هفتمین کنفرانس ملی مباحث نوین در کامپیوتر و فناوری اطلاعات
7th National Conference on Advanced Topics in Computer and Information Technology
۲ آذر ماه ۱۴۰۲

استفاده از الگوریتم سنجاب پرنده جهت تشخیص انواع نرم‌افزارهای بدخواه در سیستم‌های کامپیوتری

علیرضا صفائی
گروه کامپیوتر، واحد ماهشهر، دانشگاه آزاد اسلامی، ماهشهر، ایران
alisafasafaei@yahoo.com

چکیده : در سال‌های اخیر یکی از مهمترین چالش‌های امنیت اطلاعات و شبکه‌های ارتباطی افزایش روزافزون انواع نرم‌افزارهای بدخواه و به دنبال آن یافتن راه‌های مناسب جهت حفاظت سیستم‌ها در مقابل آنهاست که از مهمترین دغدغه‌های برنامه‌نویسان و متخصصین امنیت اطلاعات شناخت به موقع و یافتن راه‌های مقابله با اثرات مخرب این‌گونه نرم‌افزارهای بدخواه می‌باشد که در مقابله با این روند شرکت‌های ضدنرم‌افزارهای بدخواهی به طور جدی به دنبال راه‌های بهتری برای شناسایی نرم‌افزارهای بدخواه هستند در این راستا ایجاد روشی که بتواند به‌صورت کارا به شناسایی و جلوگیری از نفوذ نرم‌افزارهای بدخواه بپردازد همواره مورد نیاز خواهد بود. استفاده از تکنیک‌های داده‌کاوی و الگوریتم‌های هوش جمعی یکی از روش‌هایی است که محققان توانسته‌اند در سال‌های اخیر با استفاده از آن به تشخیص نرم‌افزارهای بدخواه با دقت‌های مختلفی بپردازند و در این راه به پیشرفت‌هایی در تشخیص نرم‌افزارهای بدخواه دست یافته‌اند. لذا در این تحقیق سعی کردیم تا با استفاده از الگوریتم سنجاب پرنده به تشخیص انواع نرم‌افزارهای بدخواه در سیستم‌های کامپیوتری بپردازیم.

کلمات کلیدی: تشخیص – نرم‌افزارهای بدخواه – سیستم‌های کامپیوتری - الگوریتم سنجاب پرنده

۱ – مقدمه

در دنیای امروز سیستم‌های کامپیوتری به جز ضروری زندگی ما تبدیل شده‌اند. در کنار مزیت‌های فراوان این سیستم‌ها، فرصت‌هایی برای سوء استفاده‌کنندگان به وجود آمده است. آنها با طراحی نرم افزارهای مخرب[1] که نرم‌افزارهای بدخواه[2] نامیده می‌شوند سعی در آسیب رساندن به سیستم‌ها و امنیت کاربران دارند. نرم‌افزارهای بدخواه، مجموعه‌ای از دستورالعمل‌هایی هستند که در نرم‌افزار کاربردی قرار داده می‌شوند و امنیت سیستم را بدون اطلاع صاحب آن تهدید می‌کنند[۱]. طبق گزارشات انجام شده تعداد نرم‌افزارهای بدخواه هر ساله بسیار افزایش یافته است به نحوی که هر ۴.۶ ثانیه یک نمونه نرم‌افزارهای بدخواه جدید در سال پیدا خواهد شد، ویندوز بعنوان پلتفرم اصلی برای نرم‌افزارهای بدخواه بیش از ۹۹ درصد از انواع نرم‌افزارهای بدخواه را به خود اختصاص می‌دهد[۲]. باتوجه به اینکه مراکز دفاع سایبری اهمیت بسیار زیادی در اکثر کشورها دارد و همانند مرزهای یک کشور که می‌توانند مورد تهدید و هجوم قرار گیرند لذا در فضای مجازی نیز این امکان وجود دارد که مورد هجوم این تهدیدها قرار گرفته شود[۳]. باتوجه به مطالعات صورت گرفته می‌توان گفت اغلب حمله‌ها در حوزه فضای مجازی توسط نرم‌افزارهای بدخواه صورت می‌گیرد که تشخیص به موقع حمله‌های امنیتی در فضای مجازی از اهمیت بالایی در حفاظت از منابع برخوردار است. بنابراین برای تشخیص چنین نرم‌افزارهای بدخواهیی نیاز به روش‌هایی جهت تشخیص رفتارهای خوب و بد نرم‌افزار است تا مشخص گردد کدام نرم‌افزار مشکل‌زا و کدام مشکل‌زا نیست و بتوانیم از آن نرم‌افزار استفاده کنیم[۴]. بنابراین در سال‌های اخیر یکی از مهمترین چالش‌های امنیت اطلاعات و شبکه‌های ارتباطی، افزایش روز افزون انواع نرم‌افزارهای بدخواه و به دنبال آن یافتن راه‌های مناسب جهت حفاظت سیستم‌ها در مقابل آنهاست که از مهمترین دغدغه‌های

[1] Malicious sofware
[2] malware

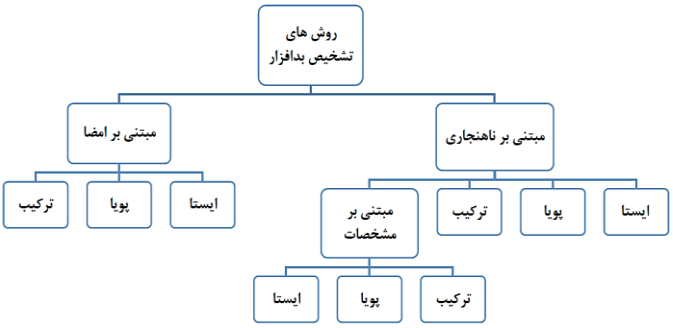
برنامه‌نویسان و متخصصین امنیت اطلاعات، شناخت به موقع و یافتن راه‌های مقابله با اثرات مخرب این‌گونه نرم‌افزارهای بدخواه می‌باشد[۵]. در مقابله با این روند، شرکت‌های ضدنرم‌افزارهای بدخواهی به طور جدی به دنبال راه‌های بهتری برای شناسایی نرم‌افزارهای بدخواه هستند[۶].

