Passive Defence
Vol. 17, No.2, Summer 2026, Serial No. 66
ISSN: 2980-8030 , E-ISSN: 2008-6849

A Hybrid Framework for Android Malware Detection Using an Improved Particle Swarm Optimization and Fuzzy Kernel Support Vector Machine

Aliakbar Tajari Siahmarzkooh*
Correspondence: Assistant Professor, Department of Computer Science, Golestan University, Gorgan, Iran, Email Address: tajari1987@gmail.com

ARTICLE INFO

Article history:

Article Type: Research paper

Received: 8 October 2025

Received in revised form: 4 January 2026

Accepted: 13 July 2026

Available online: 21 August 2026

Keywords:

Malware Detection

Android

Particle Swarm Optimization

Support Vector Machine

Fuzzy Kernel

Feature Selection.

ABSTRACT

This paper presents a novel hybrid framework for Android malware detection that integrates an Improved Particle Swarm Optimization (IPSO) algorithm with a Fuzzy Kernel Support Vector Machine (FKSVM). The primary objective is to overcome the challenges of traditional methods, including high feature dimensionality, premature convergence, and the inability to model the inherent uncertainty in malware data. The IPSO algorithm, leveraging a dynamic inertia weight and a random mutation operator, prevents premature convergence and selects an optimal feature subset. Subsequently, the FKSVM model effectively manages data uncertainty through a novel fuzzy kernel. Evaluation of the proposed framework on the CIC-AndMal2020 dataset demonstrates that the method achieves a binary classification accuracy of 99.98% and an extremely low false alarm rate of 0.01%. These results clearly indicate the superior performance of the proposed framework compared to existing methods in terms of accuracy, speed, and robustness against adversarial attacks, thereby confirming its effectiveness.

Cite this article: A. Tajari Siahmarzkooh, "A Hybrid Framework for Android Malware Detection Using an Improved Particle Swarm Optimization and Fuzzy Kernel Support Vector Machine," Journal of Passive Defence, vol. 17, no. 2, pp. 75-90, 2026. DOI: https://doi.org/10.47176/pd.2026.1576

Publisher: Imam Hossein University.

Introduction

Due to its geographical location and its distinctive cultural, political, and social conditions, Iran has always been vulnerable to a wide range of natural and man-made crises. Over time, crisis management has evolved from a purely reactive approach into a systematic and institutionalized discipline that relies on structured planning, command and control systems, and coordinated emergency operations. In this context, the Ministry of Health has developed and communicated a national program entitled "Health System Response to Disasters," which is intended to guide and standardize crisis management within Iran's health sector and universities of medical sciences. The Emergency Operations Center (EOC) and the Incident Command System (ICS) constitute the core frameworks through which this program is expected to function during emergencies. Despite several years of implementation, evidence from the Qom Provincial Health Center suggests that the program has not been able to fully satisfy the requirements of crisis management and passive defense in this area. Ineffective resource allocation, fragmented inter-organizational coordination, and challenges in staff performance and data recording remain significant obstacles. Accordingly, this study aims to identify and analyze the challenges of implementing the "Health System Response to Disasters" program within the Qom Provincial Health Center. The study also seeks to propose practical exit strategies and solutions to overcome these barriers. To this end, the research applies a qualitative descriptive-analytical method and draws upon the Burke-Litwin organizational change model to frame and categorize the findings. The findings are expected to enhance the effectiveness, efficiency, and productivity of the response program and to improve crisis management and public participation, ultimately reducing the risks and consequences of disasters at the community level.

Results and Discussion

1. Research Approach and Data Sources

This study employed a qualitative descriptive-analytical design. Data were collected from three primary sources: (a) the existing research literature on crisis/disaster management, (b) official organizational documents of the Qom Provincial Health Center (including technical unit job descriptions and the Ministry of Health's "Health System Response to Disasters" program), and (c) semi-structured interviews with crisis-management experts. All collected data were analyzed and categorized using the **Burke–Litwin model of organizational change**, which provided 12 analytical criteria.

2. Research Questions (Framework for Discussion)

The analysis was organized around four guiding research questions: (1) How should the health care system response program be implemented in response to disasters? (2) What is the current status of implementing the response program in the Qom Provincial Health Center? (3) What are the reasons and challenges of the program's implementation approach? (4) What is the proposed model for preparing, implementing, and evaluating the health care system's response program?

3. Key Findings — The Extracted Challenges

From the categorized analysis, the study extracted **45 important challenges**, which were structured into **4. main categories** and **12 subcategories**. The Burke–Litwin framework grouped these findings across its 12 criteria — including external environment, leadership, mission/strategy, organizational culture, structure, management practices, systems, and individual/group variables.

The 4 main challenge categories and their typical focus areas were:

- **Structural and organizational challenges** — relating to command-and-control structure, the Incident Command System (ICS) implementation, and the Emergency Operations Center (EOC) operational readiness.
- **Inter-organizational coordination and integration challenges** — including the coordination of health system units with other crisis-management organizations and the management of inter-organizational relationships.
- **Resource and management-practice challenges** — covering human resources, staff performance, resource allocation, and management systems.
- **Information, monitoring, and community-related challenges** — including accurate data recording, program monitoring, public participation, and community behavior.

4. Figures Referenced in the Study

The article presents several supporting figures and frameworks: (1) an organizational chart of the Qom Provincial Health Center's crisis-management structure and its covered population; (2) a model of the "Health System Response to Disasters" phases adapted from Ministry of Health guidelines (Reference [3]); (3) a diagram showing the Incident Command System (ICS) structure, which includes the incident commander and functional sections such as operations, planning, logistics, and finance/administration; and (4) a flowscheme of the crisis-management process covering planning, response, and evaluation stages.

5. Expected Theoretical Results (from the solutions/exit strategies)

According to the abstract, the provision of exit strategies and solutions is expected to yield the following results: program implementation, promotion, effectiveness, efficiency, and productivity; improved crisis response; accurate information recording; enhanced staff performance; improved community behavior; strengthened crisis management; greater public participation; and better program monitoring. Finally, all findings were consolidated into a final conceptual model.

6. Discussion — Interpretation of the Data

The findings indicate that, despite several years of program implementation, the Qom Provincial Health Center still exhibits gaps in fully meeting crisis-management and passive-defense requirements. The concentration of challenges across structural, coordination, resource, and information domains confirms that the obstacles are systemic rather than isolated. By framing these challenges through the Burke–Litwin model, the study links organizational-level factors (structure, strategy, culture) with individual/group performance, providing a coherent theoretical basis for the proposed solutions. The expected outcomes point toward effective improvement of the health care system, promotion of disaster-preparedness components, and reduction of disaster risks and consequences at the community level.

Conclusion

This study set out to identify and analyze the challenges of implementing the "Health System Response to Disasters" program within the Qom Provincial Health Center, and to propose practical exit strategies for overcoming these barriers. Using a qualitative descriptive-analytical approach and the Burke–Litwin organizational change model as the analytical framework, the research triangulated data from three sources—the existing literature, official organizational documents, and semi-structured interviews with crisis-management experts.

