Journal of Information and Communication Technology

Volume 17, Issue 66-67, Autumn and Winter 2025, pp. 178-188

Presenting a Hybrid Model on Machine Learning and Principal Component Analysis for Action Detection in the Internet of Things

Zahra Shahpar[1*], Mohammadreza Badragheh[2]

[1] Department of Computer Engineering, Zabol Branch, Islamic Azad University, Zabol, Iran
[2] Department of Computer Engineering, Ferdows Branch, Islamic Azad University, Ferdows, Iran

Received: 25 March 2025, Revised: 20 October 2025, Accepted: 10 February 2026
Paper type: Research

Abstract

With the rapid expansion of the Internet of Things and the increase in the number of devices connected to the Internet, the security of Internet of Things systems has become a serious challenge. Due to the easy access to these devices and existing security weaknesses, we are witnessing various attacks and an increasing penetration of these systems. One of the effective tools in dealing with these threats is intrusion detection systems. In this study, a hybrid model for intrusion detection in Internet of Things networks is presented that uses machine learning methods (logistic regression, support vector machine, nearest neighbor, random forest, decision tree, and multilayer neural network) along with principal component analysis (PCA) to reduce data dimensions. The proposed method was implemented and investigated on the UNSW-NB15 dataset. Based on the results obtained; The logistic regression model with a single-class accuracy of 97.84% and a multi-class accuracy of 89.81%, the support vector machine model with a single-class accuracy of 97.85% and a multi-class accuracy of 89.89%, the nearest neighbor model with a single-class accuracy of 98.31% and a multi-class accuracy of 88.55%, the decision tree model with a single-class accuracy of 98.11% and a multi-class accuracy of 85.45%, and the multilayer neural network model with a single-class accuracy of 98.39% and a multi-class accuracy of 89.94% have been able to identify different types of attacks. In particular, the results obtained indicate that the random forest model with a single-class accuracy of 98.63% and a multi-class accuracy of 89.06% has the best performance among the models. Also, the false positive rate was reduced to about 4% and the processing time was less than 1 millisecond. Comparison of the proposed method with other methods showed that the proposed method, with an accuracy of 84% provides significant improvement in accuracy, efficiency, and speed. Overall, the proposed model can be used as an effective and reliable method for detecting attacks in IoT networks, especially in resource-constrained environments.

Keywords: IoT, Intrusion Detection System, Principal Component Analysis (PCA), Machine Learning

[*] Corresponding Author's email: zahrashahpar@iau.ac.ir

دوفصلنامه
فناوری اطلاعات و ارتباطات ایران

سال هفدهم، شماره‌های ۶۶ و ۶۷، پاییز و زمستان ۱۴۰۴، صفحه ۱۷۸ الی ۱۸۸

ارائه مدلی ترکیبی مبتنی بر یادگیری ماشین و تحلیل مولفه‌های اصلی برای تشخیص حملات در اینترنت اشیا

زهرا شهپر[1*]، محمدرضا بدرقه[2]

[1] گروه مهندسی کامپیوتر، واحد زابل، دانشگاه آزاد اسلامی، زابل، ایران

[2] گروه مهندسی کامپیوتر، واحد فردوس، دانشگاه آزاد اسلامی، فردوس، ایران

تاریخ دریافت: ۱۴۰۴/۰۵/۰۱ تاریخ بازبینی: ۱۴۰۴/۰۷/۲۸ تاریخ پذیرش: ۱۴۰۴/۱۱/۲۱

نوع مقاله: پژوهشی

چکیده

با گسترش سریع اینترنت اشیا و افزایش تعداد دستگاه‌های متصل به اینترنت، امنیت سیستم‌های اینترنت اشیا به چالشی جدی تبدیل شده است. به دلیل دسترسی ساده به این دستگاه‌ها و ضعف‌های امنیتی موجود، شاهد حملات متنوع و افزایش نفوذ به این سیستم‌ها هستیم. یکی از ابزارهای مؤثر در مقابله با این تهدیدات، سیستم‌های تشخیص نفوذ است. در این پژوهش مدلی ترکیبی برای تشخیص نفوذ در شبکه‌های اینترنت اشیا ارائه می‌شود که از روش‌های یادگیری ماشین (رگرسیون لجستیک، ماشین بردار پشتیبان، نزدیک‌ترین همسایه، جنگل تصادفی، درخت تصمیم و شبکه عصبی چندلایه) در کنار تحلیل مؤلفه‌های اصلی (PCA) برای کاهش ابعاد داده‌ها بهره می‌برد. روش پیشنهادی بر روی مجموعه داده UNSW-NB15 پیاده‌سازی و بررسی شد. بر اساس نتایج به دست آمده؛ مدل رگرسیون لجستیک با دقت تک کلاسی ۹۷٫۸۴٪ و دقت چند کلاسی ۸۹٫۸۱٪، مدل ماشین بردار پشتیبان با دقت تک کلاسی ۹۷٫۸۵٪ و دقت چند کلاسی ۸۹٫۸۹٪، مدل نزدیک‌ترین همسایه با دقت تک کلاسی ۹۸٫۳۱٪ و دقت چند کلاسی ۸۸٫۵۵٪، مدل درخت تصمیم با دقت تک کلاسی ۹۸٫۱۱٪ و دقت چند کلاسی ۸۵٫۴۵٪ و مدل شبکه عصبی چند لایه با دقت تک کلاسی ۹۸٫۳۹٪ و دقت چند کلاسی ۸۹٫۹۴٪ توانسته اند انواع حملات مختلف را شناسایی کنند. به طور خاص نتایج به دست آمده حاکی از آن است که ، مدل جنگل تصادفی با دقت تک کلاسی ۹۸٫۶۳٪ و دقت چند کلاسی ۸۹٫۰۶٪. بهترین عملکرد را در میان مدل‌ها دارد. همچنین نرخ مثبت کاذب به حدود ۴٪ کاهش یافت و زمان پردازش به کمتر از ۱ میلی ثانیه رسید. مقایسه روش پیشنهادی با سایر روش‌ها از نشان داد که روش پیشنهادی با دقت ٪۸۴، بهبود قابل توجهی در دقت، کارایی و سرعت ارائه می‌دهد. به طور کلی، مدل پیشنهادی می‌تواند به عنوان روشی مؤثر و قابل اعتماد برای تشخیص حملات در شبکه‌های اینترنت اشیا به ویژه در محیط‌های با منابع محدود مورد استفاده قرار گیرد.