۲- نرم‌افزارهای بدخواه

نرم‌افزار بدخواه یا بدافزار یکی از ابزارهای اقدامات ضد امنیتی می‌باشد که بدون اجازه صاحب سیستم قصد انجام کارهای ناخواسته یا خرابکارانه را در سیستم دارد. اصطلاح نرم‌افزارهای بدخواه از دو واژه بدخواهانه[۳] و نرم‌افزار[۴] تشکیل شده است. از دیدگاه تخصصی نرم‌افزارهای بدخواه در حالت کلی برنامه‌ای کامپیوتری بوده که با هدف آسیب‌رسانی به سیستم‌های کامپیوتری طراحی شده است. نرم‌افزارهای بدخواه علاوه بر اینکه با عملیات خرابکارانه به منابع و اطلاعات سیستم‌ها دسترسی پیدا می‌کنند به سرویس‌دهی و فعالیت‌های سیستم‌ها آسیب وارد کرده و با سرقت غیرمجاز اطلاعات شخصی افراد؛ سبب وارد کردن هزینه‌های گزاف به قربانیان می‌شوند. از اوایل پیدایش نرم‌افزارهای بدخواه نبرد بزرگی بین نرم‌افزارهای بدخواه و ضد نرم‌افزارهای بدخواه به وجود آمده است به طور کلی می‌توان گفت جامعه نرم‌افزارهای بدخواه همیشه از رقیب دیرینه‌شان یعنی ضد نرم‌افزارهای بدخواه یک قدم جلوتر هستند. از انواع نرم‌افزارهای بدخواه می‌توان به اسب تروا، نرم‌افزارهای جاسوسی، درهای پشتی، روت کیت و کرم اشاره کرد این نرم‌افزارهای بدخواه وقتی که راهی برای ورود به سیستم پیدا کنند سیستم را آلوده کرده و به سرعت به منظور یافتن قربانیان جدید شروع به اسکن مجدد کرده و خود را درون شبکه تکثیر می‌کنند. یکی از بزرگترین چالش‌ها در این زمینه مقابله با سرعت زیاد انتشار نرم‌افزارهای بدخواه است زیرا یک سیستم آلوده به نرم‌افزارهای بدخواه قادر خواهد بود صدها سیستم دیگر را آلوده کند و هرکدام از این صدها سیستم دیگر نیز می‌توانند سیستم‌های دیگری را آلوده نمایند [۷].

۳- روش‌های تشخیص نرم‌افزارهای بدخواه

روش‌های تشخیص نرم‌افزارهای بدخواه برای تشخیص بدافزارها مورد استفاده قرار گرفته و از آلوده شدن سیستم‌های کامپیوتری و از دست رفتن اطلاعات و سازش سیستم جلوگیری می‌کنند بطور کلی تشخیص نرم‌افزارهای بدخواه به دو دسته کلی تشخیص مبتنی بر امضا[۵] و تشخیص مبتنی بر ناهنجاری[۶] تقسیم می‌شود هریک از روش‌های تشخیص نرم‌افزارهای بدخواه خود شامل روش‌های پویا[۷]،روش‌های ایستا[۸] و روش‌های ترکیب[۹] هستند. در شکل (۱) طبقه‌بندی روش‌های تشخیص بدافزار یا نرم‌افزار بدخواه بطورکلی به نمایش درآمده است [۸].



شکل ۱: طبقه‌بندی انواع روش‌های تشخیص نرم‌افزارهای بدخواه

روش مبتنی بر امضا براساس بررسی کدهای مشکوک و جمع‌آوری اطلاعات به منظور توصیف اهداف مغرضانه نرم‌افزارهای مخرب است هدف اصلی این روش استخراج توالی بایت‌های ویژه از کدها به عنوان امضا است همچنین یکی دیگر از اهداف این روش جستجو برای یک امضا در داخل فایل‌های مشکوک است. بیشتر ضدنرم‌افزارهای بدخواهی تجاری امروزی از مجموعه‌ای از امضاها جهت تشخیص برنامه‌های مخرب استفاده می‌کنند این کدهای مشکوک با یک توالی منحصر به فرد از ساختارهای برنامه یا بایت‌ها مقایسه می‌شود اگر این امضا در داخل پایگاه داده وجود نداشته باشد به این معناست که این فایل خوش‌خیم بوده و آلوده به نرم‌افزارهای بدخواه نیست بنابراین یکی از محدودیت‌هایی که در روش تشخیص مبتنی بر امضا وجود دارد نیاز به دخالت انسان در بروزرسانی پایگاه داده امضا با امضاهای جدید است همچنین گروهی از محققان نیز نشان دادند که برخی از نویسندگان نرم‌افزارهای بدخواهی چندریختی قادر به شکست روش مبتنی بر امضا بوسیله تکنیک‌های مبهم‌سازی هستند بنابراین این روش تشخیص مستعد ابتلا به منفی کاذب[۱۰] است [۹،۱۰].

