پایین	سرعت پایین- حجم محاسبات بالا
۹	[۱۵]	۲۰۱۷	منطق فازی- الگوریتم ژنتیک	بهبود زمان مورد نیاز برای آموزش و اختصاص کمتر حافظه	حساسیت به تابع عضویت- حجم محاسبات بالا
۱۰	[۱۶]	۲۰۱۵	منطق فازی- درخت رگرسیون	بهره‌گیری از مکانیزم بهینه‌سازی- دقت تشخیص بالا	حساسیت به تابع عضویت فازی- ناتوانی در کار با حجم بالای داده-ها
۱۱	[۱۸]	۲۰۱۶	انتخاب مشخصه- بهینه‌سازی اجتماع ذرات	بهره‌گیری از مکانیزم بهینه‌سازی به‌منظور دست‌یابی به عملکرد بالا	واستگی به مقداردهی پارامترهای اجتماع ذرات سرعت پایین یادگیری
۱۲	[۱۹]	۲۰۱۵	درخت تصمیم‌گیری و برچسب-گزاری	دقت تشخیص بالا	پیچیدگی نتایج با افزایش سطح درخت- مشکل کار با حجم بالای داده
۱۳	[۲۰]	۲۰۱۵	درخت رگرسیون- درخت تصمیم گیری بیزین	دقت بالا	پیچیدگی محاسبات بالا- ناکارامدی در کار با کلان داده‌ها
۱۴	[۲۱] [۲۲] [۲۳]	۲۰۱۷ ۲۰۱۸	درخت تصمیم‌گیری- ماشین بردار پشتیبان	سهولت استفاده- نرخ مثبت غلط پایین	ناکارامدی با افزایش سطح درخت- استفاده از یک هسته عدم تحلیل پارامترهای تصمیم گیری
۱۵	[۲۶]	۲۰۱۸	انتخاب مشخصه- الگوریتم ژنتیک- ماشین بردار پشتیبان	بهره‌گیری از چندهسته	تاثیر یکسان هر یک از هسته‌ها در نظر نگرفتن مقدار پارامترهای تصمیم‌گیری
۱۶	[۲۷]	۲۰۱۷	الگوریتم ژنتیک- ماشین بردار پشتیبان	بهره‌گیری از چندهسته	تاثیر یکسان هر یک از هسته‌ها در نظر نگرفتن مقدار پارامترهای تصمیم‌گیری



شکل(۱): فرایند روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای خود تطبیق

Figure (1): Process of the proposed hybrid method based on auto econcer and improved support vector machine with Self-Adaptive Island Genetic Algorithm

گام های شکل (۱) به‌صورت زیر تشریح می‌شود.

گام اول: پس از جمع‌آوری داده‌ها و یکپارچه‌سازی آن‌ها، داده‌های مربوط به اتصال‌های صورت گرفته پیش‌پردازش می‌شوند. که فعالیت‌های پیش‌پردازش عبارتند از:

تمیزسازی داده‌ها ازنظر مقادیر خارج از رفتار یا حدی، مشخصه‌های تکراری، داده‌هایی که در فرم مناسب برای مدل‌سازی با حذف آن‌ها.

نرمال‌سازی داده‌ها به‌منظور تسریع فاز یادگیری و اجتناب از مسائل عددی مانند از دست دادن صحت به خاطر سرریز شدن، با استفاده از روش Min-Max و رابطه زیر:

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{x} - \mathbf{x_{min}}}{\mathbf{x_{max}} - \mathbf{x_{min}}} & \text{و } \mathbf{x_{max}} \neq \mathbf{x_{min}} \\ \mathbf{0} & \text{و } \mathbf{x_{max}} = \mathbf{x_{min}} \end{cases} \quad (۱)$$

گام دوم: مجموعه داده به دو مجموعه داده آموزشی و آزمایشی تقسیم‌بندی می‌شود. کاهش حجم داده‌ها به‌منظور بازنمایش فشرده داده‌ها و کاهش ابعاد داده‌ها با استفاده از اتوانکودر انجام می‌شود. تعداد لایه مخفی در اتوانکودر یک پارامتر تأثیرگذار بر کیفیت کاهش حجم داده‌ها است که تعیین مقدار مناسب برای آن می‌تواند عملکرد طبقه‌بندی و تشخیص، را بهبود دهد. ما در اینجا از رویکرد سعی و خطا به‌منظور تعیین مقدار مناسب آن استفاده می‌کنیم. معیار ارزیابی هر یک از تعداد لایه مخفی بر اساس صحت حاصل از طبقه‌بندی با ماشین بردار پشتیبان بر روی مجموعه داده حاصل محاسبه می‌شود. نتایج نشان می‌دهد که تعداد لایه مخفی کمتر از ۵ و بیشتر از ۱۵ نتایج خوبی را نمی‌دهند و درنتیجه ما این بازه را موردبررسی قرار می‌دهیم.

گام سوم: تعداد زیرجمعیت ها در اکوسیستم به صورت تصادفی تعیین می شود.

گام چهارم: موجودات اولیه تشکیل می‌شود و نرخ برش، نرخ جهش، نرخ مهاجرت برای هر زیرجمعیت به‌طور مستقل تعیین می‌گردد.

تعیین احتمال جهش هر زیرجمعیت p_i به ۰/۰۰۰۱

تعیین احتمال برش هر زیرجمعیت p_i برابر ۰/۵

تعیین درصد موجودات بدون تغییر زیرجمعیت p_i برابر ۰/۵

گام پنجم: اجرای عملگرهای برش و جهش الگوریتم ژنتیک برای هر زیر جمعیت و بهنگام سازی عملگرها

گام ششم: برازندگی هر یک از موجودات در الگوریتم ژنتیک بر اساس دقت حاصل از تحلیل ROC[۱۷]محاسبه می‌شود.

گام هفتم: بهترین موجود بر اساس تابع برازندگی و میانگین برازندگی برای هر زیرجمعیت تعیین می‌شود.

گام هشتم: انتقال بهترین موجود هر زیر جمعیت با بهترین تابع برآزندگی به نسل بعدی با توجه به توپولوژی‌های ارتباطی و تولید بقیه‌ی موجودات زیرجمعیت به کمک اعمال برش و جهش با توجه به احتمال‌های تعیین شده.

گام نهم: عملگرهای الگوریتم ژنتیک بهنگام سازی می‌شود.

گام دهم: دقت تشخیص بر روی مجموعه داده آزمایشی محاسبه می‌شود.

گام یازدهم: عملکرد دقت طبقه‌بندی مورد ارزیابی قرار می‌گیرد.