The analysis revealed that the program, despite several years of implementation, has not fully satisfied the requirements of crisis management and passive defense in the studied center. A total of 45 important challenges were extracted and structured into 4 main categories and 12 subcategories, spanning structural and organizational dimensions, inter-organizational coordination, resource and management practices, and information, monitoring, and community engagement. The concentration of obstacles across all four domains indicates that the deficiencies are systemic rather than isolated, requiring coherent and integrated solutions.

The principal product of this research is a consolidated conceptual model that links the extracted challenges with the proposed exit strategies. The model, built upon the twelve criteria of the Burke–Litwin framework, treats organizational-level factors and individual/group performance as interconnected, thereby providing a theoretical basis for targeted interventions.

Ultimately, the provision of exit strategies is expected to yield improvement in program implementation, effectiveness, efficiency, and productivity; accurate information recording; enhanced staff performance; improved community behavior; stronger crisis management; greater public participation; and effective program monitoring. In doing so, the study contributes to the effective improvement of the health care system, the promotion of disaster-preparedness components, and the reduction of disaster risks and consequences at the community level.

For future work, the proposed model should be tested and operationalized at other provincial health centers and universities of medical sciences to verify its generalizability. In addition, further research is recommended to strengthen Emergency Operations Center readiness, enhance inter-organizational coordination with passive-defense and crisis-management entities, establish robust data-recording and monitoring mechanisms, and expand public participation in disaster response. Implementing these recommendations will help institutionalize an effective and resilient response program across Iran's health system.

پدافند غیرعامل

سال هفدهم، شماره ۲، تابستان ۱۴۰۵، (پیاپی ۶۶): صص ۹۰-۷۵

شاپای چاپی: ۶۹۴۹-۲۰۰۸ | شاپای الکترونیکی: ۸۰۳۰-۲۹۸۰

ارائه چارچوب ترکیبی تشخیص بدافزارهای اندرویدی با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته و ماشین بردار پشتیبان با هسته فازی

علی‌اکبر تجری سیاه‌مرزکوه*

استادیار، گروه علوم کامپیوتر، دانشگاه گلستان، گروه علوم کامپیوتر، گرگان، ایران (نویسنده مسئول). رایانامه: tajari1987@gmail.com

چکیده

این مقاله یک چارچوب ترکیبی نوآورانه برای تشخیص بدافزار اندروید ارائه می‌دهد که یک الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته (IPSO) را با یک ماشین بردار پشتیبان فازی با هسته (FKSVM) تلفیق می‌کند. هدف اصلی، غلبه بر چالش‌های روش‌های سنتی، از جمله ابعاد بالای ویژگی‌ها، همگرایی زودرس و ناتوانی در مدل‌سازی عدم قطعیت ذاتی در داده‌های بدافزاری است. الگوریتم IPSO با بهره‌گیری از یک وزن پویا و یک عملگر جهش تصادفی، از همگرایی زودرس جلوگیری کرده و یک زیرمجموعه ویژگی بهینه را انتخاب می‌کند. در ادامه، مدل FKSVM با استفاده از یک هسته فازی جدید، عدم قطعیت داده‌ها را به‌طور مؤثر مدیریت می‌نماید. ارزیابی چارچوب پیشنهادی بر روی مجموعه‌داده CIC-AndMal2020 نشان می‌دهد که این روش در طبقه‌بندی دودویی به دقت ۹۹/۹۸ درصد و نرخ هشدار نادرست بسیار پایین ۰/۰۱ درصد دست می‌یابد. این نتایج، عملکرد برتر چارچوب ارائه‌شده را در مقایسه با روش‌های موجود از نظر دقت، سرعت و مقاومت در برابر حملات متخاصم به وضوح نشان می‌دهد و اثربخشی آن را تأیید می‌کند.

مشخصات مقاله

تاریخچه مقاله:

نوع مقاله: علمی پژوهشی

دریافت: ۱۴۰۴/۰۷/۱۶

بازنگری: ۱۴۰۴/۱۰/۱۴

پذیرش: ۱۴۰۵/۰۴/۲۲

ارائه آنلاین: ۱۴۰۵/۰۵/۳۰

کلیدواژه‌ها:

تشخیص بدافزار

اندروید

بهینه‌سازی ازدحام ذرات

ماشین بردار پشتیبان

هسته فازی

انتخاب ویژگی

استناد: تجری سیاه‌مرزکوه، علی‌اکبر، " ارائه چارچوب ترکیبی تشخیص بدافزارهای اندرویدی با استفاده از الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته و ماشین بردار پشتیبان با هسته فازی"، نشریه پدافند غیرعامل، دوره ۱۷، شماره۲، صفحات ۹۰-۷۵، ۱۴۰۵.

DOI: https://doi.org/10.47176/pd.2026.1584

ناشر: دانشگاه جام امام حسین(ع).

۱- مقدمه

در عصر حاضر، تلفن‌های همراه هوشمند به بخشی جدایی‌ناپذیر از زندگی روزمره میلیاردها نفر در سراسر جهان تبدیل شده‌اند. در این میان، سیستم عامل اندروید با سهم بازاری غالب، به پلتفرمی محبوب برای توسعه‌دهندگان و کاربران بدل گشته است. این محبوبیت و گستردگی، در کنار ماهیت باز[1] این سیستم عامل، آن را به هدفی جذاب و سودآور برای مهاجمان سایبری تبدیل کرده است. بر اساس گزارش‌های معتبر مراکز امنیتی همچون «سوفوس» و «کاسپرسکی لاب»، سالانه میلیون‌ها نمونه بدافزار جدید اندرویدی شناسایی می‌شود که هر یک با اهداف مخرب متنوعی از جمله سرقت اطلاعات حساس مالی و شخصی، اخاذی، جاسوسی، ایجاد اختلال در سرویس‌ها و استفاده از منابع دستگاه در باتنت‌ها طراحی می‌شوند. این روند صعودی، تهدیدی جدی برای حریم خصوصی افراد، امنیت سازمان‌ها و حتی امنیت ملی کشورها محسوب می‌شود.

بدافزارهای اندرویدی در قالب‌های گوناگونی از جمله «اسب‌های تروجان[2]» ، «باج‌افزارها[3]»، «جاسوس‌افزارها[4]»، «ماینرها» و «آگهی‌افزارها[5]» ظاهر می‌شوند. آنچه مقابله با این تهدیدات را پیچیده‌تر می‌کند، استفاده مهاجمان از تکنیک‌های پیشرفته «پوشش کد[6]» و «چندریختی[7]» است. در تکنیک پوشش کد، ساختار و کد منبع برنامه مخرب بدون تغییر در عملکرد آن، دستخوش تغییر می‌شود تا از شناسایی توسط آنتی‌ویروس‌ها جلوگیری کند [۱]. تکنیک چندریختی نیز امکان ایجاد نسخه‌های متعدد و خودتغییردهنده از یک بدافزار را فراهم می‌آورد، به گونه‌ای که هر نمونه دارای امضای منحصربه‌فردی باشد. این روش‌ها، اثرگذاری روش‌های سنتی تشخیص را به شدت کاهش می‌دهند.