تشخیص مبتنی بر ناهنجاری عموما در دو فاز آموزش و تشخیص یا

[3] Malicious
[4] Software
[5] signature-based methods
[6] anomaly-based methods
[7] dynamic methods
[8] satic methods
[9] hybrid methods
[10] False negative

نظارت اتفاق می‌افتد. در طول فاز آموزش کشف کننده سعی در یادگیری رفتار نرمال دارد. کشف کننده در طول فاز آموزش قادر است رفتار را از میزبان یا برنامه تحت نظارت یا ترکیبی از هر دو یاد بگیرد. یکی از مزایای این روش توانایی تشخیص حملات صفر-روز می‌باشد اما از محدودیت‌های این روش می‌توان به نرخ کاذب بالای آن و پیچیدگی در تعیین اینکه کدام ویژگی باید در فاز آموزش، مورد آموزش قرار داده شود می‌باشد در واقع باتوجه به اینکه در طول فاز آموزش امکان نمایش رفتارهایی که قبلا دیده نشده است برای یک سیستم صفر نیست لذا احتمال اینکه یک روش مبتنی بر ناهنجاری مثبت کاذب را بالا ببرد نیز صفر نخواهد بود بنابراین می‌توان گفت فاز تشخیص برنامه تحت بازرسی را در زمان اجرا مورد بررسی قرار داده و آن را باتوجه به آنچه که در فاز آموزش یاد گرفته است از نظر ناپایداری مورد بررسی قرار می‌دهد به عبارتی دیگر می‌توان گفت روش تشخیصی مبتنی بر ناهنجاری جهت تصمیم‌گیری مخرب بودن یا نبودن یک برنامه تحت بررسی از دانش خود برای اینکه چه چیزی رفتار عادی را تشکیل داده است استفاده می‌کند [۱۱].

جدول ۱: مزایا و معایب روش‌های مبتنی بر امضا و ناهنجاری

تجزیه و تحلیل	مزایا	معایب
مبتنی بر ناهنجاری	زمان اسکن بسیار پایین است	ناتوان در تشخیص نرم‌افزارهای بدخواه ناشناخته
	میزان نرخ مثبت غلط پایین است	توانایی اندک در شناسایی موارد مبهم و گیج‌کننده
مبتنی بر امضا	ارائه بهترین نتایج در تشخیص نرم‌افزارهای بدخواهی چندریختی	قادر به تشخیص بسیاری از ویروس‌های بسته‌بندی شده نیست

۵– داده کاوی

بطور عمومی داده‌کاوی به فرآیند استخراج خودکار مدل‌ها از میان انبوه داده‌ها اطلاق می‌شود. اخیراً پیشرفت‌های قابل توجهی در زمینه داده‌کاوی به دست آمده است و داده‌کاوی مجموعه الگوریتم‌های گسترده‌ای را از علومی چون آمار، تشخیص الگو، یادگیری ماشین و پایگاه داده‌ها بکار گرفته است. داده‌کاوی دو نوع هدایت شده و غیرهدایت شده می‌باشد داده‌کاوی هدایت شده دارای هدفی خاص و از پیش تعیین شده دارد که به دنبال الگویی خاص است در حالی که هدف داده‌کاوی غیرهدایت شده یافتن الگوها یا شباهت‌های بین گروه‌هایی از اطلاعات بدون داشتن هدفی خاص و یا مجموعه‌ای از دسته‌ها و الگوهای از پیش تعیین شده است. در داده‌کاوی هدایت شده همیشه یک متغیر هدف دارد چیزی که باید دسته‌بندی، تخمین یا پیش‌بینی گردد اما در داده‌کاوی غیرهدایت شده متغیر هدفی وجود نداریم و وظیفه داده‌کاوی یافتن الگوهای کلیدی که متعلق به یک متغیر خاص نیستند می‌باشد [۱۲]. از رایج‌ترین مدل‌های داده‌کاوی طبقه‌بندی می‌باشد که می‌تواند با استفاده از نمونه‌های از قبل طبقه‌بندی شده برای توسعه یک مدل تعداد زیادی نمونه را طبقه‌بندی کند. فرایند طبقه‌بندی داده‌ها شامل یادگیری و طبقه‌بندی می‌باشد؛ در یادگیری داده‌های آموزشی توسط الگوریتم طبقه‌بندی تجزیه و تحلیل می‌شوند اما در طبقه‌بندی داده‌های ارزیابی برای تخمین دقت طبقه‌بندی مورد استفاده قرار می‌گیرند. اگر دقت تخمین زده شده توسط طبقه‌بندی قابل قبول باشد مجموعه داده‌های جدید را نیز می‌توان به طبقه‌بندی اعمال کرد. برای بکارگیری تکنیک‌های داده کاوی در سیستم‌های تشخیص نرم‌افزارهای بدخواه متخصصان داده کاوی روش‌های مختلفی را بکار می‌گیرند[۱۳].

۶ – شبکه عصبی مصنوعی

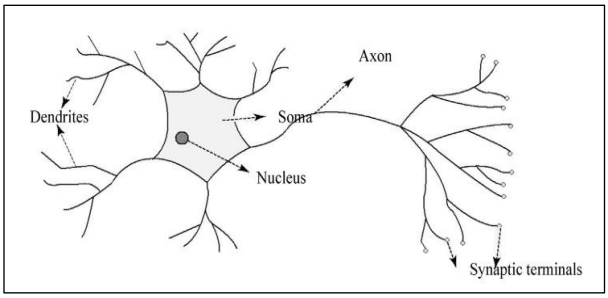
در سال ۱۹۴۳ شبکه‌های عصبی مصنوعی برای اولین بار توسط مک کولاک و پیتس معرفی شدند. فلسفه اصلی شبکه عصبی مصنوعی، مدل کردن ویژگی‌های پردازشی مغز انسان برای تقریب زدن روش‌های معمول محاسباتی با روش پردازش زیستی است به بیان دیگر شبکه عصبی مصنوعی روشی است که دانش ارتباط بین چند مجموعه داده را از طریق آموزش فرا گرفته و برای استفاده در موارد مشابه ذخیره می‌کند. این پردازنده از دو جهت مشابه مغز انسان عمل می‌کند:

- یادگیری شبکه عصبی از طریق آموزش صورت می‌گیرد
- وزن‌دهی مشابه با سیستم ذخیره‌سازی اطلاعات در شبکه عصبی مغز انسان انجام می‌گیرد

شبکه عصبی مصنوعی یکی از تکنیک‌های پرکاربرد داده‌کاوی در کشف دانش و تشخیص الگوهای پنهان به شمار می‌رود که براساس مغز و سیستم عصبی انسان الگوبرداری شده است. شبکه عصبی ازشبیه‌سازی نورون عصبی[11] برای تقلید رفتار مغز انسان استفاده می‌کند. نورون‌های عصبی سلول‌های طویل در سیستم عصبی می‌باشند که وظیفه آنها

[11] Neurons

انتقال و پردازش اطلاعات دریافتی توسط سایر نورون‌های شبکه عصبی مصنوعی بوده و در شکل (۲) شان داده شده است [۱۴]:

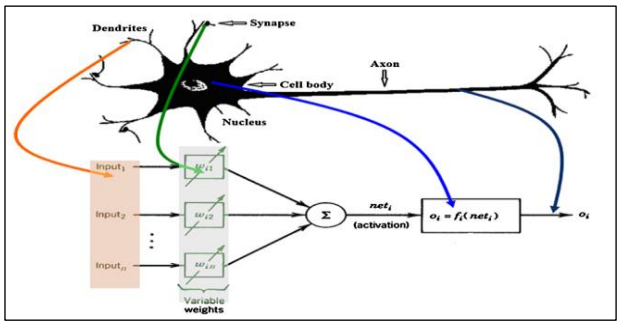


شکل ۲: یک نورون عصبی

یک نورون عصبی شامل اجزاء و بخش‌های ذیر جهت دریافت، پردازش و انتقال داده‌ها می‌باشد:

۱. دندریت[۱۲]: یک نورون عصبی توسط دندریت‌ها می‌تواند پیام‌های عصبی را از سایر نورون‌های عصبی دریافت و تحویل سایر قسمت‌های سلول عصبی نماید.
۲. جسم سلولی[۱۳]: این جزء را می‌توان محلی برای تقویت یا تضعیف سیگنال‌های دریافتی هر نورون به شمار آورد.
۳. آکسون[۱۴]: توسط این بخش می‌توان پیام‌های عصبی هر نورون را به سایر سلول‌های بدن انتقال داد.

در شکل(۳) مفهوم یک سلول عصبی برای مدل‌سازی یک شبکه عصبی مصنوعی نشان داده شده است:



شکل ۳: مفهوم یک سلول عصبی و شبکه عصبی مصنوعی

شکل (۳) نشان می‌دهد که دندریت سلول عصبی معادل ورودی‌های شبکه عصبی مصنوعی، فضای سیناپسی[۱۵] معادل ضرایب به کار رفته در یال‌های شبکه، تابع فعالیت[۱۶] معادل جسم سلولی و آکسون معادل خروجی یک شبکه عصبی مصنوعی تعریف می‌شوند. برای آنکه یک شبکه عصبی مصنوعی بتواند تصمیم‌گیری خود را از حالت خطی[۱۷] به حالت غیرخطی[۱۸] تبدیل نماید از مفاهیم نگاشت[۱۹] استفاده می‌نماید که ورودی‌های شبکه و خروجی‌های نورون‌های عصبی را تحت تأثیر خود قرار می‌دهد. شبکه عصبی مصنوعی چندلایه یکی از انواع شبکه عصبی مصنوعی به شمار می‌رود که نورون‌های مصنوعی در قالب لایه‌های پنهان آن سازماندهی شده‌اند در این نوع شبکه‌ها ورودی هر نورون تحت تأثیر ضرایبی به نام اوزان[۲۰] قرار گرفته و در نهایت مجموع ضرایب با هم جمع شده و با مقدار مشخصی در هر لایه به نام آستانه‌ها[۲۱] جمع نهایی می‌شوند. استفاده از ضرایب عددی و جمع کننده آستانه در واقع شبیه‌سازی شدت سیگنال‌های انتقال یافته بین نورون‌های مصنوعی شبکه است[۱۵].

۷– الگوریتم سنجاب پرنده

الگوریتم سنجاب پرنده[۲۲] به عنوان یک الگوی بهینه‌سازی جدید با الهام از طبیعت ارائه شده است[۱۶]. این بهینه‌ساز از رفتار کاوشگرایانه پویا از سنجاب‌های پرنده[۲۳] جنوبی و روش موثر حرکت آنها که همچون یک گلایدر است تقلید می‌کند. پریدن[۲۴]، مکانیسم موثری است که توسط پستانداران کوچک بر ای پیمودن مسافت‌های طولانی استفاده می‌شود. روند جستجو وقتی شروع می‌شود که سنجاب‌ها شروع به کاوش می‌کنند. در طول فصل پاییز که دمای هوا برای جستجوی منابع غذایی مساعد است سنجاب‌ها با پریدن از یک درخت به درخت دیگر منابع غذایی را جستجو می‌کنند. در حالی که این کار را انجام می‌دهند مکان خود را تغییر داده و مناطق مختلف جنگل را کشف می‌کنند. از آنجا که شرایط آب و هوایی به اندازه کافی گرم است آنها می‌توانند نیازهای انرژی روزانه خود را با سرعت بیشتری در رژیم غذایی میوه بلوط به وفور یافت کنند و از این رو بلافاصله پس از پیدا کردن میوه بلوط[۲۵] آن را مصرف می‌کنند.