گام دوازدهم: در صورتی‌که جمعیت‌ها به شرط همگرایی (مشابه بودن بهترین موجودات در ۱۰ نسل متوالی) نرسیده بود، گام پنجم اجرا می شود.

گام سیزدهم: نمایش بهترین موجود در بین زیرجمعیت‌ها به‌عنوان جواب نهایی.

گام چهاردهم: ماشین بردار پشتیبان با پارامترهای بهینه بدست آمده طبقه‌بندی می‌شود.

۲-۱- تشریح گام‌ها

پس از جمع‌آوری اطلاعات و پیش پردازش داده‌ها برای کاهش مشخصه‌های داده از اتوانکودر[۱۸]استفاده می‌شود. سپس برای طبقه‌بندی و دسته‌بندی اطلاعات مبنی بر حمله یا نرمال بودن سراغ ماشین بردار پشتیبان می‌رویم ولی شکل ساده svm[۱۹] دارای محدودیت‌هایی ا ست که برای رفع این محدویت‌ها از مجموع وزنی ه سته‌های ما شین بردار پ شتیبان ا ستفاده می‌کنیم.

برای بهینه سازی پارامترهای mkl از ژنتیک جزیره‌ای خود تطبیق استفاده نمودیم که بتواند پارامترهای ماشین بردار پشتیبان را بهینه نماید. متداول‌ترین توابع هسته در رابطه (۱) آورده شده است. که $K(x_i,x_j)$ هسته‌های ماشین بردار پشتیبان هستند. x_i,x_j، ورودی i ام و j ام و p ، γ ، σ ، c و k پارامترهای توابع هسته هستند.

$$
\begin{aligned}
&\text{چندجمله‌ای} && K(x_i,x_j) = (x_i . x_j + 1)^p \\
&\text{چندجمله‌ای نرمال شده} && K(x_i,x_j) = \frac{(x_i . x_j)^p}{\sqrt{(x_i . x_i)^p * (x_j . x_j)^p}} \\
&\text{تابع بنیادی شعاعی}[21] && K(x_i,x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \ \forall \gamma > 0 \\
&\text{تابع بنیادی شعاعی گاوس} && K(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \\
&\text{حلقوی}[22] && K(x_i,x_j) = \tanh(k x_i . x_j + c) \text{ و } k < 0 \ \& \ c < 0
\end{aligned}
\qquad (۱)
$$

در ادامه از ماشین بردار پشتیبان برای تعیین کلاس‌های مختلف نفوذ استفاده می‌کنیم که با توسعه آن و استفاده از توابع کرنل، به عنوان طبقه‌بندی کننده چندکالسی و غیر خطی از آن استفاده می‌کنیم. استفاده از هسته در ساختار طبقه‌بند SVM باعث می شود که داده‌هایی که دارای الگوی غیرخطی هستند با نگاشت به فضایی با ابعاد بیش‌تر بتوانند با مدل خطی تفکیک‌پذیر شوند. این کار باعث افزایش دقت تفکیک‌کنندگی مدل یادگیری ماشین می شود. استفاده از توابع هسته‌های مختلف به دلیل اینکه توابع هسته در هر کاربرد اختصاصی هستند لذا نمی‌توان از یک هسته عمومی برای همه کاربردها استفاده کرد. برای این منظور از مجموع وزنی هسته‌ها استفاده شده و برای بهبود پارامترهای ماشین بردار پشتیبان چند هسته‌ای از الگوریتم ژنتیک پویای خود تطبیق استفاده می‌کنیم. رابطه (۲) جهت مدل ارایه شده در این مقاله استفاده شده است:

$$
\begin{aligned}
\mathbf{K_\eta(x_i \, . \, x_j)} = \ & \mathbf{w_1 (x_i . x_j + 1)^p} + \mathbf{w_2} \frac{\mathbf{(x_i . x_j)^p}}{\sqrt{\mathbf{(x_i . x_i)^p * (x_j . x_j)^p}}} \\
& + \mathbf{w_3} \exp\left(-\boldsymbol{\gamma} \|\mathbf{x_i - x_j}\|^2\right) + \mathbf{w_4} \exp\left(-\frac{\|\mathbf{x_i - x_j}\|^2}{\mathbf{2\sigma^2}}\right) + \mathbf{w_5} \tanh(\mathbf{k x_i . x_j + c})
\end{aligned}
\qquad (۲)
$$

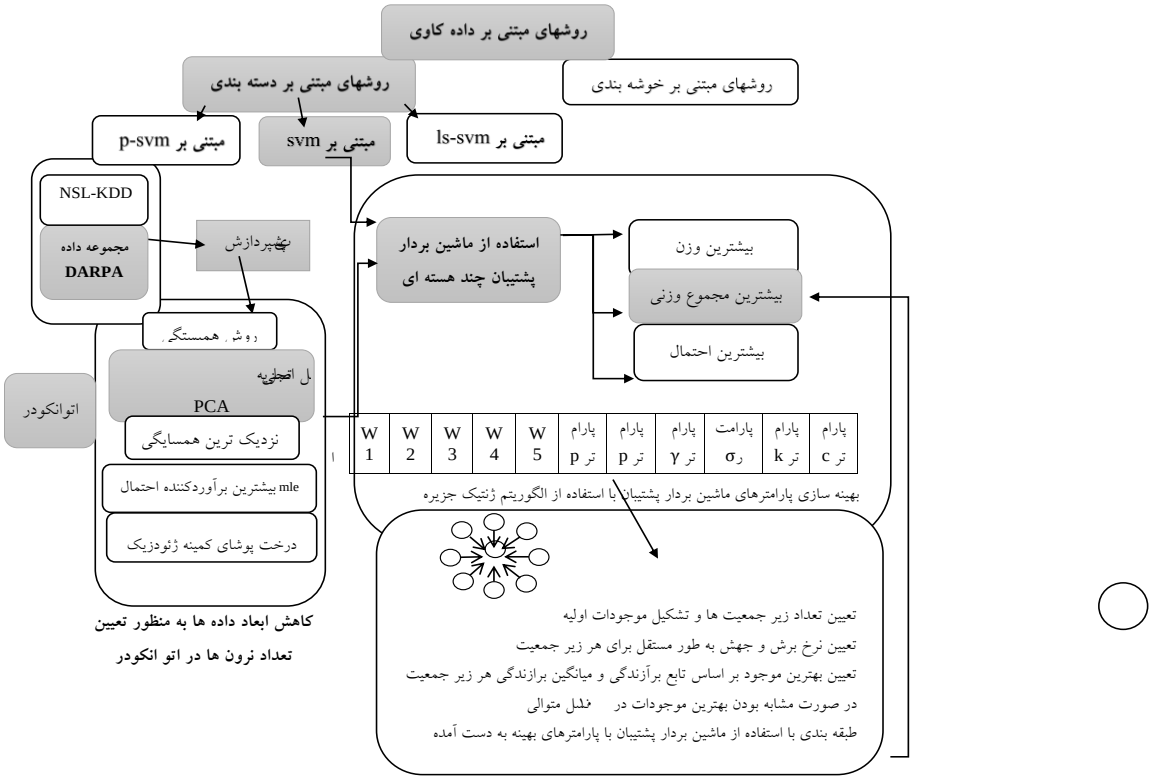
w_iها وزن‌های تابع کرنل رابطه (۱) هستند. مدل‌سازی مسأله تعیین بهترین وزن و پارامتر هسته‌ها در روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق ما از دقت طبقه‌بندی با هسته‌ای که بیشترین مجموع وزنی را در کروموزوم دارد به عنوان دقت روش پیشنهادی استفاده می‌کنیم که در جدول (۲) آمده است.