کلیدواژگان: اینترنت اشیا، سیستم تشخیص نفوذ، تحلیل مؤلفه‌های اصلی(PCA) ، یادگیری ماشین

* رایانامهٔ نویسنده مسؤول: zahrashahpar@iau.ac.ir

۱- مقدمه

اینترنت اشیا[1] شبکه‌ای از دستگاه‌های متصل است که امکان تبادل اطلاعات و تصمیم‌گیری بدون دخالت انسان را فراهم می‌کند. این دستگاه‌ها در حوزه‌هایی مانند خانه‌های هوشمند برای مدیریت خودکار سیستم‌های روشنایی و گرمایشی، مراقبت‌های بهداشتی برای نظارت از راه دور بر سلامت بیماران با استفاده از دستگاه‌های پوشیدنی، و صنعت برای بهبود بهره‌وری کاربرد دارند [۱]. با این حال، محدودیت‌های منابع مانند قدرت پردازش پایین، حافظه محدود، ظرفیت باتری کم و ابعاد فیزیکی کوچک، این دستگاه‌ها را در برابر حملات سایبری آسیب‌پذیر می‌سازد. برای مثال، حملات می‌توانند دسترس‌پذیری سیستم‌ها را مختل کنند، منابع انرژی را مصرف کنند و منجر به خاموشی‌های گسترده شوند. علاوه بر این، مسائل حریم خصوصی و محرمانگی داده‌ها چالش‌های دیگری ایجاد می‌کند، زیرا روش‌های رمزنگاری سنتی برای دستگاه‌هایی با منابع محدود مناسب نیستند و نیاز به رویکردهای متفاوت مانند رمزنگاری سبک وجود دارد [۵-۲]. بات‌نت میرا[2] نمونه‌ای عینی از این حملات است که در سال ۲۰۱۶ هزاران دستگاه متصل به اینترنت اشیاء مانند دوربین‌های نظارتی و روترهای خانگی را با استفاده از رمزهای عبور پیش‌فرض آلوده کرد و از آن‌ها برای اجرای حملات منع سرویس[3] علیه وب‌سایت‌هایی مانند توئیتر[4] استفاده نمود. در نمونه‌ای دیگر، آسیب‌پذیری در پمپ‌های انسولین و ضربان‌سازهای هوشمند امکان کنترل از راه دور و تغییر دوز دارو توسط مهاجم را فراهم کرد، که تهدیدی مستقیم برای جان بیماران محسوب می‌شد [۹-۶]. این گونه از حملات نشان‌دهنده اهمیت و عمق این آسیب‌پذیری‌ها است. سیستم‌های تشخیص نفوذ[5] ابزار مؤثری برای مقابله با این تهدیدات هستند، زیرا با تحلیل مداوم ترافیک شبکه، فعالیت‌های غیرمجاز را شناسایی و از نفوذ جلوگیری می‌کنند. در شبکه‌های اینترنت اشیاء با محدودیت منابع، نیاز به سیستم‌های تشخیص نفوذ بهینه که بدون کاهش سرعت یا مصرف بالای انرژی عمل کنند؛ احساس می‌شود [۱۰-۱۲]. از این رو در این پژوهش مدلی ترکیبی مبتنی بر یادگیری ماشین[6] و تحلیل مؤلفه‌های اصلی[7] برای کاهش ابعاد داده‌ها ارائه می‌شود. هدف پژوهش افزایش دقت تشخیص حملات، کاهش نرخ مثبت کاذب برای کاهش هشدارهای نادرست، و کاهش زمان پردازش برای سازگاری با محیط‌های واقعی است. مراحل انجام پژوهش عبارت‌اند از جمع‌آوری داده‌ها، پیش‌پردازش، اعمال PCA برای استخراج ویژگی‌های کلیدی، آموزش و ارزیابی مدل‌ها. برای این منظور از مجموعه داده UNSW-NB15 استفاده شده است.

در ادامه، بخش دوم به پیشینه پژوهش می‌پردازد، بخش سوم روش پیشنهادی را توصیف می‌کند، بخش چهارم نتایج تجربی را ارائه می‌دهد، و بخش پنجم به بحث و نتیجه‌گیری اختصاص دارد.

۲- پیشینه پژوهش

پژوهش‌های متعددی در زمینه تشخیص نفوذ در شبکه‌های اینترنت اشیا انجام شده است. احمد[8] و همکاران [۱۳] با بررسی روش‌های تشخیص ناهنجاری نشان دادند که مدل‌های یادگیری عمیق می‌توانند در شناسایی حملات شبکه عملکرد مناسبی داشته باشند. دیرو[9] و همکاران [۱۴] با به‌کارگیری رویکردهای مبتنی بر یادگیری عمیق توزیع‌شده برای شناسایی حملات در محیط‌های اینترنت اشیا بر روی مجموعه داده‌هایی مانند KDDCUP '99، NSL-KDD و ISCX، توانستند نرخ تشخیص بالاتری نسبت به روش‌های کلاسیک به دست آورند. آن‌ها نشان دادند که یادگیری عمیق می‌تواند به طور مؤثر در تشخیص فعالیت‌های ناهنجار عمل کند، هرچند چالش‌هایی مانند توزیع نامتوازن داده‌ها همچنان وجود دارد.

بابار[10] و همکاران [۱۵] در زمینه امنیت شبکه‌های حسگر بی‌سیم مکانیسم مدیریت کلید و مبتنی بر هش برای خوشه‌بندی پیشنهاد دادند که از پیش‌توزیع کلید تصادفی استفاده می‌کند. این رویکرد با ارزیابی شاخص‌هایی مانند نرخ از دست دادن بسته، مصرف انرژی و تأخیر، امنیت ارتباطات تک‌هاپ و چندهاپ را افزایش می‌دهد، اما محدودیت‌هایی مانند عدم تحرک سرخوشه‌ها دارد.

در حوزه انتخاب و استخراج ویژگی، فتانی[11] و همکاران [۱۶] با استفاده از الگوریتم‌های هوش جمعی و روش‌های بهینه‌سازی رویکردی ترکیبی برای بهبود کارایی سیستم‌های تشخیص نفوذ پیشنهاد کردند.

دیشا[12] و همکاران [۱۷] با توجه به اثربخشی روش‌های یادگیری ماشین، چندین مدل یادگیری ماشین را برای سیستم تشخیص نفوذ اعمال کردند. آن‌ها برای تحلیل عملکرد مدل‌های یادگیری ماشین

[1] Internet of things (IoT)
[2] Mirai
[3] Distributed Denial of Service (DDoS)
[4] Twitter
[5] Intrusion Detection System (IDS)
[6] Machine learning (ML)
[7] Principal Component Analysis (PCA)
[8] Ahmed
[9] Diro
[10] Babar
[11] Fatani
[12] Disha

مانند درخت تصمیم[1] و شبکه‌های عصبی بازگشتی از مجموعه داده‌های UNSW-NB 15 و Network TON_IoT استفاده کردند. نتایج این پژوهش نشان داد، استفاده از تکنیک انتخاب ویژگی مبتنی بر ناخالصی جینی در جنگل تصادفی وزنی، عملکرد تشخیص نفوذ در داده‌های نامتعادل را بهبود می‌بخشد.

محی‌الدین و همکاران [۱۸] از ضریب همبستگی پیرسون[2] و جنگل جداسازی[3] برای کاهش ابعاد و حذف موارد پرت در مجموعه داده Bot-IoT استفاده کردند. این روش نرخ تشخیص و دقت را ارتقا داده و عملکرد بهتری نسبت به مدل‌های دیگر نشان می‌دهد.

پژوهش‌های جدید نیز تمرکز ویژه‌ای بر بهینه‌سازی مدل‌ها برای محیط‌های کم‌منبع داشته‌اند. به عنوان نمونه، تیواری[4] و همکاران [۱۹] تکنیک تشخیص نفوذ مبتنی بر یادگیری ماشین برای IoT ارائه کردند که توانست با استفاده از تکنیک‌های انتخاب ویژگی و کاهش ابعاد، دقت خوبی را بر روی مجموعه داده WUSTL-IIOT-2021 به دست آورد. آن‌ها نشان دادند ترکیب بهینه‌سازی ازدحام ذرات با PCA و مدل‌های رگرسیون، دقت ٪۱۰۰ را در برخی ترکیب‌ها به دست می‌آورد و دسترسی به منابع را تسریع می‌کند.