[12] Dendrites
[13] Nucleus
[14] Axon
[15] Synaptic space
[16] Activity Function
[17] Linear
[18] Nonlinear
[19] Kernel
[20] Weights
[21] Threshold
[22] Squirrel Search Algorithm (SSA)
[23] Flying Squirrels (FS)
[24] Gliding
[25] Acorn nut tree

آنها پس از تحقق نیاز روزانه انرژی خود به جستجوی منبع بهینه مواد غذایی برای زمستان (آجیل‌های هیکوری[۲۶] (درخت گردوی آمریکایی)) می‌پردازند. ذخیره آجیل‌های هیکوری به آنها کمک می‌کند تا انرژی خود را در هوای بسیار سخت حفظ کنند و باعث کاهش هزینه‌های کاوش و در نتیجه افزایش احتمال بقا شوند. در طول زمستان از بین رفتن پوشش برگ در جنگل‌ها باعث افزایش خطر شکار می‌شود و از این رو سنجاب‌ها کمتر فعال می‌شوند اما بطور کامل در زمستان خواب نمی‌روند. در پایان فصل زمستان سنجاب‌ها دوباره شروع به فعالیت می‌کنند. این روند تکراری است و در طول عمر یک سنجاب ادامه دارد و پایه و اساس الگوریتم جستجوی سنجاب را تشکیل می‌دهد.

۸- پیشینه تحقیق

در این قسمت به کارهای انجام شده مرتبط با موضوع کار که توسط محققان داخلی و خارجی صورت گرفته می‌پردازیم. Jamal و همکاران در سال ۲۰۲۲ استفاده از الگوریتم‌های نزدیکترین همسایگی، NaïveBayes و شبکه عصبی مصنوعی جهت تشخیص آلودگی نرم‌افزارهای بدخواهی در شبکه اینترنت اشیا را پیشنهاد کردند. نتایج بدست آمده با استفاده از مجموعه داده دریافتی از سایت kaggle که شامل ۱۰۰۰۰۰ رکورد و ۳۲ ویژگی می‌باشد نشان داد که شبکه عصبی با دقت ۹۷.۰۸ درصد قادر به تشخیص آلودگی نرم‌افزارهای بدخواه است[۱۷]. Siahmarzkooh در سال ۲۰۲۱ جهت تشخیص آلودگی نرم‌افزارهای بدخواه از الگوریتم درخت تصمیم و جهت انتخاب ویژگی‌های مهم از الگوریتم kmeans و شبکه بیزین استفاده کرد. نتایج بدست آمده با استفاده از مجموعه داده دریافتی از سایت kaggle که شامل ۱۰۰۰۰۰ رکورد و ۳۲ ویژگی می‌باشد نشان داد که روش پیشنهادی قادر است با دقت ۹۴.۶۵ درصد انواع آلودگی نرم‌افزارهای بدخواه را تشخیص دهد[۱۸]. صف‌آرا و همکاران در [illegible] تشخیص وب‌سایت‌های جعلی از الگوریتم فرا[illegible] کفتار به [illegible] انتخاب ویژگی‌های مناسب جهت طبقه‌بندی وب[illegible]های جعلی است [illegible]کردند. نتایج بدست آمده با استفاده از مجموعه [illegible] دریافتی از سایت kag[illegible] که شامل ۱۰۰۰۰۰ رکورد و ۳۲ ویژ[illegible]د نشان [illegible]وش پیشنهادی قادر است با دقت ۹۴.۶۴ [illegible]صد [illegible]سایت‌های [illegible]ل را تشخیص دهد[۱۹]. Anggraini و [illegible]ش در سال [illegible]ریتم NaïveBayes و تکنیک گسسته‌س[illegible] فواصل [illegible]erval [illegible] و 5-Interval جهت تشخیص حملات نرم‌افزارهای بدخواه استفاده کردند. نتایج بدست آمده با استفاده از مجموعه داده دریافتی از سایت kaggle که شامل ۱۰۰۰۰۰ رکورد و ۳۲ ویژگی می‌باشد نشان داد که دقت روش پیشنهادی با استفاده از الگوریتم Naïve Bayes زمانی که داده‌ها از مرحله تشخیص و گسسته‌سازی عبور نکرده‌اند برابر با ۶۹.۷۲ درصد و زمانی که داده‌ها با فاصله 3-Interval از مرحله تشخیص عبور کنند برابر با ۷۸.۱۶ درصد خواهد بود و در زمانی که میزان فاصله برابر با 5-interval باشد دقت برابر با ۷۹.۹۷ درصد است[۲۰]. Samantray و همکارش در سال ۲۰۲۰ جهت تشخیص نرم‌افزارهای بدخواه از روش‌های انتخاب ویژگی جنگل تصادفی و k-best و برای طبقه‌بندی داده‌ها از طبقه‌بندهای Naive Bayes، Logistic Regression و SVM استفاده کردند. نتایج بدست آمده با استفاده از مجموعه داده دریافتی از سایت kaggle که شامل ۱۰۰۰۰۰ رکورد و ۳۲ ویژگی می‌باشد نشان داد که دقت بدست آمده برای طبقه‌بندهای Naive Bayes، Logistic Regression و SVM پس از انتخاب ویژگی بوسیله روش جنگل تصادفی به ترتیب برابر با ۷۳.۰۱ درصد، ۹۵.۵۳ درصد و ۷۵.۹۷ درصد می‌باشد اما با استفاده از روش انتخاب ویژگی k-best میزان دقت به ترتیب برابر ۶۹.۷ درصد، ۹۳.۳۳ درصد و ۸۷.۳ درصد می‌باشد[۲۱].