Table (2): Different methods of combining nuclei based on the corresponding weight in chromosome genes

جدول (۲): روش‌های مختلف ترکیب هسته‌ها بر اساس وزن متناظر در ژن‌های کروموزوم

بیشترین وزن	استفاده از هسته متناظر با بیشترین وزن اختصاص داده شده
بیشترین مجموع وزنی	هر هسته برای هررکورد یک برچسب پیش بینی می کند و ترکیب هسته ها در این روش بر اساس مرتب سازی بر اساس برچسب و مجموع وزن هایی است که هسته های متناظر به آن برچسب طبقه بندی کرده اند و انتخاب بیشترین مجموع وزنی به عنوان برچسب رکورد
بیشترین احتمال	محاسبه احتمال انتخاب یک کلاس با استفاده از قانون بیز و انتخاب کلاس با استفاده از روش چرخ رولت

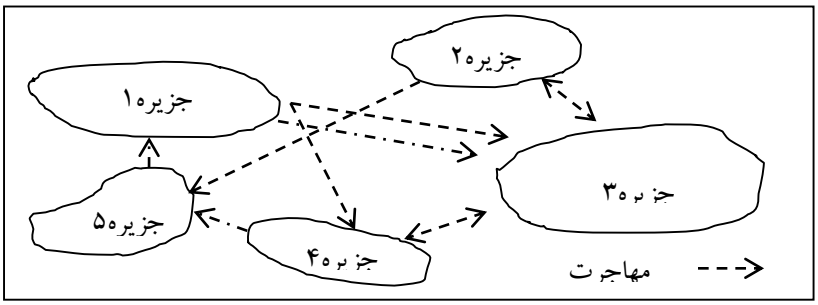
هر یک از روش‌های ترکیب هسته‌های مورد بررسی عملکرد متفاوتی در طبقه‌بندی دارند. شکل (۲) شمای روش پیشنهادی را در حوزه تشخیص نفوذ همراه با روند کار روش پیشنهادی که استفاده از الگوریتم ژنتیک جزیره‌ای خود تطبیق برای حل ماشین بردار پشتیبان است، را نشان می‌دهد.



شکل (۲): شمای روش پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای خود تطبیق با نرخ مهاجرت

Figure (2): Process of the proposed hybrid method based on autoconcer and improved support vector machine with Self-Adaptive Island genetic algorithm to migration rate

در روش پیشـنهادی مبتنی بر الگوریتم ژنتیک به‌منظور بهبود عملکرد ماشـین بردار پشـتیبان از طبیعت موازی بودن الگوریتم ژنتیک ا ستفاده کرده‌ایم. این ویژگی الگوریتم ژنتیک باعث می شود که هم‌زمان چندین جمعیت پردازش شده و در تکامل نقش داشته باشند. این کار علاوه بر افزایش فضای جواب باعث افزایش قدرت اکتشاف می‌شود. در این استراتژی به‌جای پردازش یک جمعیت از پردازش چندین جمعیت با اندازه‌های مختلف اسـتفاده می‌شـود. این کار علاوه بر افزایش قدرت اکتشـاف می‌تواند امکان اعمال اسـتراتژی‌های متفاوت را برای هر زیر جمعیت و مسـتقل از دیگر زیرجمعیت‌ها را فراهم آورد. مدل‌های توزیعی نه تنها باعث افزایش سرعت می شوند، بلکه تنوعی از کروموزم‌ها را نگه می‌دارند. نگهداری تنوعی از کروموزم‌ها، مهم‌ترین عامل در بهبود کیفیت جواب‌ها در محاسبات تکاملی است.



شکل (۳): یک مدل جزیره‌ای پیشنهادی

Figure (2): A proposed island model

همان‌طور که در شـکل (۳) نمایش داده شـده، زیرجمعیت‌ها هر یک مانند یک جزیره رفتار می‌کنند و مسـتقل از یکدیگر عمل کرده، عملگرها و پارامترهای الگوریتم ژنتیک بر یک زیرجمعیت اعمال می‌شوند؛ به وضعیت زیرجمعیت‌های دیگر وابسته نیست.

نرخ جهش، برش و مهاجرت پویا و مستقل را برای هر زیرجمعیت انجام می‌دهیم. با توجه به این‌که جهش عملگری است که ساختار کروموزوم را به‌هم ریخته و باعث می‌شود جستجو به مناطق جدید از فضای جواب منتقل شود، نرخ جهش در زیرجمعیت‌هایی که دارای موجودات با برازندگی بدتر هستند افزایش‌یافته و برای زیرجمعیت‌های با برازندگی بهتر کاهش یابد. بنابراین با این نرخ جهش پویا می‌توان قدرت اکتشاف الگوریتم را بالا برد. عملگر برش در الگوریتم ژنتیک قادر به استفاده از ویژگی‌های موجودات قوی به‌منظور بهبود جواب‌ها در تکامل می‌باشد. با توجه به این‌که این عملگر سعی بر بردن الگوریتم به سمت جواب‌های بهتر دارد بنابراین ما نرخ برش در زیرجمعیت‌های قوی را بیشتر کرده و برای زیرجمعیت‌های ضعیف کاهش می‌دهیم. بدین ترتیب قدرت بهره‌برداری را افزایش می‌دهیم. از طرف دیگر به‌منظور استفاده از موجودات دیگر زیرجمعیت‌ها در تکامل هر زیرجمعیت، امکان مهاجرت موجودات بین زیرجمعیت‌ها را فراهم می‌آوریم. این نرخ برای زیرجمعیت‌های با بهترین میانگین برازندگی پایین است و دیگر زیرجمعیت‌ها بهترین موجود را با نرخ‌های متناسب با میانگین برازندگی زیرجمعیت بر اساس توپولوژی ارتباطی ارسال کرده و بدترین موجود زیرجمعیت مقصد به مبدأ ارسال می‌شود. این نحوه برش، جهش و مهاجرت بهبود بهره‌برداری از مناطق امیدبخش و به‌طور هم‌زمان بهبود اکتشاف از دیگر نواحی را در پی خواهد داشت. این کار را به کمک الگوریتم ژنتیک جزیره‌ای[۲۸] انجام داده‌ایم که هر جزیره یک زیر جمعیت در اکوسیستم می‌باشد و امکان مهاجرت موجودات بین زیرجمعیت‌ها بر اساس توپولوژی‌های ارتباطی مختلف ستاره‌ای، خطی و حلقوی وجود دارد. تعیین تعداد موجودت منتقل شده بدون تغییر به نسل بعد می‌تواند بر کارایی الگوریتم تأثیر به سزایی داشته باشد که آن را به یک چالش تبدیل کرده است. ما در الگوریتم ژنتیک پیشنهادی با در نظر گرفتن این پارامتر، یک نرخ پویا برای آن در طی تکامل در نظر می‌گیریم. نرخ موجودات منتقل شده بدون تغییر به نسل بعدی متناسب با متوسط برازندگی هر جمعیت است و در زیرجمعیت‌های با موجودات قوی این پارامتر مقدار کمتری داشته و برای زیرجمعیت‌های با موجودات ضعیف مقدار بالاتری دارد. این کار باعث می‌شود به ساختار موجودات قوی آسیبی نرسد.