روش‌های سنتی تشخیص بدافزار عمدتاً بر پایه «امضا[8]» استوار هستند. در این روش‌ها، یک پایگاه داده از امضاهای منحصربه‌فرد (مانند درهم‌سازی فایل یا رشته‌های ثابت در کد) که از نمونه‌های شناخته‌شده بدافزار استخراج شده‌اند، نگهداری می‌شود. سپس، هر فایل جدید با این امضاها مقایسه می‌شود و در صورت مطابقت، به عنوان مخرب شناسایی می‌گردد. اگرچه این روش در شناسایی بدافزارهای شناخته‌شده سریع و کارآمد است، اما دارای محدودیت‌های بنیادینی می‌باشد: نخست آنکه در برابر بدافزارهای نوظهور روز صفر[9] و نمونه‌های چندریخت کاملاً ناتوان است، چرا که امضای آن‌ها در پایگاه داده وجود ندارد. دوم آنکه نیازمند به‌روزرسانی مستمر و پرهزینه پایگاه داده امضاها است. و سوم آنکه با حجم انبوه بدافزارهای جدید، اندازه این پایگاه داده به طور تصاعدی افزایش یافته و فرآیند جستجو و مقایسه را کند می‌سازد.

برای فائق آمدن بر این چالش‌ها، محققان به سمت استفاده از روش‌های «هوشمند» مبتنی بر یادگیری ماشین و یادگیری عمیق گرایش یافته‌اند. این روش‌ها به جای جستجوی یک امضای ثابت، به تحلیل الگوهای رفتاری و ویژگی‌های ذاتی برنامه‌ها می‌پردازند. آن‌ها قادرند با یادگیری از داده‌های تاریخی، مدلی را ایجاد کنند که می‌تواند نمونه‌های مخرب را حتی اگر قبلاً دیده نشده باشند، بر اساس شباهت‌های ساختاری و رفتاری به نمونه‌های مخرب قبلی، شناسایی کند. این رویکرد، تحولی اساسی در عرصه امنیت سایبری ایجاد کرده است.

اگرچه روش‌های مبتنی بر یادگیری ماشین موفقیت‌های چشمگیری داشته‌اند، اما خود با چالش‌های جدیدی روبرو هستند. یکی از اصلی‌ترین این چالش‌ها، «بعد بالای ویژگی‌ها[10]» است [۲]. در مجموعه داده‌های بدافزاری مدرن، هزاران ویژگی از هر نمونه استخراج می‌شود. وجود این تعداد زیاد ویژگی، چندین مشکل ایجاد می‌کند: نیاز به قدرت پردازشی و حافظه بسیار زیاد، افزایش زمان آموزش مدل، و کاهش دقت طبقه‌بندی به دلیل وجود ویژگی‌های زائد، نامربوط و پرت. بسیاری از این ویژگی‌ها ممکن است هیچ اطلاعات مفیدی برای تمایز بین بدافزار و نرم‌افزار عادی ارائه ندهند و حتی باعث «بیش‌برازش[11]» مدل شوند؛ یعنی مدل بر روی داده‌های آموزشی بسیار خوب عمل کند، اما در مواجهه با داده‌های جدید و دیده نشده عملکرد ضعیفی داشته باشد. از این رو، انتخاب ویژگی به عنوان یک گام حیاتی در فرآیند ساخت مدل‌های کارآمد مطرح می‌شود. هدف از انتخاب ویژگی، یافتن یک زیرمجموعه بهینه و حداقلی از

[1] Open-Source
[2] Trojans
[3] Ransomware
[4] Spyware
[5] Adware
[6] Obfuscation
[7] Polymorphism
[8] Signature-Based
[9] Zero-Day
[10] Curse of Dimensionality
[11] Overfitting

ویژگی‌ها است که بیشترین قدرت تبعیض را بین کلاس‌ها داشته باشد. این کار نه تنها پیچیدگی محاسباتی را کاهش می‌دهد، بلکه دقت، کارایی و قابلیت تفسیر مدل نهایی را افزایش می‌دهد. الگوریتم‌های بهینه‌سازی فرااکتشافی به دلیل توانایی در جستجوی کارآمد فضاهای بزرگ و پیچیده، گزینه‌ای ایده‌آل برای مسئله انتخاب ویژگی هستند. در میان این الگوریتم‌ها، الگوریتم بهینه‌سازی ازدحام ذرات[1] یا PSO به دلیل سادگی مفهومی، سرعت همگرایی بالا و کارایی اثبات‌شده، مورد توجه گسترده‌ای قرار گرفته است. با این حال، نسخه استاندارد PSO دارای ضعف اصلی همگرایی زودرس و گیر کردن در بهینه‌های محلی است. هنگامی که این اتفاق می‌افتد، ازدحام ذرات تنوع خود را از دست داده و در نقطه‌ای از فضای جستجو که لزوماً بهترین جواب ممکن نیست، متوقف می‌شوند. این امر می‌تواند منجر به انتخاب یک زیرمجموعه بهینه از ویژگی‌ها گردد.

چالش دیگر، انتخاب یک طبقه‌بند قدرتمند و مقاوم است. ماشین بردار پشتیبان[2] یا SVM یکی از پرکاربردترین و مؤثرترین طبقه‌بندها در حوزه تشخیص بدافزار است که به دلیل پایه نظری قوی و عملکرد خوب در فضاهای با بعد بالا شناخته می‌شود. هسته در SVM این امکان را فراهم می‌آورد تا داده‌هایی که به صورت خطی جداپذیر نیستند، به یک فضای با بعد بالاتر نگاشت شده و در آن فضا به صورت خطی جدا شوند. با این حال، هسته‌های مرسوم مانند RBF یا Polynomial در مواجهه با عدم قطعیت و نویز ذاتی در داده‌های بدافزاری ممکن است عملکرد مطلوبی نداشته باشند [۳ و ۴]. برای مثال، ممکن است در برچسب‌گذاری اولیه داده‌ها خطا وجود داشته باشد، یا برخی نمونه‌ها دارای ویژگی‌هایی باشند که در مرز بین دو کلاس قرار می‌گیرند. یک هسته سنتی نمی‌تواند این عدم قطعیت را به طور مستقیم مدلسازی کند.