پژوهش‌های صورت گرفته در زمینه تشخیص نفوذ به ترتیب زمانی و با ذکر روش مورد استفاده، مزیت کلیدی و با توجه به مجموعه داده‌های مورد استفاده در، جدول ۱ خلاصه شده‌اند. با وجود دستاوردهای حاصل شده در پژوهش‌های پیشین همچنان چالش‌هایی باقی است:

- بسیاری از مدل‌ها نیازمند منابع پردازشی و حافظه بالا هستند و اجرای آن‌ها در دستگاه‌های IoT دشوار است.
- در برخی پژوهش‌ها، نرخ مثبت کاذب همچنان بالاست و دقت مدل‌ها در شناسایی حملات ناشناخته کافی نیست.
- مقایسه مستقیم با مجموعه‌داده‌های جدیدتر مانند UNSW-NB15 کمتر مورد توجه قرار گرفته است

۳- روش پیشنهادی

۳-۱- معماری سیستم تشخیص نفوذ

سیستم‌های تشخیص نفوذ مبتنی بر یادگیری ماشین ابزارهایی مؤثر برای حفاظت از شبکه‌های اینترنت اشیا در برابر حملات سایبری به ویژه در محیط‌هایی با منابع محدود هستند. در معماری پیشنهادی، ابتدا لاگ‌های ترافیکی شامل شامل اطلاعاتی مانند سرآیند بسته‌ها، درخواست‌های سرویس و فعالیت‌های دستگاه از دستگاه‌های IoT جمع‌آوری شده و به سرورهای ذخیره‌سازی توزیع‌شده ارسال می‌شوند تا دو ویژگی مقیاس‌پذیری و دسترسی سریع به داده‌ها تضمین شود. سپس، داده‌ها به موتور تحلیل مبتنی بر یادگیری ماشین که از روش‌های ترکیبی (تحلیل امضا و ناهنجاری) برای شناسایی فعالیت‌های مشکوک استفاده می‌کنند؛ منتقل می‌شوند. در صورت تشخیص نفوذ، سیستم گزارش‌هایی شامل اطلاعات نفوذکننده، زمان و نوع حمله تولید شده را به مدیر شبکه هشدار می‌دهد. دیکشنری نفوذ به‌صورت خودکار با داده‌های جدید به‌روزرسانی می‌شود تا الگوهای نوظهور را شناسایی کند. این معماری با استفاده از تحلیل مولفه‌های اصلی ابعاد داده‌ها را کاهش می‌دهد تا فشار بر منابع سخت‌افزاری کم شود [۲۰]. شکل ۱ معماری سیستم تشخیص نفوذ پیشنهادی شامل بخش‌های جمع‌آوری لاگ‌های ترافیکی، موتور تحلیل و دیکشنری نفوذ را نشان می‌دهد.

جدول ۱. مرور پژوهش‌های پیشین

شماره مرجع	سال	روش	مجموعه داده	مزیت کلید	محدودیت
۱۳	۲۰۱۶	یادگیری عمیق	DARPA KDD	مرور جامع ناهنجاری‌ها	پچیدگی- عدم ارزیابی کمی
۱۵	۲۰۱۶	مدیریت کلید مبتنی بر هش	KDD	کاهش مصرف انرژی و تأخیر	عدم تحرک سرخوشه‌ها
۱۴	۲۰۱۸	یادگیری عمیق توزیع‌شده	KDD'99, NSL-KDD ، ISCX	بهبود تشخیص حملات	توزیع نامتوازن داده‌ها
۱۶	۲۰۲۱	هوش جمعی و روش‌های بهینه‌سازی	CIC2017 , NSL-KDD ، BoT-IoT, KDD99	دقت بالا در F1-score	وابستگی به داده‌ها
۱۷	۲۰۲۲	انتخاب ویژگی با جنگل تصادفی وزنی	UNSW-NB15 Network TON_IoT	عملکرد برتر در داده‌های نامتعادل	تمرکز بر طبقه‌بندی باینری
۱۸	۲۰۲۳	PCC و جنگل جداسازی	Bot-IoT	کاهش هزینه محاسبات	داده‌های نامتعادل
۱۹	۲۰۲۴	PSO با PCA و MARS	WUSTL-IIOT-2021	دقت ٪۱۰۰ در ترکیب‌ها	بار محاسباتی در

[1] Decision tree (DT)

[2] Pearson Correlation Coefficients (PCC)

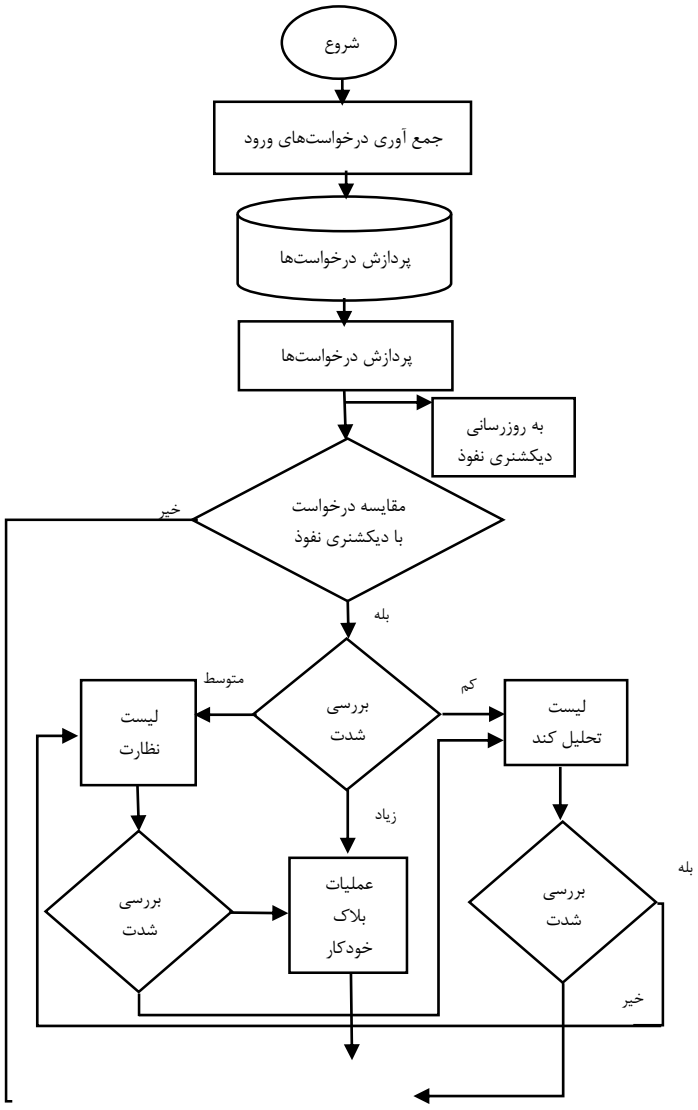
[3] Isolation Forest (IF)