۹- پیاده سازی روش پیشنهادی

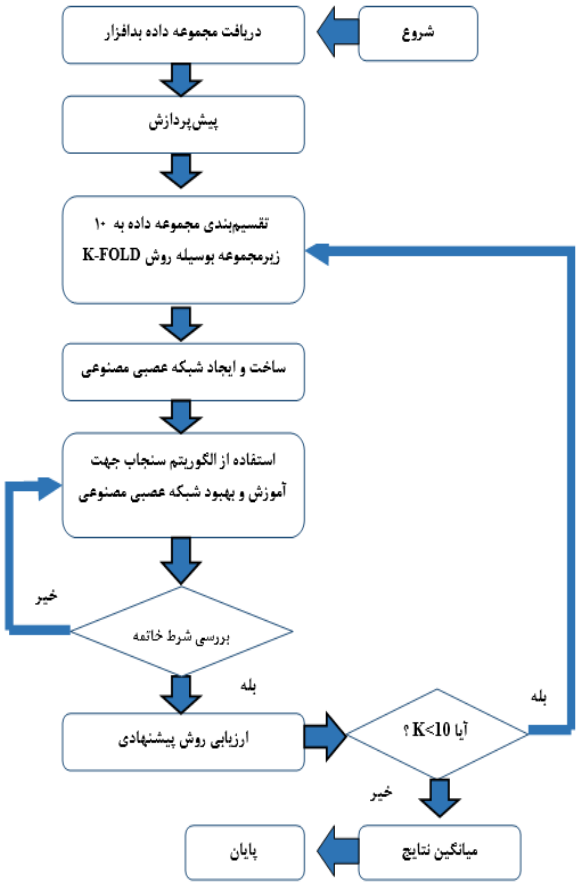
در این تحقیق قصد داریم روشی جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری با استفاده از شبکه عصبی مصنوعی بهبود یافته بوسیله الگوریتم بهینه‌سازی سنجاب ارائه دهیم. در جدول ۴-۱ مقادیر پارامترهای بکار رفته در این تحقیق بیان شده است.

جدول ۴-۱: مقایر پارامترهای مورد استفاده در روش پیشنهادی

K-Fold (تعداد حلقه تکرار برنامه)	10 = K
تعداد نرون لایه ورودی	**۳۲ نورون**
تعداد نرون لایه خروجی	**یک نورون**
تعداد لایه‌هاو نرون مخفی	**یک لایه با ۲۰ نورون**
الگوریتم یادگیری	**پس انتشار خطا**
جمعیت اولیه	**۵۰ سنجاب**
تعداد تکرار (شرط خاتمه)	**۱۰۰۰ بار آموزش شبکه عصبی**

[illegible]Hickory

فلوچارت روش پیشنهادی در شکل ۳-۱ نمایش داده شده است.

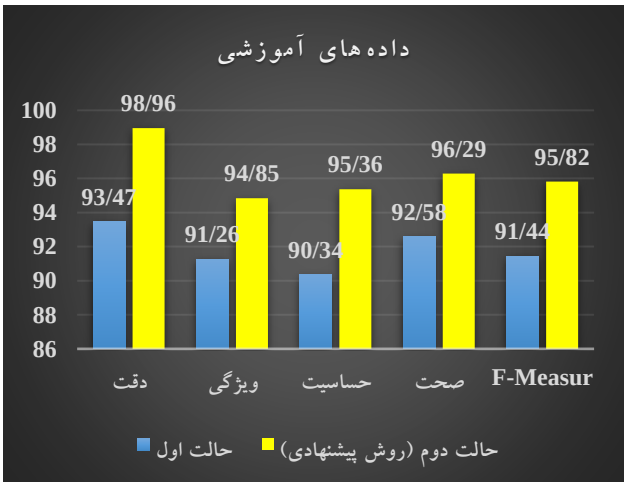


شکل ۳-۱: فلوچارت روش پیشنهادی

برای انجام این منظور مجموعه داده مورد نظر پس از فراخوانی در بازه بین صفر و یک مورد نرمال‌سازی قرار گرفت و در ادامه داده‌های موجود با استفاده از روش اعتبارسنجی Kfold به دو دسته داده‌های آموزشی و آزمایشی تقسیم شدند. سپس شبکه عصبی پیشنهادی ایجاد و با داده‌های آموزشی به وسیله الگوریتم بهینه‌سازی سنجاب مورد آموزش قرار گرفت و در ادامه شبکه عصبی بهبودیافته با استفاده از داده‌های آزمایشی ارزیابی شد و عملکرد آن با استفاده از از معیارهای دقت[۲۷]، حساسیت[۲۸]، ویژگی[۲۹]، صحت[۳۰]، F-Measure مورد بررسی قرار گرفت در پایان با میانگین از نتایج بدست آمده برای داده‌های آموزشی و آزمایشی در هر فولد پس از ۱۰۰۰ بار تکرار در هر بار اجرای برنامه نتایج بدست آمده بعنوان پاسخ کلی بیان گردید.