برای پر کردن این شکاف‌های تحقیقاتی، این مقاله یک روش ترکیبی نوین را پیشنهاد می‌دهد. نوآوری اصلی این تحقیق در سه محور اساسی خلاصه می‌شود:

۱. ارائه یک الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته برای انتخاب ویژگی. در این بهبود، از یک ضریب اینرسی پویا استفاده می‌شود که به تدریج در طول فرآیند بهینه‌سازی کاهش می‌یابد تا تعادل مناسبی بین جستجوی سراسری و جستجوی محلی برقرار کند. علاوه بر این، یک عملگر جهش تصادفی نیز به الگوریتم اضافه شده است که در صورت کاهش تنوع جمعیت، به ذرات اجازه می‌دهد تا با جهش به موقعیت‌های جدید، از دام بهینه‌های محلی فرار کنند.

۲. طراحی و پیاده‌سازی یک هسته فازی نوین برای ماشین بردار پشتیبان. این هسته با تلفیق نظریه مجموعه‌های فازی با ماشین بردار پشتیبان، یک ماشین بردار پشتیبان با هسته فازی ایجاد می‌کند. وظیفه این هسته، تخصیص درجه عضویت به هر نمونه داده است که میزان تعلق آن نمونه به یک کلاس خاص را نشان می‌دهد. این امر باعث می‌شود مدل در برابر داده‌های پرت و نویزی مقاوم‌تر شده و بتواند عدم قطعیت موجود در مرزهای بین کلاس‌ها را به خوبی مدیریت کند.

۳. تلفیق هوشمندانه این دو جزء در یک چارچوب یکپارچه. در این چارچوب، ابتدا الگوریتم IPSO بر روی مجموعه داده آموزشی اجرا شده و بهینه‌ترین زیرمجموعه ویژگی را شناسایی می‌کند. سپس، تنها این ویژگی‌های انتخاب‌شده به مدل FKSVM تغذیه شده و مدل نهایی آموزش می‌بینند. این همکاری منجر به ایجاد یک سیستم تشخیصی می‌شود که هم از نظر دقت و هم از نظر سرعت و مقاومت، برتری قابل توجهی نسبت به روش‌های موجود دارد.

ساختار این مقاله به این صورت است که در بخش ۲، پیشینه تحقیق و کارهای مرتبط به طور مفصل مرور می‌شوند. در بخش ۳، روش پیشنهادی شامل جزئیات الگوریتم IPSO، هسته فازی و چارچوب ترکیبی به تفصیل شرح داده می‌شود. بخش ۴ به ارائه نتایج تجربی گسترده، شامل مجموعه داده مورد استفاده، معیارهای ارزیابی، تنظیم پارامترها و مقایسه کمی و کیفی با روش‌های پیشین اختصاص دارد. در نهایت، در بخش ۵، نتیجه‌گیری کلی از پژوهش ارائه شده و راهکارهایی برای تحقیقات آینده پیشنهاد می‌گردد.

۲- ادبیات و پیشینه تحقیق

با توجه به اهمیت روزافزون امنیت دستگاه‌های اندرویدی، تحقیقات گسترده‌ای در زمینه تشخیص بدافزار با استفاده از روش‌های هوشمند انجام شده است. این تحقیقات را می‌توان بر اساس نوع ویژگی‌های استخراج‌شده، الگوریتم‌های بهینه‌سازی

[1] Particle Swarm Optimization
[2] Support Vector Machine

مورد استفاده و طبقه‌بندهای به کار رفته دسته‌بندی نمود. در این بخش، به بررسی مهم‌ترین و مرتبط‌ترین کارهای انجام شده در این حوزه می‌پردازیم.

در حوزه روش‌های مبتنی بر ویژگی‌های ایستا، آرپ و همکاران [۵]، در سال ۲۰۱۴ روش DREBIN را معرفی کردند که یک چارچوب تشخیصی مبتنی بر تحلیل ایستا است. در این روش، ویژگی‌هایی همچون مجوزهای درخواستی، فراخوانی‌های API خطرناک، آدرس‌های شبکه و اجزای سخت‌افزار مستقیماً از فایل APK استخراج شده و در یک فضای ویژگی یکپارچه ادغام می‌شوند. سپس از یک ماشین بردار پشتیبان خطی برای دسته‌بندی نمونه‌ها استفاده می‌شود. اگرچه DREBIN به دلیل سادگی و قابلیت تفسیرپذیری بالا مورد استقبال قرار گرفت، اما در برابر تکنیک‌های پیشرفته پوشش کد و بدافزارهای چندریخت مقاومت کمی از خود نشان می‌دهد. همچنین، حجم بالای ویژگی‌های استخراج‌شده می‌تواند منجر به افزایش بار محاسباتی شود.

میلانی و همکاران [۶]، در سال ۲۰۱۹ از الگوریتم جنگل تصادفی برای تشخیص بدافزارهای اندرویدی استفاده کردند. آن‌ها بر استخراج ویژگی‌های ساختاری از فایل مانیفست، از قبیل مجوزها و intentها تمرکز نمودند. نتایج پژوهش آن‌ها نشان داد که این روش می‌تواند به دقت بالایی دست یابد. با این حال، نقطه ضعف اصلی مدل پیشنهادی آن‌ها، آسیب‌پذیری در برابر بدافزارهای چندریختی و نیز عدم استفاده از مکانیزمی برای کاهش ابعاد فضای ویژگی بود که این امر می‌توانست بر کارایی مدل در مجموعه‌داده‌های بسیار بزرگ تأثیر منفی بگذارد.

در زمینه استفاده از الگوریتم‌های بهینه‌سازی برای انتخاب ویژگی، چن و همکاران [۷]، در سال ۲۰۲۰ از الگوریتم ژنتیک برای انتخاب زیرمجموعه بهینه‌ای از ویژگی‌ها از یک مجموعه داده بزرگ متشکل از هزاران برنامه اندرویدی استفاده کردند. ویژگی‌های انتخابی سپس به یک طبقه‌بند بیز ساده تغذیه شد. نتایج تجربی آن‌ها حاکی از بهبود قابل توجه دقت در ازای کاهش چشمگیر تعداد ویژگی‌ها بود. با این وجود، سرعت همگرایی الگوریتم ژنتیک در مقایسه با سایر الگوریتم‌های فرااکتشافی می‌تواند در برخی موارد کند باشد و همچنین تنظیم پارامترهای متعدد آن چالشی اساسی محسوب می‌شود.

لیو و همکاران [۸]، در سال ۲۰۲۱ به طور خاص از الگوریتم بهینه‌سازی ازدحام ذرات برای حل مسئله انتخاب ویژگی در تشخیص بدافزار اندروید بهره بردند. در این پژوهش، هر ذره نمایانگر یک زیرمجموعه احتمالی از ویژگی‌ها بود و تابع برازش بر اساس دقت یک طبقه‌بند ماشین بردار پشتیبان طراحی شده بود. یافته‌های آن‌ها به وضوح نشان داد که PSO قادر است یک زیرمجموعه بهینه از ویژگی‌ها را شناسایی کند که منجر به عملکرد بهتر طبقه‌بند نهایی می‌شود. با این حال، نویسندگان به مسئله همگرایی زودرس در الگوریتم استاندارد PSO و گیر کردن آن در بهینه‌های محلی اذعان داشتند، که این همان شکافی است که پژوهش حاضر قصد پر کردن آن را دارد.