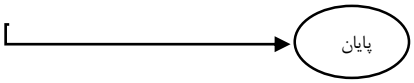
[4] Tiwari

محیط‌های بزرگ					

۳-۲- چارچوب PCA و کاربرد آن در تشخیص نفوذ

تحلیل مولفه‌های اصلی تکنیکی برای کاهش ابعاد داده‌ها و استخراج ویژگی‌های کلیدی است که زمان پردازش را کاهش داده و دقت مدل‌ها را بهبود می‌بخشد. در این روش، داده‌های جمع‌آوری‌شده از شبکه IoT ابتدا استانداردسازی می‌شوند (کسر میانگین و تقسیم بر انحراف معیار). سپس، ماتریس همبستگی برای تحلیل روابط متغیرها محاسبه می‌شود. در ادامه، مقادیر ویژه و بردارهای ویژه استخراج شده و مؤلفه‌های اصلی با بیشترین واریانس انتخاب می‌شوند. این فرآیند داده‌ها را به فضای جدید با ابعاد کمتر تبدیل می‌کند، در حالی که اطلاعات کلیدی حفظ می‌شود. این روش به‌ویژه برای دستگاه‌های IoT با منابع محدود مناسب است، زیرا مصرف انرژی و زمان پردازش را کاهش می‌دهد [۲۱]. شکل ۲ مراحل PCA شامل استانداردسازی، محاسبه مقادیر ویژه و انتخاب مؤلفه‌ها در سیستم تشخیص نفوذ را نشان می‌دهد.

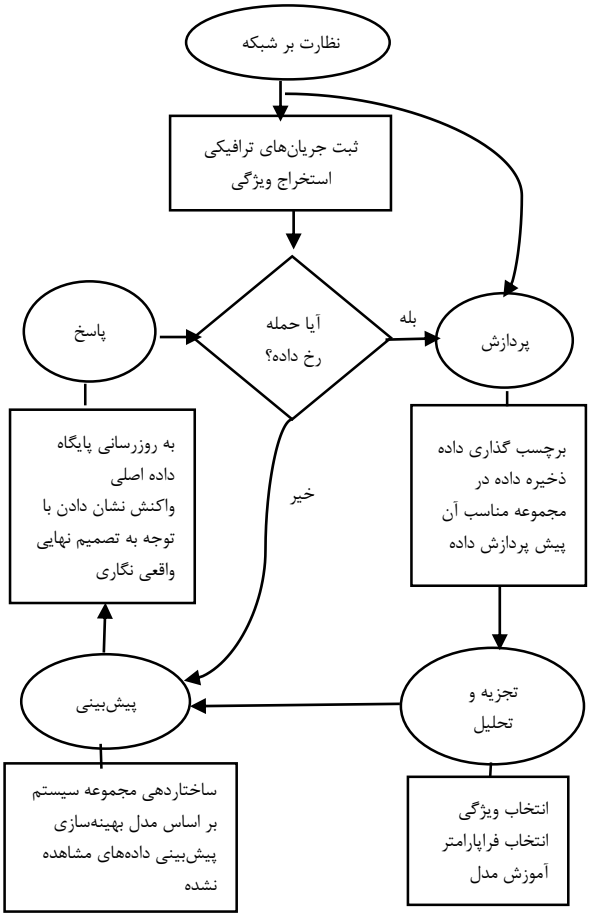




شکل ۱. معماری سیستم تشخیص نفوذ [۲۰]

استفاده از PCA در طراحی سیستم‌های تشخیص نفوذ به مجموعه‌ای از مزایا منجر می‌شود [۲۱]:

- **کاهش زمان پردازش:** با کاهش ابعاد داده و تمرکز بر ویژگی‌های مهم، زمان پردازش به طور قابل توجهی کاهش می‌یابد.
- **بهبود دقت مدل:** با استفاده از داده‌های کمتر و مهم‌تر، مدل‌های یادگیری ماشین می‌توانند دقت بالاتری در شناسایی نفوذها داشته باشند.
- **کاهش مصرف منابع:** سیستم‌های سبک‌تر با نیاز به منابع کمتر می‌توانند به راحتی بر روی دستگاه‌های IoT پیاده‌سازی شوند



شکل ۲. مراحل روش پیشنهادی با استفاده از PCA [۲۱]

۳-۳- مراحل روش پیشنهادی

روش پیشنهادی شامل پنج مرحله زیر است که به طور دقیق و گام

به گام به منظور دستیابی به اهداف پژوهش طراحی شده‌اند.

۳-۳-۱- جمع‌آوری داده‌ها

داده‌های ترافیکی از شبکه IoT، شامل آدرس‌های IP، پورت‌ها، پروتکل‌ها، تعداد بایت‌های منتقل‌شده و ویژگی‌های جیتر، از مجموعه داده UNSW-NB15 جمع‌آوری می‌شوند. این مجموعه شامل ۲،۵۴۰،۰۴۴ جریان (٪۸۷,۳۵ خوش‌خیم و ٪۱۲,۶۵ حمله) با ۴۹ ویژگی است.

۳-۳-۲- پیش‌پردازش داده‌ها

داده‌ها نرمال‌سازی شده و نویز حذف می‌شود تا برای تحلیل آماده شوند.

۳-۳-۳- تحلیل مولفه‌های اصلی

داده‌ها و ویژگی‌های کلیدی با حفظ حداکثر واریانس انتخاب می‌شوند تا ابعاد داده‌ها کاهش یابد. در فرآیند آموزش، داده‌های کاهش‌یافته با PCA به نسبت ۷۰:۳۰ به مجموعه‌های آموزش و آزمایش تقسیم شدند. اعتبارسنجی متقاطع ۵-تایی[1] برای جلوگیری از بیش‌برازش و بهینه‌سازی مدل‌ها استفاده می‌شود. تنظیم‌های بهینه برای هر مدل با آزمایش‌های اولیه بر روی مجموعه داده UNSW-NB15 صورت گرفته تا دقت و سرعت پردازش بهینه شوند. این فرآیند امکان شناسایی الگوهای حملات پیچیده مانند DDoS را با حداقل منابع محاسباتی فراهم می‌کند.

۳-۳-۴- آموزش مدل‌های یادگیری ماشین

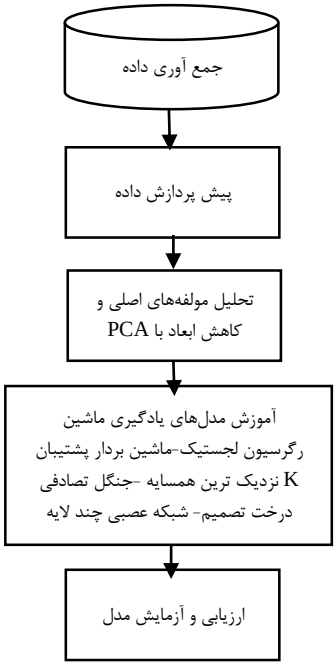
مدل‌های یادگیری ماشین شامل رگرسیون لجستیک[2]، ماشین بردار پشتیبان خطی[3]، K نزدیک‌ترین همسایه[4]، جنگل تصادفی[5]، درخت تصمیم و شبکه عصبی چندلایه[6] بر روی داده‌های کاهش‌یافته آموزش می‌بینند.