۱۰- ارزیابی نتایج

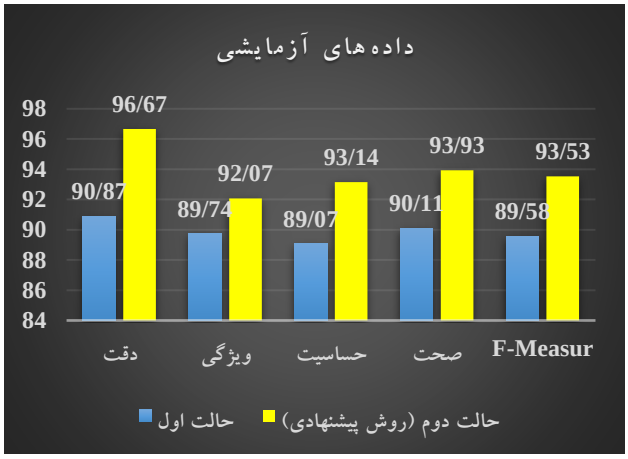
روش پیشنهادی در این تحقیق جهت تشخیص انواع نرم‌افزارهای بدخواه در سیستم‌های کامپیوتری در دو حالت بررسی و نتایج آن مورد تحلیل قرار گرفت. در حالت اول شبکه عصبی مصنوعی طراحی شده را بدون استفاده از الگوریتم بهینه‌سازی سنجاب به تعداد ۱۰۰۰ مرتبه مورد آموزش و ارزیابی قرار داده و در ادامه در حالت دوم روش پیشنهادی را توسط الگوریتم بهینه‌سازی سنجاب بهبود داده و به تعداد ۱۰۰۰ مرتبه مورد آموزش و ارزیابی قرار داده و اجرا کردیم.



شکل ۵: مقایسه نتایج بدست آمده برای داده‌های آموزشی

باتوجه به نتایج بدست آمده از پیاده‌سازی روش پیشنهادی که در شکل ۴-۱ نشان داده شده است دقت بدست آمده جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری در حالت دوم که از الگوریتم سنجاب پرنده استفاده نمودیم نسبت به حالت اول ۵.۴۹ درصد افزایش داشته است همچنین مقادیر معیارهای ویژگی، حساسیت، صحت و F-Measur نیز در حالت دوم به ترتیب ۳.۵۹ درصد، ۵.۰۲ درصد، ۳.۷۱ درصد و ۴.۳۸ درصد نسبت به حالت اول افزایش داشته است.

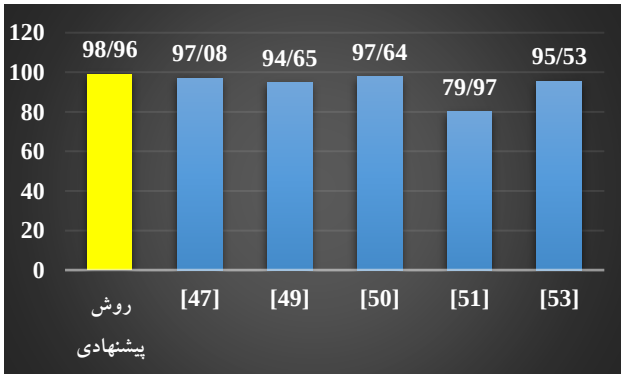
[27] Accuracy
[28] Sensitivity
[29] Specificity
[30] Precision



شکل ۶: مقایسه نتایج بدست آمده برای داده‌های آزمایشی

باتوجه به نتایج بدست آمده از پیاده‌سازی روش پیشنهادی که در شکل ۴-۲ نشان داده شده است دقت بدست آمده جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری در حالت دوم که از الگوریتم سنجاب پرنده جهت بهبود شبکه عصبی پیشنهادی استفاده نمودیم نسبت به حالت اول ۵.۸ درصد افزایش داشته است همچنین مقادیر معیارهای ویژگی،حساسیت، صحت و F-Measur نیز در حالت دوم به ترتیب ۲.۳۳ درصد، ۴.۰۷ درصد، ۳.۸۲ درصد و ۳.۹۵ درصد نسبت به حالت اول افزایش داشته است.

در نتیجه براساس نتایج بدست آمده می‌توان گفت حالت دوم یا روش پیشنهادی عملکرد بهتری نسبت به حالت اول داشته است. در ادامه دقت روش پیشنهادی در این تحقیق را با مقالات بخش کارهای پیشین مورد ارزیابی و مقایسه قرار دهیم.



شکل ۴-۳: مقایسه دقت روش پیشنهادی و مقالات پیشین

باتوجه به مقایسات صورت گرفته در شکل ۴-۳ می‌توان گفت روش پیشنهادی در این تحقیق نسبت به مقالات مورد مقایسه از دقت بالاتر و در نتیجه عملکرد بهتری جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری برخوردار است.