در حوزه روش‌های مبتنی بر یادگیری عمیق، وو و همکاران [۹]، در سال ۲۰۲۲ یک شبکه عصبی کانولوشنی یک‌بعدی را برای پردازش مستقیم دنباله‌های opcode به عنوان داده ورودی پیشنهاد دادند. این روش با حذف نیاز به استخراج دستی ویژگی‌های پیچیده، توانایی یادگیری خودکار ویژگی‌های انتزاعی و سلسله‌مراتبی را دارا بود. اگرچه این معماری نتایج امیدوارکننده‌ای ارائه داد، اما اشکال اصلی آن نیاز مبرم به حجم بسیار عظیمی از داده برای آموزش و همچنین نیاز به منابع سخت‌افزاری قدرتمند (مانند GPU) برای رسیدن به زمان آموزش معقول است که این امر استقرار آن را روی دستگاه‌های موبایل با منابع محدود با چالش مواجه می‌سازد.

ژانگ و همکاران [۱۰]، در سال ۲۰۲۱ از یک شبکه عصبی بازگشتی (RNN) با حافظه بلند-کوتاه‌مدت (LSTM) برای تحلیل رفتار پویای برنامه‌های اندرویدی استفاده کردند. در این روش، دنباله‌ای از فراخوانی‌های سیستمی که در زمان اجرای برنامه در یک محیط sandbox ثبت شده بود، به عنوان ورودی به مدل ارائه می‌شد. مدل LSTM با توجه به ماهیت دنباله‌ای داده‌ها، توانایی مدلسازی وابستگی‌های بلندمدت در رفتار برنامه را داشت و در شناسایی بدافزارهای مبتنی بر رفتارهای پیچیده و زمان‌بر موفق عمل کرد. با این حال، فرآیند استخراج لاگ‌های پویا بسیار پرهزینه، زمان‌بر و نیازمند محیط‌های شبیه‌سازی پیچیده است که این موضوع مقیاس‌پذیری روش را تحت تأثیر قرار می‌دهد.

برای حل مسئله عدم قطعیت ذاتی در داده‌های بدافزاری، برخی از محققان به سمت تلفیق منطق فازی با روش‌های تشخیص رفته‌اند. علی‌زاده و همکاران [۱۱] در سال ۲۰۱۹ یک سیستم استنتاج فازی برای تشخیص بدافزار طراحی کردند. در

این سیستم، از مجموعه‌ای از قواعد "اگر-آنگاه" (If-Then) فازی که بر اساس دانش متخصصان حوزه امنیت تعریف شده بود، استفاده گردید. این سیستم در مدیریت عدم قطعیت و ارائه تصمیم‌های تفسیرپذیر موفق بود. اما عیب اصلی این رویکرد، وابستگی شدید به دانش دامنه و متخصص انسانی برای طراحی قواعد است که این امر، باعث می‌شود سیستم برای مجموعه‌داده‌های بسیار بزرگ و پیچیده به خوبی مقیاس‌پذیر نباشد.

ترکیب الگوریتم‌های بهینه‌سازی و یادگیری ماشین نیز در کانون توجه پژوهشگران قرار داشته است. وانگ و همکاران [۱۲] در سال ۲۰۲۰ از الگوریتم کلونی زنبور عسل برای انتخاب ویژگی و سپس از ماشین بردار پشتیبان برای طبقه‌بندی نهایی استفاده کردند. آن‌ها گزارش کردند که این ترکیب به نتایج قابل قبولی از نظر دقت دست یافته است. اگرچه ABC یک الگوریتم قدرتمند است، اما برخی مطالعات نشان داده‌اند که سرعت همگرایی آن در مقایسه با PSO در برخی مسائل بهینه‌سازی کندتر است. همچنین، عملکرد آن می‌تواند به شدت به تنظیم پارامترهای کنترل‌کننده جستجو وابسته باشد.

دارما و همکاران [۱۳]، در سال ۲۰۲۲ از الگوریتم گرگ خاکستری برای انتخاب ویژگی در یک چارچوب تشخیص بدافزار استفاده نمودند. آن‌ها ادعا کردند که GWO به دلیل مکانیزم شکار سلسله‌مراتبی خود، می‌تواند تعادل بهتری بین جستجوی سراسری و محلی برقرار کند. با این حال، GWO نیز در مسائل پیچیده ممکن است با چالش همگرایی زودرس روبرو شود.

در زمینه استفاده از ماشین بردار پشتیبان پیشرفته، لی و همکاران [۱۴] در سال ۲۰۲۱ استفاده از یک هسته ترکیبی جدید برای SVM را پیشنهاد دادند. اگرچه عملکرد مدل آن‌ها بهبود یافت، اما هسته پیشنهادی آن‌ها به طور صریح مسئله عدم قطعیت در داده‌ها را مورد توجه قرار نداده بود. از سوی دیگر، اسمیت و جانسون [۱۵]، در سال ۲۰۱۹ یک چارچوب تشخیص ناهنجاری مبتنی بر SVM با هسته RBF ارائه دادند که اگرچه در شناسایی حملات ناشناخته مفید بود، اما نرخ هشدار نادرست نسبتاً بالایی داشت.

علاوه بر این، پژوهش‌های ترکیبی نیز مورد بررسی قرار گرفته‌اند. به عنوان مثال، گارسیا و همکاران [۱۶] در سال ۲۰۲۳ یک مدل هیبرید متشکل از CNN برای استخراج ویژگی و LSTM برای مدل‌سازی توالی زمانی رفتار برنامه ارائه کردند. این مدل اگرچه از دقت بسیار بالایی برخوردار بود، اما پیچیدگی محاسباتی فوق‌العاده زیادی داشت و برای استقرار بلادرنگ مناسب نبود. کیم و همکاران [۱۷] نیز در سال ۲۰۲۲ از یادگیری انتقال با استفاده از یک مدل از پیش آموزش دیده BERT برای تحلیل کد اسمبلی برنامه‌های اندرویدی استفاده کردند که نیازمند منابع محاسباتی عظیم بود.

با بررسی دقیق و مقایسه تطبیقی پژوهش‌های پیشین، می‌توان شکاف‌های تحقیقاتی اصلی را به صورت زیر خلاصه کرد:

۱. عدم وجود یک مکانیزم انتخاب ویژگی کارآمد و مقاوم در برابر همگرایی زودرس که بتواند به طور پایدار بهینه سراسری را در فضای ویژگی‌های با بعد بالا پیدا کند.

۲. ناتوانی طبقه‌بندهای مرسوم مانند SVM استاندارد در مدلسازی صریح عدم قطعیت و نویز موجود در داده‌های بدافزاری واقعی.