- رگرسیون لجستیک: روشی آماری است که برای پیش‌بینی احتمال وقوع یک رویداد دوتایی (مثل بله یا خیر) استفاده می‌شود. این مدل از تابع لجستیک برای تبدیل ورودی‌های خطی به خروجی‌های میان ۰ و ۱ استفاده می‌کند؛ و برای طبقه‌بندی باینری مناسب است [۱۷].
- ماشین بردار پشتیبان خطی: با ایجاد مرز تصمیم، داده‌ها را با حداکثر حاشیه تفکیک می‌کند[۱۷].
- K نزدیک‌ترین همسایه: بر اساس فاصله اقلیدسی، نزدیک‌ترین همسایگان را برای طبقه‌بندی انتخاب می‌کند[۱۷].
- جنگل تصادفی: با ترکیب چندین درخت تصمیم، بیش‌برازش را کاهش داده و دقت بالایی دارد[۱۷].
- درخت تصمیم: داده‌ها را به صورت درختی تقسیم کرده و ویژگی‌های کلیدی را شناسایی می‌کند [۱۷].
- شبکه عصبی چندلایه: با لایه‌های نورونی، الگوهای پیچیده را شناسایی می‌کند[۱۷].

۳-۳-۵- ارزیابی و آزمایش

عملکرد مدل‌ها با معیارهای دقت تک‌کلاسی و چندکلاسی، نرخ مثبت کاذب و زمان پردازش ارزیابی می‌شود.

مراحل روش پیشنهادی در شکل ۳ نشان داده شده است.



شکل ۳. مراحل روش پیشنهادی

۴- نتایج تجربی روش پیشنهادی

همان طور که در بخش قبل بیان شده برای شبیه‌سازی روش پیشنهادی پژوهش از مجموعه داده UNSW-NB15 که توسط آزمایشگاه سایبری مرکز امنیت سایبری استرالیا در سال ۲۰۱۵ منتشر شد، استفاده شده است. این مجموعه شامل ۲،۵۴۰،۰۴۴ جریان داده (٪۸۷,۳۵ خوش‌خیم و ٪۱۲,۶۵ حمله) با ۴۹ ویژگی

[1] K(5)-fold cross-validation
[2] Logistic regression
[3] Support vector machines (SVM)
[4] K-Nearest Neighbors (KNN)
[5] Random forest (RF)
[6] Multilayer perceptron (MLP)

مانند آدرس‌های IP، پورت‌ها، پروتکل‌ها، تعداد بایت‌های منتقل شده، Sload، Dload، مقادیر جیتر و اطلاعات حملات است. این ویژگی‌ها امکان تحلیل دقیق ترافیک شبکه را فراهم می‌کنند [۲۳، ۲۲]. هدف این بخش ارزیابی عملکرد مدل‌های یادگیری بیان شده در ترکیب با PCA برای دستیابی به اهداف کمی پژوهش شامل افزایش دقت و کاهش نرخ مثبت کاذب و زمان پردازش است.

۴-۱- فرآیند آزمایش

مطابق مراحل بیان شده در روش پیشنهادی داده‌های نرمال‌سازی شده ابتدا با PCA پردازش می‌شوند تا ابعاد کاهش یابد. بدین ترتیب که استانداردسازی داده‌ها (کسر میانگین و تقسیم بر انحراف معیار) انجام شده، سپس ماتریس همبستگی و مقادیر ویژه محاسبه می‌گردد تا مؤلفه‌های اصلی با حداکثر واریانس انتخاب شوند. داده‌ها به دو حالت تک‌کلاسی (یک ویژگی کلیدی) و چندکلاسی (چند ویژگی) تقسیم می‌شوند. مجموعه داده به نسبت ۷۰:۳۰ (آموزش:آزمایش) تقسیم شده و اعتبارسنجی متقاطع ۵-تایی برای بهینه‌سازی مدل‌ها به کار گرفته شد. مدل‌های یادگیری ماشین بر PCA ترکیب می‌شوند. پارامترهای هر یک از مدل‌های یادگیری ماشین بر اساس مقادیر ارائه شده در جدول ۲ تنظیم شده است؛ تا تکرارپذیری فرآیند تضمین شود. این تنظیمات بر اساس آزمایش‌های اولیه بهینه‌سازی شده‌اند تا دقت و کارایی در محیط‌های IoT بهبود یابد. در پایان مدل‌ها با معیارهای دقت تک‌کلاسی، دقت چندکلاسی، نرخ مثبت کاذب و زمان پردازش ارزیابی شدند.

۴-۲- نتایج مدل‌های یادگیری ماشین

در این بخش نتایج حاصل از پیاده‌سازی هر یک از مدل‌های یادگیری ماشین ارائه می‌شود.

۴-۲-۱- رگرسیون لجستیک

در این پژوهش از مدل رگرسیون لجستیک برای تشخیص خطا استفاده شده است که تأثیر PCA بر عملکرد رگرسیون لجستیک با مقایسه دقت قبل و بعد از کاهش ابعاد در شکل ۴ نشان داده شده است.

جدول ۲. پارامترهای مدل‌های یادگیری به کار گرفته شده در روش پیشنهادی

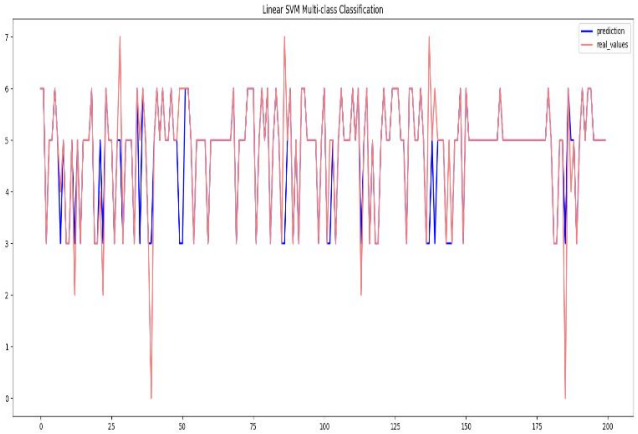
مدل یادگیری ماشین	پارامترها	مقدار
رگرسیون لجستیک	نرخ یادگیری، حداکثر تکرار	۰،۰۱، ۱۰۰۰
SVM خطی	پارامتر منظم‌سازی(C)	۱،۰
KNN	تعداد همسایگان(k)	۵
جنگل تصادفی	تعداد درختان، حداکثر عمق	۱۰۰، ۱۰
درخت تصمیم	حداکثر عمق، معیار تقسیم	۱۰، جینی
شبکه عصبی چندلایه	تعداد لایه‌های مخفی نرخ یادگیری	۲ لایه (۱۰۰ نورون)، ۰،۰۰۱

۴-۲-۲- ماشین بردار پشتیبان خطی

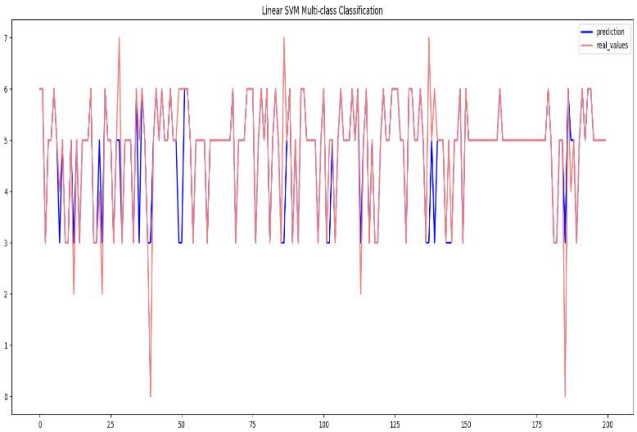
ماشین بردار پشتیبان یکی از روش‌های یادگیری ماشین است که برای تفکیک داده‌ها به دو کلاس مختلف استفاده می‌شود. این مدل با ایجاد یک خط (یا ابرصفحه در ابعاد بالاتر) به نام "مرز تصمیم" کار می‌کند که داده‌ها را به بهترین نحو از یکدیگر جدا می‌کند. هدف SVM حداکثر کردن فاصله بین این مرز و نزدیک‌ترین نقاط داده از هر کلاس است که به آن‌ها "پشتیبان" می‌گویند. شکل ۵ تأثیر PCA بر عملکرد SVM خطی را نشان می‌دهد.