۳- تحلیل و ارزیابی

به منظور ارزیابی روش ترکیبی مبتنی بر اتوانکودر و ماشین بردار پشتیبان به عنوان روش پیشنهادی ما در این تحقیق، آن را با روش ارایه شده توسط ویجایایاند و همکارانش [۲۶] با هسته تابع هسته شعاعی مورد مقایسه قرار می دهیم.

۳-۱ فرضیات

کامپیوتر مورداستفاده به‌منظور اجرای روش‌های مختلف در این مقاله دارای پردازنده i5 با ۸ گیگابایت حافظه RAM می‌باشد. مجموعه داده مورد استفاده در این مقاله به‌منظور بررسی عملکرد روش پیشنهادی مجموعه داده DARPA است. این داده ها در جدول (۳) آمده اند و به‌عنوان KDD CUP2009 نیز نامیده می شود[۲۹]. این مجموعه داده حاوی ۱۰۴۸۵۷۶ رکورد و ۴۱ مشخصه به همراه یک مشخصه نوع کلاس است و دارای بیشتری استفاده به‌عنوان مجموعه داده موردبررسی در تحقیقات تشخیص نفوذ می‌باشد. ۴۲ مشخصه مجموعه داده DARPA نشان‌دهنده یک برچسب کلاس است که نوع اتصالات هر نمونه را نشان می‌دهد. مشخصه چهل و دوم می‌تواند ۲۲ مقدار متفاوت به‌عنوان نوع اتصال بگیرد که عبارتند از:

ftp-write, imap, land, loadmodule, multihop, phf, rootkit, spy, teardrop, warezclient, warezmaster, pod, back, buffer_overflow and normal. smurf, perl, guess_passwd, ipsweep, neptune, nmap, portsweep, satan,

Table (3): Attack categories in the dataset. KDDCUP99[29]

جدول (۳) دسته های حمله در مجموعه داده های .KDDCUP99 [۲۹]

نوع حمله	انواع حمله
Ipsweep, nmap, portsweep, satan	Prob ing
Back,land,Neptune,pod,smurf,teardrop	DoS
Rootkit,perl,loadm odule,buffer-over flow	U2R
ftp-write,spy,phf,guess-passwd,imap,warezclient,warezmaster,multihop	R2L

در الگوریتم ژنتیک جزیره ای از زیر جمعیت‌ها به جای یک جمعیت اســتفاده می‌کنیم. در این روش ما تعداد ۱۰ زیر جمعیت در نظر گرفته شده است و موجودات هر زیر جمعیت به صورت تصادفی تشکیل می‌شوند. نرخ برش و جهش را به طور مستقل برای هر زیر جمعیت تعیین می‌شود. هر زیر جمعیت یک گره از این شبکه ارتباطی می باشد. کروموزوم‌ها در روش پیشنهادی در جدول (۴) نمایش داده شده است.

Table (4): Chromosome display in the proposed method

جدول(۴) نمایش کروموزوم در روش پیشنهادی

پارامتر c در هسته حلقوی	پارامتر k در هسته حلقوی	پارامتر σ در تابع بنیادی شعاعی گاوس	پارامتر γ در هسته تابع بنیادی شعاعی	پارامتر p در هسته چندجمله‌ای نرمال شده	پارامتر p در هسته چندجمله‌ای	وزن هسته حلقوی	وزن هسته تابع بنیادی شعاعی گاوس	وزن هسته تابع بنیادی شعاعی	وزن هسته چندجمله‌ای نرمال شده	وزن هسته چندجمله‌ای

۳-۲ متغیرهای قابل اندازه گیری

به منظور محاســبه دقت طبقه بندی از تحلیل ROC [۱]و متریک‌هایی اســتاندارد که در جدول (۵) آورده شــده اســت، اســتفاده می‌شود.

Table (5): Standard metrics to assess intrusion detection [5]

جدول (۵) متریک‌های استاندارد برای ارزیابی تشخیص نفوذ [۵]

متریک‌های استاندارد		برچسب ارتباط پیشگویی شده	
		نرمال	حمله
برچسب ارتباط واقعی	نرمال	نرمال صحیح (TN)	تشخیص اشتباه (FP)
	حمله	نرمال اشتباه (FN)	تشخیص صحیح (TP)

نرمال صحیح[۴](TN)، تشخیص اشتباه[۵](FP)، نرمال اشتباه[۶](FN)، تشخیص صحیح[۷](TP)

دقت (Precision) : دقت مجموعه‌ای از اندازه‌گیری‌ها، به صــورت میزان نزدیک بودن نتایج آن اندازه‌گیری‌ها به یکدیگر، توســط رابطه (۳) تعریف می شود. در واقع، هرچه نتایج به دست آمده فاصله کمتری با یکدیگر داشته باشند، دقت آن سیستم بیشتر اســت. به عبارت دیگر، دقت را می‌توان تابعی از انحراف معیار داده‌ها به حســاب آورد. هرچه انحراف معیار داده‌ها کمتر باشــد، سیستم دقیق‌تر است .

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (۳)$$

بازیابی (sensitivity یا recall) که در ۴ نشان داده شده عبارتست از کسری از جوابهای مثبت که تشخیص داده شده‌اند.