۳. کمبود یک چارچوب یکپارچه که بتواند به طور همزمان مسئله انتخاب ویژگی بهینه و طبقه‌بندی مقاوم را با در نظر گرفتن محدودیت‌های منابع در محیط‌های عملیاتی حل کند.

بنابراین، این مقاله با ارائه یک چارچوب ترکیبی نوین متشکل از الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته (IPSO) برای انتخاب ویژگی و ماشین بردار پشتیبان با هسته فازی (FKSVM) برای طبقه‌بندی، به طور مستقیم به پر کردن این شکاف‌ها می‌پردازد. رویکرد پیشنهادی حاضر نه تنها از مزایای هر دو جزء تشکیل‌دهنده خود بهره می‌برد، بلکه با ایجاد یک همکاری سینرژیک، محدودیت‌های هر یک را نیز پوشش می‌دهد.

۳- روش تحقیق

در این بخش، چارچوب پیشنهادی برای تشخیص بدافزارهای اندرویدی که مبتنی بر ترکیب الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته (IPSO) و ماشین بردار پشتیبان با هسته فازی (FKSVM) است، به تفصیل شرح داده می‌شود. این چارچوب از چهار مرحله اصلی تشکیل شده است: (۱) پیش‌پردازش داده و استخراج ویژگی، (۲) انتخاب ویژگی بهینه با استفاده از الگوریتم IPSO، (۳) دسته‌بندی با استفاده از FKSVM، و (۴) ارزیابی مدل. شکل (۱) نمای شماتیک کلی این چارچوب را نشان می‌دهد.

شکل (۱): چارچوب کلی سیستم پیشنهادی

۳-۱- پیش‌پردازش داده و استخراج ویژگی

در این مقاله از مجموعه داده جامع CIC-AndMal2020 [۱۸] استفاده شده است که شامل ۲۰۰،۰۰۰ نمونه برنامه اندرویدی (۱۰۰،۰۰۰ بدافزار از ۱۴۵ خانواده مختلف و ۱۰۰،۰۰۰ برنامه عادی) است. این مجموعه داده شامل ۷۷۰ ویژگی استخراج‌شده از تحلیل ایستا و پویا است که شامل مجوزها، فراخوانی‌های API ریسکی، دسترسی به سخت‌افزار، فعالیت‌های شبکه و سایر ویژگی‌های رفتاری است. توزیع کلاس‌ها در مجموعه داده CIC-AndMal2020 از نظر تعداد نمونه‌ها نسبتاً متوازن است. این مجموعه داده شامل ۱۰۰،۰۰۰ نمونه بدافزار و ۱۰۰،۰۰۰ نمونه نرم‌افزار عادی است. با این حال، به دلیل تفاوت‌های رفتاری بین خانواده‌های مختلف بدافزار و ناهمگونی ذاتی داده‌های امنیتی، تحلیل توازن کلاس‌ها تنها به برابری تعداد نمونه‌ها محدود نمی‌شود. از این رو، در ارزیابی عملکرد مدل، علاوه بر دقت کلی، از معیارهای مقاوم به عدم‌توازن نظیر امتیاز F1 و AUC نیز استفاده شده است تا ارزیابی جامع‌تری ارائه گردد.

مراحل پیش‌پردازش داده‌ها به شرح زیر است:

حذف داده‌های تکراری: نمونه‌های کاملاً یکسان از نظر تمامی ویژگی‌ها با استفاده از تابع drop_duplicates() در کتابخانه pandas حذف شدند.

مدیریت مقادیر گمشده: ویژگی‌هایی که بیش از ۶۰٪ مقادیرشان گمشده بود، به طور کامل حذف شدند. برای مقادیر گمشده پراکنده در سایر ویژگی‌ها، از میانه هر ستون برای پر کردن استفاده گردید [۱۹].

کدگذاری برچسب‌ها: برچسب‌های چندکلاسه به صورت One-Hot Encoding کدگذاری شدند.

نرمال‌سازی: از نرمال‌سازی Min-Max برای مقیاس‌بندی همه ویژگی‌ها در بازه [۰،۱] استفاده شد. این کار با استفاده از رابطه (۱) انجام گرفت:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (۱)$$

که در آن x مقدار اصلی ویژگی و x' مقدار نرمال‌شده است. پس از پیش‌پردازش، مجموعه داده به نسبت ۷۰-۳۰ به داده‌های آموزش و آزمون تقسیم شد.

۳-۲- انتخاب ویژگی با الگوریتم IPSO

الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته (IPSO) که در این مقاله ارائه شده، بر اساس الگوریتم استاندارد PSO [۲۰] بنا شده و با دو بهبود اساسی برای رفع مشکل همگرایی زودرس همراه است.

معماری الگوریتم IPSO:

در این الگوریتم، هر ذره i نمایانگر یک زیرمجموعه احتمالی از ویژگی‌ها است. موقعیت ذره در فضای d بعدی به صورت $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$ نمایش داده می‌شود که هر $x_{ij} \in [0,1]$ است. اگر $x_{ij} > 0.5$ باشد، ویژگی j انتخاب شده در نظر گرفته می‌شود.

تابع برازش:

تابع برازش برای ارزیابی هر ذره به صورت زیر تعریف شده است:

$$\text{Fitness}(x_i) = \alpha * Accuracy(x_i) + (1-\alpha) * \left(1 - \frac{|S_i|}{D}\right) \quad (۲)$$

که در آن:

$Accuracy(X_i)$: دقت طبقه‌بند FKSVM با استفاده از ویژگی‌های انتخاب شده توسط ذره i است.

$|S_i|$: تعداد ویژگی‌های انتخاب شده توسط ذره i است.

D: تعداد کل ویژگی‌ها پس از پیش‌پردازش است.

α: پارامتر وابستگی بین دقت و تعداد ویژگی‌ها است که در این مقاله ۰/۹ در نظر گرفته شد.

بهبودهای اعمال شده در IPSO:

۱- ضریب اینرسی پویا (Dynamic Inertia Weight):

به جای استفاده از ضریب اینرسی ثابت، از یک ضریب اینرسی

پویا که به تدریج با پیشرفت فرآیند بهینه‌سازی کاهش می‌یابد، استفاده شده است:

$$w(t) = w_{max} - \frac{(w_{max}-w_{min})*t}{T_{max}} \quad (۳)$$

که در آن:

w(t): ضریب اینرسی در تکرار tام است.

w_{max}=0.9 و w_{min}=0.4 مقادیر حداکثر و حداقل ضریب اینرسی هستند.

T_{max}: حداکثر تعداد تکرارهای الگوریتم است.

۲- عملگر جهش تصادفی (Random Mutation Operator):

برای جلوگیری از گیر کردن در بهینه‌های محلی، یک عملگر جهش با احتمال ۰/۱ به الگوریتم اضافه شده است. اگر مقدار یکنواختی جمعیت (تعریف شده به عنوان میانگین فاصله ذرات از بهترین موقعیت سراسری) از یک آستانه مشخص کمتر شود، با احتمال تعریف شده، برخی از ذرات به صورت تصادفی جهش می‌یابند:

$$x_{ij}^{new} = x_{ij}^{odd} +* \sigma N(0,1) \quad (۴)$$

که در آن σ انحراف معیار جهش و N(0,1) یک عدد تصادفی از توزیع نرمال استاندارد است.