۴-۲-۳- K نزدیک ترین همسایه

این مدل بر اساس نزدیکی داده‌ها به یکدیگر عمل می‌کند و برای طبقه‌بندی یک نمونه جدید، به k نزدیک‌ترین همسایه‌اش نگاه می‌کند. با استفاده از معیارهای فاصله مانند فاصله اقلیدسی، KNN می‌تواند الگوها و انحرافات ناشناخته را شناسایی کند. تأثیر PCA بر KNN در شکل ۶ نشان داده شده است.



شکل ۴. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در مدل رگرسیون لجستیک



شکل ۵. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در ماشین بردار پشتیبان خطی

۴-۲-۴- جنگل تصادفی

جنگل تصادفی شامل ساخت تعدادی درخت تصمیم به‌صورت تصادفی و ترکیب نتایج آن‌ها است که می‌تواند ویژگی‌های پیچیده داده‌ها را شناسایی کند. RF به‌دلیل قابلیت مدیریت داده‌های بزرگ و کاهش احتمال بیش‌برازش به‌ویژه در کاربردهای امنیت سایبری، به محبوبیت رسیده است. نتایج اعمال این مدل در شکل ۷ نشان داده شده است.

۴-۲-۵- درخت تصمیم

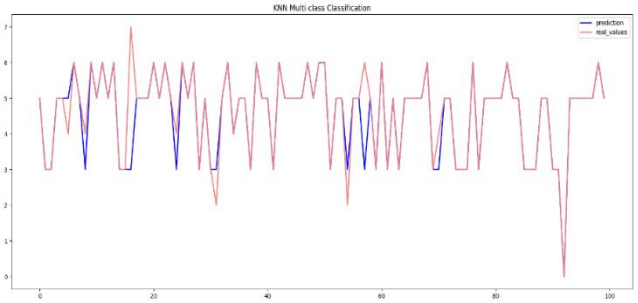
درخت تصمیم با استفاده از تقسیم داده‌ها به صورت درختی، ویژگی‌های مهم را شناسایی کرده و تصمیم‌گیری می‌کند. در فرآیند آموزش، شرایط و ویژگی‌های مختلف به صورت گام‌به‌گام بررسی می‌شوند تا بتوانند انواع مختلف حملات را شناسایی کنند. نتایج حاصل از این مدل در شکل ۸ نشان داده شده است.

۴-۲-۶- شبکه‌ی عصبی چند لایه

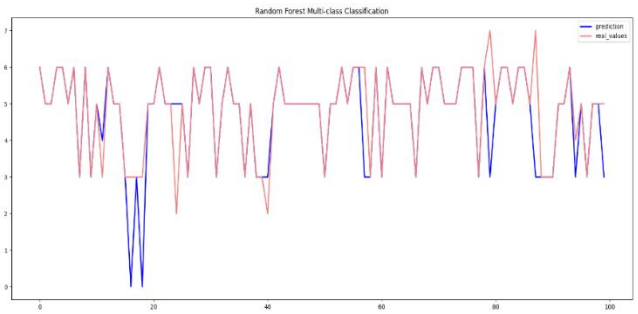
این مدل با استفاده از لایه‌های مختلف نورون‌ها می‌تواند الگوهای پیچیده داده‌ها را شناسایی کند. MLP از ورودی‌های چندگانه استفاده کرده و با اعمال توابع فعال‌ساز، روابط غیرخطی موجود در داده‌ها را بررسی می‌کند. با آموزش مناسب بر روی داده‌های نفوذ و غیرنفوذ، این الگوریتم قادر است به‌طور مؤثر حملات سایبری را شناسایی و تفکیک کند. نتایج تشخیص توسط این الگوریتم قبل و بعد از اجرای روش پیشنهادی در شکل ۹ نشان داده شده است.

۴-۲-۷- مقایسه مدل‌های یادگیری ماشین

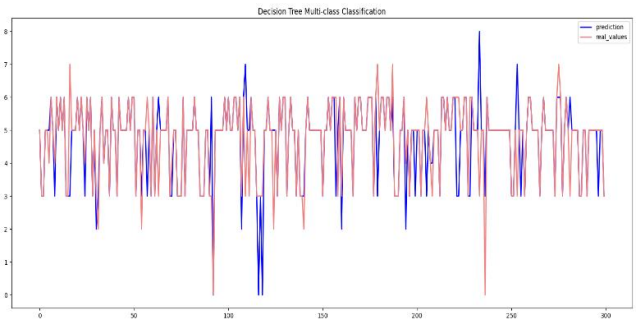
نتایج حاصل از به‌کارگیری شش مدل یادگیری ماشین در ترکیب با PCA در روش پیشنهادی بر حسب معیارهای دقت تک کلاسی، دقت چند کلاسی، نرخ مثبت کادب و زمان پردازش در جدول ۳ ارائه شده است. همان طور که در این جدول بیان شده، مدل رگرسیون لجستیک دقت تک‌کلاسی ۹۷,۸۴٪ و دقت چندکلاسی ۸۹,۸۱٪ را به ثبت رسانده است. مدل SVM خطی با دقت تک‌کلاسی ۹۷,۸۵٪ و دقت چندکلاسی ۸۹,۸۹٪ عملکرد خوبی را از خود نشان داده است. مدل KNN دقت تک‌کلاسی قابل توجهی با ۹۸,۳۱٪ دارد، با این حال دقت چندکلاسی آن کاهش می‌یابد و به ۸۸,۵۵٪ می‌رسد. جنگل تصادفی به عنوان یکی از قدرتمندترین مدل‌های یادگیری ماشین، دقت تک‌کلاسی ۹۸,۶۳٪ و دقت چندکلاسی ۸۹,۰۶٪ را نشان داده است. مدل درخت تصمیم دقت تک‌کلاسی ۹۸,۱۱٪ و دقت چندکلاسی ۸۵,۴۵٪ دارد. MLP به عنوان مدلی از شبکه‌های عصبی، دقت تک‌کلاسی ۹۸,۳۹٪ و دقت چندکلاسی ۸۹,۹۴٪ را به ثبت می‌رساند.