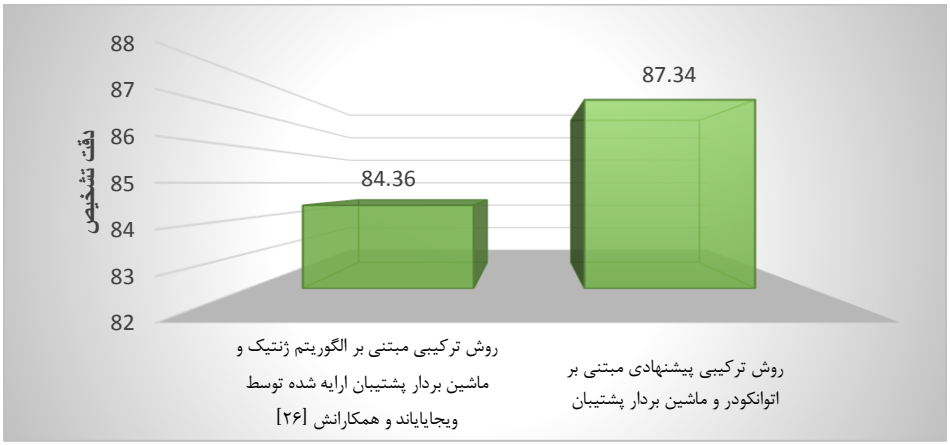
$$Recall = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (۴)$$

صحت (accuracy)، در رابطه (۵)، بیان‌گر آن است که مقدار اندازه‌گیری شده چقدر به مقدار واقعی نزدیک است.

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

۳-۳ شبیه‌سازی

نتایج حاصل از روش‌های ترکیبی مبتنی بر اتوانکودر و ماشین بردار پشتیبان و ویجایاباند و همکارانش در شکل (۴) نمایش داده شده است. روش پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان از نظر دقت عملکرد بهتری تشخیص نسبت به مرجع [۲۶] دارد. این بهبود دقت می‌تواند دلیل استفاده از مشخصه‌ها در روش ویجایاباند و ترکیب مشخصه‌ها در روش پیشنهادی باشد؛ چرا که این امر میزان اطلاعات داده‌ها را افزایش داده و منجربه بهبود دقت طبقه بندی می‌شود.



شکل (۴): ارزیابی روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان با مقایسه با روش ترکیبی مبتنی بر الگوریتم ژنتیک و ماشین بردار پشتیبان ارایه شده توسط ویجایاباند و همکارانش [۲۶]

Figure (4): Evaluation of the proposed hybrid method based on support vector autowankodo machine with comparison with the combined method based on Genetic Algorithm and support vector machine [26]

به منظور بررسی عملکرد روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان و روش ویجایاباند و همکارانش [۲۶]، آنها را از نظر تشخیص کلاس نفوذهای مختلف مقایسه می‌کنیم. برای این منظور در مرحله آزمایش روش‌ها، دقت تشخیص را برای هر کلاس نفوذ به دست می آوریم ، در جدول (۶) نمایش داده شده است.

Table (6): Evaluation of the proposed hybrid method based on the support vector machine autowanker based on the accuracy of different intrusion class detection [26]

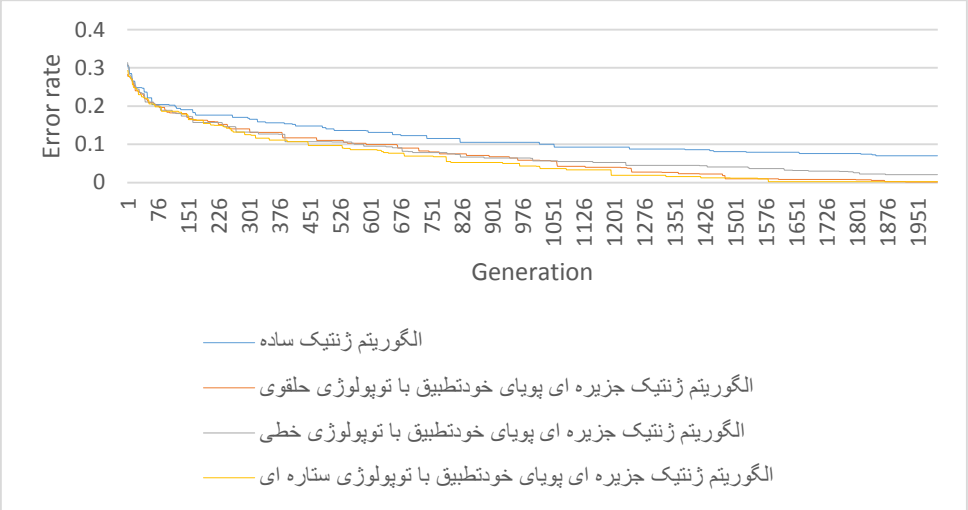
جدول (۶): ارزیابی روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان و روش ویجایاباند و همکارانش [۲۶] بر اساس دقت تشخیص کلاس نفوذهای مختلف

کلاس	روش ترکیبی پیشنهادی متنی بر اتوانکودر و ماشین بردار پشتیبان		روش ترکیبی مبتنی بر الگوریتم ژنتیک و ماشین بردار پشتیبان ارایه شده توسط ویجایاباند و همکارانش [۲۶]	
	فراخوانی	دقت	فراخوانی	دقت
Normal	۸۱/۴۲	۸۱/۵۹	۸۰/۲۸	۸۰/۶۴
Probe	۸۸/۵۱	۸۸/۱۲	۸۶/۴۱	۸۳/۴۲
DOS	۸۷/۸۴	۸۸/۱	۸۷/۲۵	۸۴/۳۴
U2R	۸۹/۲۵	۸۹/۸۴	۸۹/۰۱	۸۷/۷۴
R2L	۸۸/۶۵	۸۹/۰۵	۸۸/۱۱	۸۵/۵۶
میانگین	۸۷/۱۳۴	۸۷/۳۴	۸۶/۲۱۲	۸۴/۳۴

نتایج نشان داده شده در جدول (۶) کارایی بالای روش پیشنهادی ترکیبی مبتنی بر اتوانکودر و ماشین بردار پشتیبان را نسبت به روش ویجایاباند برای همه کلاس‌های نفوذ اثبات می‌کند. اســـتفاده از هســـته های مختلف منجر به نتایج عملکردی متفاوت می‌شود.