۱۱- نتیجه گیری

در سال‌های اخیر یکی از مهمترین چالش‌های امنیت اطلاعات و شبکه‌های ارتباطی افزایش روزافزون انواع نرم‌افزارهای بدخواه و به دنبال آن یافتن راه‌های مناسب جهت حفاظت سیستم‌ها در مقابل آنهاست که از مهمترین دغدغه‌های برنامه‌نویسان و متخصصین امنیت اطلاعات شناخت به موقع و یافتن راه‌های مقابله با اثرات مخرب این‌گونه نرم‌افزارهای بدخواه می‌باشد که در مقابله با این روند شرکت‌های ضدنرم‌افزارهای بدخواهی به طور جدی به دنبال راه‌های بهتری برای شناسایی نرم‌افزارهای بدخواه هستند لذا ایجاد روشی که بتواند به‌صورت کارا به شناسایی و جلوگیری از نفوذ نرم‌افزارهای بدخواه بپردازد همواره مورد نیاز خواهد بود. لذا در این تحقیق سعی کردیم تا با استفاده از شبکه عصبی بهبود یافته بوسیله الگوریتم سنجاب پرنده به شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری با دقت بالایی بپردازیم. نتایج بدست آمده از پیاده‌سازی روش پیشنهادی برای داده‌های آموزشی حاکی از آن است که دقت بدست آمده جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری در حالت دوم که از الگوریتم سنجاب پرنده جهت بهبود شبکه عصبی پیشنهادی استفاده نمودیم نسبت به حالت اول ۵.۴۹ درصد افزایش داشته است همچنین مقادیر معیارهای ویژگی، حساسیت، صحت و F-Measur نیز در حالت دوم به ترتیب ۳.۵۹ درصد، ۵.۰۲ درصد، ۳.۷۱ درصد و ۴.۳۸ درصد نسبت به حالت اول افزایش داشته است. همچنین مقادیر معیارهای ویژگی، حساسیت، صحت و F-Measur نیز در حالت دوم برای داده‌های آزمایشی به ترتیب ۵.۸ درصد، ۲.۳۳ درصد، ۴.۰۷ درصد، ۳.۸۲ درصد و ۳.۹۵ درصد نسبت به حالت اول افزایش داشته است. همچنین در ادامه جهت بررسی روش پیشنهادی به مقایسه دقت روش پیشنهادی در این تحقیق با دقت مقالات کارهای پیشین پرداخته شد که باتوجه به مقایسه صورت گرفته و نتایج حاصل از پیاده‌سازی روش پیشنهادی و بهبود شبکه عصبی مصنوعی بوسیله الگوریتم سنجاب پرنده می‌توان گفت روش پیشنهادی در این تحقیق از دقت و در نتیجه عملکرد بهتری جهت شناسایی آلودگی نرم‌افزارهای بدخواه در زیرساخت‌های شبکه‌ای و سیستم‌های نرم‌افزاری برخوردار است.

مراجع

[1] Faruk, M. J. H., Shahriar, H., Valero, M., Barsha, F. L., Sobhan, S., Khan, M. A., ... & Wu, F. (2021, December). Malware Detection and Prevention using Artificial Intelligence Techniques. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 5369-5377). IEEE.

[2] Singh, J., & Singh, J. (2021). A survey on machine learning-based malware detection in executable files. *Journal of Systems Architecture*, *112*, 101861.

[3] Manavi, F., & Hamzeh, A. (2019, July). A new approach for malware detection based on evolutionary algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 1619-1624).

[4] AV-TEST, "Internet security threat report (ISTR)," 2019.

[5] Sharma, N., & Arora, B. (2021). Data mining and machine learning techniques for malware detection. In *Rising Threats in Expert Applications and Solutions* (pp. 557-567). Springer, Singapore.

[6] Yadav, P., Menon, N., Ravi, V., Vishvanathan, S., & Pham, T. D. (2022). EfficientNet Convolutional Neural Networks-based Android Malware Detection. *Computers & Security*, 102622.

[7] Sandeep, H. R. (2019, May). Static analysis of android malware detection using deep learning. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)* (pp. 841-845). IEEE.

[8] Hadiprakoso, R. B., Kabetta, H., & Buana, I. K. S. (2020, November). Hybrid-based malware analysis for effective and efficiency android malware detection. In *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 8-12). IEEE.

[9] Saad, S., Briguglio, W., & Elmiligi, H. (2019). The curious case of machine learning in malware detection. *arXiv preprint arXiv:1905.07573*.

[10] Shatnawi, A. S., Yassen, Q., & Yateem, A. (2022). An android malware detection approach based on static feature analysis using machine learning algorithms. *Procedia Computer Science*, *201*, 653-658.

[11] Liu, Y., & Wang, Y. (2019, March). A robust malware detection system using deep learning on API calls. In *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)* (pp. 1456-1460). IEEE.

[12] Urooj, U., Al-rimy, B. A. S., Zainal, A., Ghaleb, F. A., & Rassam, M. A. (2021). Ransomware detection using the dynamic analysis and machine learning: A survey and research directions. *Applied Sciences*, *12*(1), 172.

[13] Sabhadiya, S., Barad, J., & Gheewala, J. (2019, April). Android malware detection using deep learning. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 1254-1260). IEEE.

[14] Nakhodchi, S., Upadhyay, A., & Dehghantanha, A. (2020). A comparison between different machine learning models for iot malware detection. *Security of Cyber-Physical Systems: Vulnerability and Impact*, 195-202.

[15] Ma, Z., Ge, H., Liu, Y., Zhao, M., & Ma, J. (2019). A combination method for android malware detection based on control flow graphs and machine learning algorithms. *IEEE access*, *7*, 21235-21245.

[16] Jain, M., V. Singh, and A. Rani, *A novel nature-inspired algorithm for optimization: Squirrel search algorithm.* Swarm and evolutionary computation, 2019. **44**: p. 148-175.

[17] Jamal, A., Hayat, M. F., & Nasir, M. (2022). Malware Detection and Classification in IoT Network using ANN. *Mehran University Research Journal of Engineering and Technology*, *41*(1), 80-91.

[18] Siahmarzkooh. A, Intrusion Detection in Computer Networks Using Decision Tree and Feature Reduction, Journal of Electronical & Cyber Defence Vol. 9, No. 3, 2021, Serial No. 35

[19] Safara.M, Sabahnour. F, (2020). Increase the accuracy of detecting fake web pages using hyena optimization algorithm and artificial neural network. Journal of Command and Control, 3(4), 72-91.

[20] Anggraini, I., Kunang, Y. N., & Firdaus, F. (2020). Penerapan Naive Bayes Pada Detection Malware dengan Diskritisasi Variabel. Telematika, 13(1), pp.11-21

[21] Samantray, O. P., & Tripathy, S. N. (2020). A Knowledge-Domain Analyser for Malware Classification. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-7). IEEE.