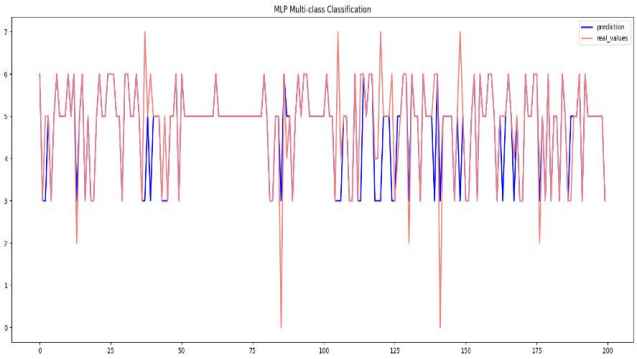
شکل ۶. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در K نزدیکترین همسایه



شکل ۷. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در جنگل تصادفی



شکل ۸. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در درخت تصمیم



شکل ۹. نتایج تشخیص نفوذ قبل و بعد ازاعمال PCA در شبکه‌ی عصبی چند لایه

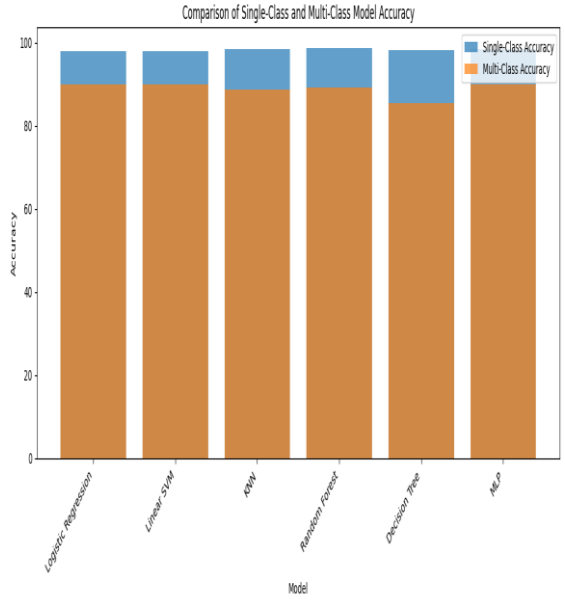
بر اساس نتایج به دست آمده جنگل تصادفی در ترکیب با PCA با دقت تک‌کلاسی ٪۹۸,۶۳ و نرخ مثبت کاذب ٪۳,۲ و زمان پردازش ۰,۶ میلی ثانیه بهترین عملکرد را از نظر دقت و زمان پردازش داشته است. در حالی که درخت تصمیم با دقت چندکلاسی ٪۸۵,۴۵ ضعیف‌ترین عملکرد را در مقایسه با سایر مدل‌ها دارد. زمان پردازش برای تمام مدل‌ها زیر ۱ میلی‌ثانیه است، که برای محیط‌های IoT با منابع محدود مناسب است. به منظور بررسی اهداف پژوهش می‌توان نتایج حاصل از روش پیشنهادی را بدین صورت ارزیابی کرد:

- **هدف اول(افزایش دقت):** چهار مدل از شش مدل مورد آزمایش به دقت بالاتر از ٪۹۸ دست یافتند. مدل جنگل تصادفی با ٪۹۸,۶۳ بهترین نتیجه را ارائه داد.
- **هدف دوم (کاهش نرخ مثبت کاذب):** همه‌ی مدل‌ها به هدف کاهش نرخ مثبت کاذب به کم‌تر از ٪۵ دست یافته‌اند. مدل‌های جنگل تصادفی و شبکه عصبی چند لایه کم‌ترین نرخ مثبت کاذب را دارند. این موضوع اهمیت به‌کارگیری PCA در کنار مدل مناسب را نشان می‌دهد.
- **هدف سوم (کاهش زمان پردازش):** زمان پردازش تمام مدل‌ها زیر ۱ میلی‌ثانیه است و مدل جنگل تصادفی زمان پردازش متوسط ۰,۶ میلی‌ثانیه را ثبت کرد که پایین‌تر از سایر مدل‌ها و قابل‌اجرا بر روی دستگاه‌های کم‌منبع است.

در پایان این بخش به منظور درک و مقایسه بهتر دقت مدل‌های یادگیری ماشین به کارگرفته شده در ترکیب با PCA نتایج دقت تک کلاسی و چند کلاسی مدل‌ها در نمودار شکل ۱۰ نیز نمایش داده شده است.

جدول ۳. نتایج حاصل از مدل‌های یادگیری ماشین

مدل یادگیری ماشین	دقت تک‌کلاسی (درصد)	دقت چندکلاسی (درصد)	نرخ مثبت کاذب (درصد)	زمان پردازش (میلی ثانیه)
رگرسیون لجستیک	۹۷,۸۴	۸۹,۸۱	۳,۵	۰,۸
SVM خطی	۹۷,۸۵	۸۹,۸۹	۳,۴	۰,۹
KNN	۹۸,۳۱	۸۸,۵۵	۴,۰	۰,۷
جنگل تصادفی	۹۸,۶۳	۸۹,۰۶	۳,۲	۰,۶
درخت تصمیم	۹۸,۱۱	۸۵,۴۵	۴,۵	۰,۷
شبکه عصبی چندلایه	۹۸,۳۹	۸۹,۹۴	۳,۳	۰,۸



شکل ۱۰. نتایج دقت مدل‌ها تک کلاسی و چند کلاسی

۵- بحث و نتیجه گیری

در این پژوهش مدلی ترکیبی مبتنی بر یادگیری ماشین و تحلیل مولفه‌های اصلی برای تشخیص نفوذ در شبکه‌های اینترنت اشیا ارائه شد که با استفاده از مجموعه داده UNSW-NB15 مورد ارزیابی قرار گرفت. نتایج تجربی پژوهش نشان‌دهنده دستیابی به اهداف کمی مدنظر، شامل دقت کلی ٪۹۸,۶۳، نرخ مثبت کاذب ۳-٪۴ و زمان پردازش کمتر از ۱ میلی‌ثانیه است. در بین مدل‌های یادگیری ماشین مورد استفاده مدل جنگل تصادفی بهترین عملکرد را با دقت تک‌کلاسی ٪۹۸,۶۳ و چندکلاسی ٪۸۹,۰۶ نشان داد، که به دلیل توانایی آن در مدیریت داده‌های پیچیده و کاهش بیش‌برازش است. استفاده از PCA ابعاد داده‌ها را کاهش داده و زمان پردازش را به کمتر از ۱ میلی‌ثانیه می‌رساند، که نسبت به روش‌های سنتی مانند SVM با زمان ۵-۱۰ میلی‌ثانیه بهبود قابل‌توجهی دارد. این ویژگی

به‌ویژه برای سیستم‌های اینترنت اشیاء که نیاز به پاسخ سریع دارند، بسیار مهم است و باعث می‌شود مدل پیشنهادی در مقایسه با دیگر روش‌ها عملکرد بهتری داشته باشد. مزایا و محدودیت‌های مدل پیشنهادی در مقایسه با با مدل‌های دیگر در جدول ۴ لیست شده است.

با توجه به بررسی‌های انجام شده می‌توان این گونه نتیجه گرفت که مدل پیشنهادی این قابلیت را دارد که بر روی دستگاه‌های با منابع محدود پیاده‌سازی شود. این ویژگی باعث می‌شود که سیستم پیشنهادی در مقایسه با سیستم‌های پیچیده‌تر و سنگین‌تر که نیاز به منابع محاسباتی بیشتری دارند، از مزیت قابل توجهی برخوردار باشد.