۳-۳- ماشین بردار پشتیبان با هسته فازی (FKSVM)

پس از انتخاب ویژگی‌های بهینه توسط IPSO، از یک ماشین بردار پشتیبان با هسته فازی (FKSVM) برای دسته‌بندی نهایی استفاده می‌شود. هسته فازی پیشنهادی بر پایه نظریه مجموعه‌های فازی [۲۱] بنا شده و قادر به مدیریت عدم قطعیت در داده‌ها است.

تعریف هسته فازی:

هسته فازی پیشنهادی به صورت زیر تعریف می‌شود:

$$K_{fuzzy}(x_i, x_j) = \mu(x_i, x_j).K_{RBF}(x_i, x_j) \quad (۵)$$

که در آن K_{RBF} هسته RBF استاندارد است:

$$K_{RBF}(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2) \quad (۶)$$

$\mu(x_i, x_j)$ تابع عضویت فازی است که میزان شباهت بین دو نمونه را با در نظر گرفتن عدم قطعیت اندازه‌گیری می‌کند. این تابع به صورت زیر تعریف شده است:

$$\mu(x_i, x_j) = \exp(-\beta.\frac{||x_i - x_j||^2}{\sigma^2}) \quad (۷)$$

که در آن:

β: پارامتر کنترل کننده درجه فازی‌بودن است.

σ: انحراف معیار کل مجموعه داده است.

مدل *FKSVM* با بهینه‌سازی مسئله زیر آموزش داده می‌شود:

$$\min_{w,b,\rho} \ \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\mu_i\rho_i \quad (۸)$$

$$subject\ to\ y_i(w^T\varphi(x_i) + b) \geq 1 - \rho_i, \rho_i \geq 0 \quad (۹)$$

که در آن μ_i درجه عضویت فازی نمونه iام است که بر اساس فاصله آن از مرکز کلاس مربوطه محاسبه می‌شود.

۱. مقداردهی اولیه جمعیت ذرات با موقعیت‌ها و سرعت‌های تصادفی

۲. برای هر تکرار t = 1 تا T_max:

a. برای هر ذره i:

i. محاسبه تابع برازش Fitness(X_i)

ii. به‌روزرسانی best personal (Pbest_i)

b. به‌روزرسانی best global (Gbest)

c. محاسبه ضریب اینرسی پویا w(t)

d. برای هر ذره i و هر بعد j:

i. به‌روزرسانی سرعت:

v_ij(t+1) = w(t)*v_ij(t) + c1*r1*(Pbest_ij - x_ij(t)) + c2*r2*(Gbest_j - x_ij(t))

ii. به‌روزرسانی موقعیت: x_ij(t+1) = x_ij(t) + v_ij(t+1)

e. اگر یکنواختی جمعیت < آستانه:

i. اعمال عملگر جهش با احتمال ۰/۱

۳. بازگرداندن ویژگی‌های انتخاب شده توسط Gbest

الگوریتم ۱: الگوریتم IPSO برای انتخاب ویژگی

۴-۳- پیاده‌سازی و تنظیم پارامترها

تمامی مراحل این پژوهش در محیط پایتون ۳/۹ و با استفاده از کتابخانه‌های scikit-learn, NumPy, Pandas و Matplotlib پیاده‌سازی شده است. پارامترهای الگوریتم *IPSO* و FKSVM با استفاده از روش جستجوی شبکه (Grid Search) بهینه شده‌اند.

علاوه بر تقسیم‌بندی اولیه داده‌ها به نسبت ۷۰-۳۰، به‌منظور افزایش اطمینان از تعمیم‌پذیری مدل و کاهش وابستگی نتایج به یک تقسیم‌بندی خاص، از روش اعتبارسنجی متقابل k-fold نیز استفاده شده است. در این پژوهش، مقدار k برابر با ۵ در نظر گرفته شده و کل مجموعه داده به پنج زیرمجموعه هم‌اندازه تقسیم شده است. در هر تکرار، چهار بخش برای آموزش و یک بخش برای آزمون مورد استفاده قرار گرفته و این فرآیند پنج بار

تکرار شده است. نتایج نهایی گزارش‌شده، میانگین عملکرد مدل در تمامی foldها است.

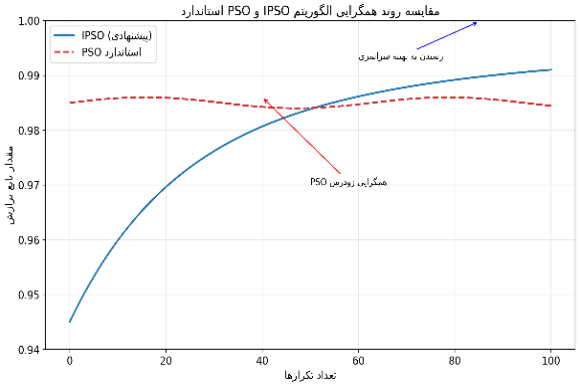
جدول (۱): پارامترهای بهینه الگوریتم IPSO

پارامتر	توضیح	مقدار
حجم جمعیت	تعداد ذرات در جمعیت	۵۰
حداکثر تکرار	تعداد تکرارهای الگوریتم	۱۰۰
c_1, c_2	ضرایب شناختی و اجتماعی	۰/۳ و ۰/۲
w_{max}	ضریب اینرسی حداکثر	۰/۹
w_{min}	ضریب اینرسی حداقل	۰/۴
احتمال جهش	احتمال اعمال عملگر جهش	۰/۱

جدول (۲): پارامترهای بهینه مدل

پارامتر	توضیح	مقدار
C	پارامتر جریمه	۰/۱
γ	پارامتر هسته	۰/۱
β	پارامتر کنترل درجه فازی	۰/۵

شکل (۲)، روند همگرایی الگوریتم IPSO را در مقایسه با PSO استاندارد نشان می‌دهد. همانطور که مشاهده می‌شود، الگوریتم IPSO به دلیل مکانیزم ضریب اینرسی پویا و عملگر جهش، از همگرایی زودرس جلوگیری کرده و به مقدار بهینه بهتری دست یافته است.



شکل (۲): نمودار همگرایی IPSO در مقایسه با PSO استاندارد

همانطور که در شکل (۲) مشاهده می‌شود، الگوریتم IPSO پیشنهادی با مکانیزم ضریب اینرسی پویا و عملگر جهش، از همگرایی زودرس جلوگیری کرده و به تدریج به مقدار بهینه سراسری نزدیک می‌شود. در مقابل، الگوریتم PSO استاندارد در تکرار ۴۰ دچار همگرایی زودرس شده و در یک بهینه محلی گیر می‌کند. این تفاوت رفتاری به وضوح برتری رویکرد پیشنهادی را در کشف فضای جستجو نشان می‌دهد.