همان‌گونه که در روش پی شنهادی بیان شد؛ در این مقاله از مجموع وزنی پنج ه سته چندجمله‌ای، چندجمله‌ای نرمال شده، تابع اســاس شــعاعی، تابع اســاس شــعاعی گاوســی و حلقوی اســتفاده می‌کنیم. از طرف دیگر هر کدام از این هســته‌ها دارای پارامترهای مربوط به خود هســتند که مقداردهی اولیه به آنها نیز بر عملکرد ماشــین بردار پشــتیبان تأثیر دارد. بنابراین، تعیین مقدار مناسب برای این پارامترها مسأله دیگری است که در مقاله به آن پرداخته شده است. به‌منظور مقداردهی مناسب به وزن و پارامترهای ه سته‌های مورد ا ستفاده، از الگوریتم ژنتیک بهره گرفته‌ایم که با توجه به ف ضای م سأله و پیچیدگی بالای م سأله مورد برر سی الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق مورد ا ستفاده در بهبود ما شین بردار پ شتیبان برای این منظور ارایه داده‌ایم. در این الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق مورد ا ستفاده در بهبود ما شین بردار پ شتیبان از خا صیت موازی بودن، پویا بودن عملگرها و پارامترهای با توپولوژی‌های مختلف ا ستفاده شده ا ست. هر زیرجمعیت در این روش قادر به اجرا بر روی یک پردازنده ا ست که با توجه به این‌که ما در این تحقیق ق صد داریم کاهش حجم محا سبات روش پی شنهادی را ن شان دهیم، تمام زیر جمعیت‌ها را دریک پردازنده در نظر می‌گیریم. از طرف دیگر، در شــکل اولیه ماکزیمم وزن هســته‌ها را در نظر می‌گیریم بدین صورت که دقت طبقه‌بندی با هسته‌ای که بیشترین وزن را دارد به عنوان نتیجه حاصل می‌دهیم.

به‌منظور بررسی کارایی الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق پیشنهادی، به مقایسه نتایج حاصل از آن با نتایج الگوریتم ژنتیک ساده پرداخته و نتایج حاصل از این مقایسه در شکل (۵) نمایش داده شده است. این نتایج نشان می‌دهد روش الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق دارای سرعت همگرایی بالاتری نسبت به الگوریتم ژنتیک است.



شکل (۵) نتایج حاصل از الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق با الگوریتم ژنتیک ساده

Figure (5): Results of self-adaptive island genetic algorithm with simple genetic algorithm

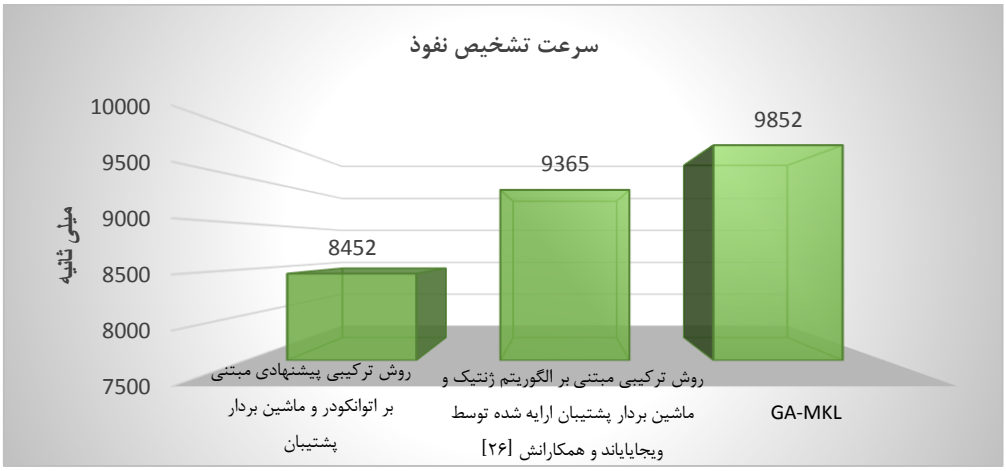
سرعت عامل دیگری برای تعیین کارایی روش پیشنهادی در سیستم‌های تشخیص نفوذ است. بنابراین، ما رویکردهای استفاده شده را براساس زمان سپری شده تجزیه و تحلیل می‌کنیم. نتایج مقایسه سرعت تشخیص روش پیشنهادی اول ماشین بردار چند هسته با الگوریتم ژنتیک ساده و روش مرجع [۲۶] با هسته تابع را در جدول (۷) نشان می‌دهد.

Table (7): Performance of approaches used based on time consumed

جدول(۷): عملکرد رویکردهای استفاده شده بر اساس زمان مصرف شده

روش پیشنهادی اول مبتنی بر اتوانکودر و ماشین بردار پشتیبان	روش مبتنی بر الگوریتم ژنتیک و ماشین بردار پشتیبان ارایه شده مرجع [۲۶]	GA-MKL
۸۴۵۲	۹۳۶۵	۹۸۵۲

همان‌طور که در جدول (۷) نشان داده شده است، روش پیشنهادی سرعت بالاتری نسبت به GA-SVM توسعه یافته توسط ویجایاباند و همکاران و روش GA-MKL دارد و روش پیشنهادی اول تقریبا ۹۱۳ میلی ثانیه سریع‌تر نسبت به روش مبتنی بر الگوریتم ژنتیک و ماشین بردار پشتیبان ارایه شده توسط ویجایاباند به جواب می‌رسد.



شکل (۶): ارزیابی سرعت تشخیص نفوذ روش پیشنهادی اول مبتنی بر اتوانکودر و ماشین بردار پشتیبان با مقایسه با روش مبتنی بر الگوریتم ژنتیک و ماشین بردار پشتیبان ارایه شده مرجع [۲۶]

Figure (6): Evaluation of intrusion detection speed of the first proposed method based on support vector atvankodro machine by comparison with the method based on genetic algorithm and backup vector machine [26]