جدول ۴. مزایا و محدودیت‌های مدل پیشنهادی

مدل	مزایا	محدودیت
رگرسیون لجستیک	ساده – مناسب برای طبقه‌بندی باینری	دقت چندکلاسی کمتر ٪۸۹,۸۱
SVM خطی	تفکیک قوی با حاشیه بالا	زمان پردازش کمی بالاتر (۰,۹ میلی‌ثانیه)
KNN	دقت تک‌کلاسی بالا ٪۹۸,۳۱	دقت چندکلاسی کمتر ٪۸۸,۵۵
جنگل تصادفی	بالاترین دقت (٪۹۸,۶۳)، نرخ مثبت کاذب پایین ٪۳,۲	نیاز به تنظیم دقیق تعداد درختان
درخت تصمیم	یادگیری سریع	دقت چندکلاسی ضعیف ٪۸۵,۴۵
شبکه عصبی چندلایه	شناسایی الگوهای پیچیده	پیچیدگی محاسباتی بالاتر
روش پیشنهادی (PCA+RF)	دقت بالا، زمان پردازش کم، مناسب برایIoT	نیاز به تنظیم دقیق پارامترها

با توجه به نتایج به دست آمده پیشنهاد می‌شود در پژوهش‌های آتی از شبکه‌های عصبی عمیق و یادگیری تقویتی استفاده شود تا شبکه ای ایجاد شود که به دقت بالاتر در شناسایی تهدیدات جدید و ناشناخته دست یابد. همچنین، گنجاندن روش‌های تحلیل پیش‌بینی حملات مانند پیش‌بینی زمانی و مدل‌های سری زمانی می‌تواند به شناسایی حملات قبل از وقوع آن‌ها کمک کند. استفاده از داده‌های واقعی از شبکه‌های MANET و اینترنت اشیاء در محیط‌های دنیای واقعی می‌تواند قابلیت‌های سیستم را بهبود بخشد و اعتبار آن را در شرایط عملیاتی افزایش دهد. علاوه بر این، بررسی کارایی مدل در برابر حملات پیچیده مانند حملات DDoS هوشمند و حملات تزریق داده و تطبیق سیستم با شبکه‌های توزیع‌شده برای مقیاس‌پذیری بهتر می‌تواند به پیشرفت‌های عمده‌ای در امنیت شبکه‌های آینده منجر شود.

مراجع

[1] S. Sadhwani, B. Manibalan, R. Muthalagu, and P. Pawar, "A lightweight model for DDoS attack detection using machine learning techniques," Applied Sciences, vol. 13, no. 17, p. 9937, 2023.

[2] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari, and A. S. Avestimehr, "Federated learning for the internet of things: Applications, challenges, and opportunities," IEEE Internet of Things Magazine, vol. 5, no. 1, pp. 24-29, 2022.

[3] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. De Alvarenga, "A survey of intrusion detection in Internet of Things," Journal of Network and Computer Applications, vol. 84, pp. 25-37, 2017.

[4] M. Ahmid and O. Kazar, "A comprehensive review of the internet of things security," Journal of Applied Security Research, vol. 18, no. 3, pp. 289-305, 2023.

[5] N. Dat-Thinh, H. Xuan-Ninh, and L. Kim-Hung, "MidSiot: A multistage intrusion detection system for internet of things," Wireless Communications and Mobile Computing, vol. 2022, no. 1, p. 9173291, 2022.

[6] L. Strous, S. von Solms, and A. Zúquete, "Security and privacy of the Internet of Things," Computers & Security, vol. 102, p. 102148, 2021.

[7] S. Pandey and B. Bhushan, "Recent Lightweight cryptography (LWC) based security advances for resource-constrained IoT networks," Wireless Networks, vol. 30, no. 4, pp. 2987-3026, 2024.

[8] P. Fusco, A. Montefusco, G. P. Rimoli, F. Palmieri, and M. Ficco, "TinyML-Based Intrusion Detection System for Handling Class Imbalance in IoT-Edge Domain Using Siamese Neural Network on MCU," in International Conference on Advanced Information Networking and Applications, 2025: Springer, pp. 389-402.

[9] K. A. Da Costa, J. P. Papa, C. O. Lisboa, R. Munoz, and V. H. C. de Albuquerque, "Internet of Things: A survey on machine learning-based intrusion detection approaches," Computer Networks, vol. 151, pp. 147-157, 2019.

[10] E. Konstantopoulou, G. Athanasiou, and N. Sklavos, "Review and Analysis of FPGA and ASIC Implementations of NIST Lightweight Cryptography Finalists," ACM Computing Surveys, vol. 57, no. 10, pp. 1-35, 2025.

[11] H. Griffioen and C. Doerr, "Examining Mirai's battle over the Internet of Things," in Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 2020, pp. 743-756.

[12] M. Kintzlinger and N. Nissim, "Keep an eye on your personal belongings! The security of personal medical devices and their ecosystems," Journal of biomedical informatics, vol. 95, p. 103233, 2019.

[13] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," Journal of Network and Computer Applications, vol. 60, pp. 19-31, 2016.

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

[15] S. D. Babar and P. N. Mahalle, "A hash key-based key management mechanism for cluster-based wireless sensor network," Journal of Cyber Security and Mobility, pp. 73-88, 2016.

[16] A. Fatani, A. Dahou, M. A. Al-Qaness, S. Lu, and M. A. Elaziz, "Advanced feature extraction and selection approach using deep learning and Aquila optimizer for IoT intrusion detection system," Sensors, vol. 22, no. 1, p. 140, 2021.

[17] R. A. Disha and S. Waheed, "Performance analysis of machine learning models for intrusion detection system using Gini Impurity-

based Weighted Random Forest (GIWRF) feature selection technique," Cybersecurity, vol. 5, no. 1, p. 1, 2022.

[18] M. Mohy-Eddine, A. Guezzaz, S. Benkirane, and M. Azrour, "An effective intrusion detection approach based on ensemble learning for IIoT edge computing," Journal of Computer Virology and Hacking Techniques, vol. 19, no. 4, pp. 469-481, 2023.

[19] R. S. Tiwari, D. Lakshmi, T. K. Das, A. K. Tripathy, and K.-C. Li, "A lightweight optimized intrusion detection system using machine learning for edge-based IIoT security," Telecommunication Systems, pp. 1-20, 2024.

[20] M. J. Awan et al., "Real-time DDoS attack detection system using big data approach," Sustainability, vol. 13, no. 19, p

[21] W. Elmasry, A. Akbulut, and A. H. Zaim, "A Design of an Integrated Cloud-based Intrusion Detection System with Third Party Cloud Service" Open Computer Science, vol. 11, no. 1, 2021, pp. 365-379. https://doi.org/10.1515/comp-2020-0214.

[22] M. Sarhan, S. Layeghy, N. Moustafa, and M. Portmann, "Netflow datasets for machine learning-based network intrusion detection systems," in Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings 10, 2021: Springer, pp. 117-135.

[23] M. S. M. AL-inizi, Y. T. Alzubaidi, S. H. Oleiwi, N. A. A. Zahra, and J. F. Yonan, "Improvement Networks Intrusion Detection System Using Artificial Neural Networks (ANN)," in International Conference On Innovative Computing And Communication, 2024: Springer, pp. 571-587.