در الگوریتم ژنتیک ساده انتخاب ویژگی برای هر نسل تکامل در رویکردهای GA-SVM و GA-MKL انجام می‌شود که این امر برای ارزیابی و بهینه‌سازی وقت‌گیر است، در حالیکه در روش پیشنهادی اول فقط یک بار همجوشی ویژگی انجام می‌شود. بنابراین سرعت بالاتری برای تشخیص نفوذ دارد. با توجه به این‌که مسئله مورد بررسی در این تحقیق یک مسئله تشخیص نفوذ است پس کارایی آن را در این حیطه نیز در شکل (۶) مورد ارزیابی قرار می‌دهیم. برای این منظور کار تحقیقاتی گاتاما رامان و همکارانش [۲۷] را در کنار روش ویجایاباند و همکارانش [۲۶] که قبلا مورد بررسی و مقایسه قرار گرفت، در نظر می‌گیریم. نتایج حاصل از ماشین بردار پشتیبان نشان می‌دهد که هسته تابع شعاعی عملکرد بهتری دارد و قادر است با دقت ۹۹/۱ درصد نفوذ را تشخیص دهد درحالی‌که این مقدار برای روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق برابر ۹۹/۸ درصد است که ۷ درصد جواب بهتری ایجاد می‌کند. برای اعتبار بیشتر دادن به ارزیابی صورت گرفته آن را از نظر تشخیص نفوذ کلاس‌های مختلف حملات مورد بررسی قرار داده و در جدول (۸) این نتایج نمایش داده شده است. نتایج نشان داده شده نیز حاکی از عملکرد بالاتر روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای پویای خودتطبیق در تشخیص نفوذ انواع کلاس نفوذ نسبت به روش مقداردهی بهین به پارامترهای هسته تابع اساس شعاعی و انتخاب مشخصه توسط الگوریتم ژنتیک ارایه شده توسط گاتاما رامان و همکارانش [۲۷] است.

۴- نتیجه‌گیری

در این مقاله مسئله را به‌صورت استفاده از چندین هسته به‌طور هم‌زمان و وزن‌دهی متفاوت به هر یک از هسته‌ها و مقادیر مختلف پارامترها مدل‌سازی کرده‌ایم. با توجه به این‌که این مسئله از پیچیدگی بالایی برخوردار است، روش‌های بهینه‌سازی به شکل مرسوم قادر به حل آن نیستند؛ بنابراین ما یک الگوریتم ژنتیک جزیره‌ای پویای خود تطبیق برای حل آن پیشنهاد کرده و از طرف دیگر با توجه به حجم بالای داده‌ها در این‌گونه مسائل، از اتوانکودر نیز برای کاهش حجم داده‌ها استفاده شد. با توجه به این که ما در روش پیشنهادی بر مسأله کار با حجم بالای داده‌ها و دقت تشخیص تمرکز کرده‌ایم، نتایج حاصل را از دو دیدگاه کاهش ابعاد داده‌ها و بهبود دقت مورد تحلیل قرار دادیم. عملکرد کار با حجم بالای داده‌های روش پیشنهادی که از کاهش ابعاد

با اتوانکودر بهره گرفته است را بدون استفاده از هسته‌های مختلف با روش کاهش ابعاد داده ارایه شده توسط مرجع [۲۶] نتایج بهبود عملکرد روش پیشنهادی تا حدود ۳/۵۳ درصد را نشان می‌دهد. دیدگاه دیگر دقت تشخیص است که ما با مقایسه روش پیشنهادی با مرجع [۲۷] تحلیل شد و نتایج این مقایسه نیز نشان دهنده بهبود دقت به اندازه تقریبا ۷ درصد است.

Table (8): Evaluation of the proposed hybrid method based on autowanker and improved support vector machine with Self-Adaptive Genetic Algorithm and reference method based on the accuracy of different intrusion class detection

جدول (۸): ارزیابی روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره‌ای پویای خودتطبیق و روش مرجع [۲۷] بر اساس دقت تشخیص کلاس نفوذهای مختلف

کلاس	روش ترکیبی پیشنهادی مبتنی بر اتوانکودر و ماشین بردار پشتیبان بهبودیافته با الگوریتم ژنتیک جزیره ای پویای خودتطبیق		روش مقداردهی بهین به پارامترهای هسته تابع اساس شعاعی و انتخاب مشخصه توسط الگوریتم ژنتیک ارایه شده توسط گاتاما رامان و همکارانش [۲۷]	
	فراخوانی	دقت	فراخوانی	دقت
Normal	۹۹/۹۴۱	۹۹/۹۶۷	۹۹/۹۱۱	۹۹/۹۰۹
Probe	۹۹/۹۴۷	۹۹/۹۸۹	۹۹/۷۹۵	۹۹/۹۰۱
DOS	۹۹/۹۴۲	۹۹/۹۸۷	۹۹/۸۵۴	۹۹/۹۳۱
U2R	۹۹/۹۲۱	۹۹/۹۸۴	۹۹/۷۴۸	۹۹/۸۹۹
R2L	۹۹/۹۲	۹۹/۹۷۵	۹۹/۹۳۸	۹۹/۹۱۴
میانگین	۹۹/۹۳۴۲	۹۹/۹۸۰۴	۹۹/۸۴۹۲	۹۹/۹۱۰۸

References

مراجع

[1] A. Almomani, M. Alauthman, F. Albalas, O. Dorgham, A. Obeidat "An online intrusion detection system to cloud computing based on NeuCube algorithms", International Journal of Cloud Applications and Computing, vol. 8, no. 2, pp.1042-1059, 2018 (doi:10.4018/IJCAC.2018040105).

[2] S. A. Mulay, P. Devale, G. Garje, "Intrusion detection system using support vector machine and decision tree", International Journal of Computer Applications,vol. 3, no.3, pp. 40-43, 2010 (doi:10.5120/758-993).

[3] W. Laftah Al-Yaseen,"Improving intrusion detection system by developing feature selection model based on firefly algorithm and support vector machine", IAENG International Journal of Computer Science, vol. 46, no. 4, pp. 534-540, 2019 (doi: IJCS_46_4_04).

[4] M. R. G. Raman, N. Somu, S. Jagarapu, T. Manghnani, T. Selvam, K. Krithivasan,V. S. S. Sriram, "An efficient intrusion detection technique based on support vector machine and improved binary gravitational search algorithm", Artificial Intelligence Review, vol. 53, pp. 3255-3286, 2019 (doi:10.1007/s.10462-019-09762-z).

[5] M. Ramkumar, M. Manikandan, K. Sathish Kumar, R. K. Kumar, "Intrusion detection in manets using support vector machine with ant colony optimization" ICTACT journals on data csience and machine learning, vol. 1, no.1, 2019.

[6] J. C. Badajena, C. Rout, "Incorporating hidden markov model into anomaly detection technique for network intrusion detection", International Journal of Computer Applications, vol. 53, no. 11, 2012 (doi: 10.5120/84-69-2395).

[7] P. Dorogovs, A. Borisov, A. Romanovs, "Building an intrusion detection system for it security based on data mining techniques", Applied Computer Systems, vol. 45, no. 1, pp. 43-48, 2011 (doi: 10.2478/v10143-011-0040-3).

[8] S. Shirbhate, S. Sherekar and, V. Thakare, " Performance evaluation of PCA filter in clustered based intrusion detection system", Proceeding of the IEEE/ICESC, pp. 217-221, Nagpur, India ,Feb. 2014 (doi: 10.1109/IC-ESC.2014.100).

[9] D. Gupta, S. Singhal, S. Malik, A. Singh, "Network intrusion detection system using various data mining techniques", Proceeding of the IEEE/(RAINS), pp. 1-6, Bangalore, India, May. 2016 (doi: 10.1109/RAIN-S.2016.7764418).

[10] E. Ariafar, R. Kiani, "Intrusion detection system using an optimized framework based on datamining techniques", Proceeding of the IEEE/KBEI, pp. 0785-0791, Tehran, Iran, Dec. 2017 (doi: 10.1109/KBEI.201-7.8324903).

[11] J. A. Sukumar, I. Pranav,MM. Neetish, J. Narayanan, "Network intrusion detection using improved genetic k-means algorithm", Proceeding of the IEEE/ICACCI, pp. 2441-2446, Bangalore, India, Sept. 2018 (doi: 10.-1109/ICACCI.2018.8554710).
[12] P. S. Bhattacharjee, A. K. M, Fujail, A. A. Begum, "Intrusion detection system for NSL-KDD data set using vectorised fitness function in genetic algorithm", Advances in Computational Sciences and Technolog.,vol. 10, no. 2, pp. 235-246, 2017.
[13] J. Ghasemi, J. Esmaeily, R. Moradinezhad, "Intrusion detection system using an optimized kernel extreme learning machine and efficient features", Sådhanå,vol. 45, no. 2, pp.1-9, 2020(doi: 10.1007/s12046-019-1230-x).
[14] D. Pal, A. Parashar, "Improved genetic algorithm for intrusion detection system", Proceeding of the IEEE/CI-CN, pp. 835-839, Bhopal, India, Nov. 2014 (doi: 10.1109/CICN.2014.178).
[15] Y. Danane, T. Parvat, "Intrusion detection system using fuzzy genetic algorithm", Proceeding of the IEEE/ ICPC , pp. 1-5, St. Louis, Missouri, USA, March. 2015 (doi: 10.1109/PERVASIVE.2015.7086963).
[16] A. F. A. Pinem, E. B. Setiawan, "Implementation of classification and regression tree (CART) and fuzzy logic algorithm for intrusion detection system", Proceeding of the IEEE/ ICoICT, pp. 266-271, Bali, Indonesia, May. 2015 (doi: 10.1109/ICoICT.2015.7231434).
[17] S. Sahu, B. M. Mehtre, "Network intrusion detection system using J48 decision tree", Proceeding of the IEEE/ ICACCI, pp. 2023-2026, Kochi, India, Aug. 2015 (doi: 10.1109/ICACCI.2015.7275914).
[18] S. M. H. Bamakan, H. Wang, Y. Shi, " Ramp loss k-support vector classification-regression; a robust and sparse multi-class approach to the intrusion detection problem", Knowledge-Based Systems, vol.126, pp. 113-126, 2017 (doi: 10.1016/j.knosys.2017.03.012).
[19] C. A. Catania, C. G. Garino, "Automatic network intrusion detection: Current techniques and open issues", Computers & Electrical Engineering, vol. 38, no. 5, pp. 1062-1072, 2012 (doi: 10.1016/j.compeleceng.2012.-05.013).
[20] S. Aljawarneh, M. Aldwairi, M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model", Journal of Computational Science, vol. 25, pp. 152-160, 2018 (doi: 10.1016/j.jocs.2017.03.006).
[21] G. Sandhya, A. Julian, "Intrusion detection in wireless sensor network using genetic K-means algorithm", Proceeding of the IEEE/ ICACCCT, pp. 1791-1794, Ramanathapuram, India, May. 2014 (doi: 10.1109/IC-ACCCT.2014.7019418).
[22] M. Sharma, K. Jindal, A.Kumar, "Intrusion detection system using Bayesian approach", International Journal of Computer Application, vol. 48, no.5, pp. 29-33,2012.
[23] G. Sandhya, A. Julian, "Intrusion detection in wireless sensor network using genetic K-means algorithm", Proceeding of the IEEE/ ICACCCT pp. 1791-1794, Ramanathapuram, India, May. 2014 (doi: 10.1109/ICA-CCCT.2014.7019418).
[24] T. Yerong, S. Sai, X. Ke, L. Zhe, "Intrusion detection based on support vector machine using heuristic genetic algorithm", Proceeding of the IEEE/CSNT, pp. 681-684, Bhopal, India,Apr. 2014 (doi: 10.1109/CSNT.201-4.143).
[25] Q. Schueller, K. Basu, M. Younas, M. Patel, F. Ball, "A hierarchical intrusion detection system using support vector machine for SDN network in cloud data center", Proceeding of the IEEE/ITNAC), pp. 1-6. Sedney, Australia, Nov. 2018 (doi: 10.1109/ATNAC.2018.8615255).
[26] R. Vijayanand, D. Devaraj, B. Kannapiran, "Intrusion detection system for wireless mesh network using multiple support vector machine classifiers with genetic-algorithm-based feature selection", Computers and Security, vol. 77, pp. 304-314, 2018 (doi: doi.org/10.1016/j.cose.2018.04.010).
[27] M. G. Raman, N. Somu, K. Kirthivasan, R. Liscano, V. S. S. Sriram, "An efficient intrusion detection system based on hypergraph-genetic algorithm for parameter optimization and feature selection in support vector machine", Knowledge-Based Systems, vol. 134, pp. 1-12, 2011 (doi: 10.1016/j.knosys.2017.07.005).
[28] S. Mirjalili, "Genetic algorithm", Evolutionary Algorithms and Neural Networks, Part of the Studies in Computational Intelligence Book Series (SCI), vol.780, pp. 43-55, 2019 (doi: 10.1007/978-3-319-93025-1_4)
[29] M. Gharaibeh, C. Papadopoulos, "Darpa-2009 intrusion detection dataset report", Tech. Rep., 2014.

زیر نویس‌ها

1. Principal component analysis
2. Linear regression
3. Improved genetic K-means algorithm (IGKM)
4. Intrusion detection system (IDS)
5. Fuzzy-genetic

6. Confusion matrix
7. Time-varying chaos particle swarm optimization (TVCPSO)
8. Multiple criteria linear programming (MCLP)
9. Support Vector Machine (SVM)
10. Chaotic
11. Misuse
12. Vote algorithm
13. Information gain
14. Decision tree
15. Feature quantile filter
16. Clique property
17. Receiver operating characteristic (ROC)
18. Auto encoder
19. Supported vector machine (SVM)
20. Multi kernel learning (MKL)
21. Radial basis function (RBF)
22. Sigmoid
23. Island genetic algorithm (IGA)
24. True normal
25. False positive
26. False normal
27. True positive