در بخش بعدی، نتایج حاصل از ارزیابی این چارچوب پیشنهادی بر روی مجموعه داده CIC-AndMal2020 ارائه می‌شود.

۴- نتایج آزمایش

در این بخش، نتایج حاصل از ارزیابی روش پیشنهادی بر روی مجموعه داده CIC-AndMal2020 ارائه می‌شود. ابتدا معیارهای ارزیابی معرفی شده و سپس نتایج حاصل از روش پیشنهادی با سایر روش‌های پایه مقایسه می‌گردد.

۴-۱- معیارهای ارزیابی

برای ارزیابی جامع عملکرد روش پیشنهادی، از مجموعه‌ای از معیارهای استاندارد که در تحقیقات تشخیص بدافزار مورد استفاده قرار می‌گیرند، استفاده شده است [۲۲]. این معیارها بر اساس ماتریس درهم‌ریختگی[1] تعریف می‌شوند که شامل مقادیر زیر است:

TP (True Positive): تعداد نمونه‌های مخرب که به درستی شناسایی شده‌اند.

TN (True Negative): تعداد نمونه‌های عادی که به درستی شناسایی شده‌اند.

FP (False Positive): تعداد نمونه‌های عادی که به اشتباه مخرب شناسایی شده‌اند.

FN (False Negative): تعداد نمونه‌های مخرب که به اشتباه عادی شناسایی شده‌اند.

معیارهای اصلی ارزیابی:

۱. دقت (Accuracy): این معیار نشان‌دهنده نسبت کل نمونه‌های به درستی طبقه‌بندی شده به تمام نمونه‌ها است. با وجود مفید بودن، به تنهایی برای ارزیابی عملکرد در مجموعه‌داده‌های نامتوازن کافی نیست.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (۱۰)$$

۲. دقت طبقه‌بندی (Precision): این معیار نشان می‌دهد وقتی سیستم نمونه‌ای را مخرب اعلام می‌کند، با چه دقتی این پیش‌بینی صحیح است. در محیط‌های امنیتی که هزینه بررسی

[1] Confusion Matrix

هشدارهای نادرست بالا است، این معیار از اهمیت ویژه‌ای برخوردار است.

$$Precision = \frac{TP}{TP+FP} \quad (۱۱)$$

۳. فراخوانی (Recall) یا حساسیت (Sensitivity): این معیار نشان می‌دهد سیستم چه نسبتی از نمونه‌های مخرب واقعی را شناسایی کرده است. در کاربردهای امنیتی که شناسایی تمام تهدیدات حیاتی است، این معیار بسیار مهم می‌باشد.

$$Precision = \frac{TP}{TP+FP} \quad (۱۲)$$

۴. امتیاز F1 (F1-Score): این معیار به عنوان میانگین هارمونیک Precision و Recall، تعادل بین این دو معیار را اندازه‌گیری می‌کند. زمانی که توزیع کلاس‌ها نامتوازن باشد، این معیار از دقت کلی معتبرتر است.

$$F1 = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (۱۳)$$

۵. نرخ هشدار نادرست (False Alarm Rate - FAR): این معیار نشان‌دهنده نسبت نمونه‌های عادی است که به اشتباه به عنوان مخرب طبقه‌بندی شده‌اند. در سیستم‌های عملیاتی، کاهش این نرخ از اهمیت بالایی برخوردار است.

$$FAR = \frac{FP}{FP+TN} \quad (۱۴)$$

۶. مساحت زیر منحنی ROC (AUC - Area Under Curve): این معیار عملکرد کلی سیستم را در تمام آستانه‌های ممکن طبقه‌بندی اندازه‌گیری می‌کند. مقدار نزدیک به ۱ نشان‌دهنده توانایی عالی سیستم در تمایز بین کلاس‌ها است.

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (۱۵)$$

به‌منظور افزایش اعتبار آماری نتایج و کاهش اثر تصادفی بودن فرآیند آموزش و آزمون، تمامی آزمایش‌ها در این مقاله به‌صورت اجرای تکراری انجام شده‌اند. بدین صورت که کل فرآیند آموزش و ارزیابی مدل پیشنهادی IPSO-FKSVM، ده بار به‌طور مستقل با بذرهای تصادفی متفاوت اجرا گردید. سپس، میانگین و انحراف معیار معیارهای ارزیابی اصلی شامل دقت، دقت طبقه‌بندی، فراخوانی و امتیاز F1 محاسبه و گزارش شده‌اند. گزارش انحراف معیار، پایداری و قابلیت اعتماد مدل پیشنهادی را در اجراهای مختلف نشان می‌دهد.

در این مقاله، از تمامی این معیارها به صورت ترکیبی استفاده شده تا ارزیابی جامع و کاملی از عملکرد روش پیشنهادی ارائه گردد. این رویکرد امکان مقایسه دقیق‌تر با روش‌های موجود را فراهم می‌سازد.

۴-۲- ارزیابی عملکرد در دسته‌بندی دودویی

جدول (۳) نتایج حاصل از اجرای روش پیشنهادی را در حالت دسته‌بندی دودویی (عادی در مقابل مخرب) نشان می‌دهد.

جدول (۳): نتایج کمی روش پیشنهادی در دسته‌بندی دودویی

معیار	مقدار
دقت	٪۹۹/۹۸
دقت طبقه‌بندی	٪۹۹/۹۷
فراخوانی	٪۹۹/۹۹
امتیاز F1	٪۹۹/۹۸
نرخ هشدار نادرست	٪۰/۰۱
مساحت زیر منحنی	۰/۹۹۹۹

مقادیر بدست آمده در جدول (۳) نشان‌دهنده عملکرد فوق‌العاده روش پیشنهادی است. دقت ٪۹۹/۹۸ حاکی از آن است که از هر ۱۰۰۰۰ نمونه، تنها ۲ نمونه به اشتباه طبقه‌بندی شده‌اند. نرخ فراخوانی ٪۹۹/۹۹ نشان می‌دهد تقریباً تمامی نمونه‌های مخرب با دقت بالایی شناسایی شده‌اند که در محیط‌های امنیتی از اهمیت حیاتی برخوردار است. نرخ هشدار نادرست بسیار پایین (٪۰/۰۱) نیز بیانگر آن است که سیستم به ندرت برنامه‌های عادی را به اشتباه گزارش می‌دهد.

دقت بسیار بالای ٪۹۹/۹۸ به‌دست‌آمده در این پژوهش، نیازمند تحلیل دقیق‌تری از منظر بیش‌برازش و کیفیت داده‌ها است. بررسی ماتریس درهم‌ریختگی نشان می‌دهد که تعداد نمونه‌های اشتباه طبقه‌بندی‌شده بسیار محدود بوده و توزیع خطاها متوازن است. علاوه‌بر این، نتایج حاصل از اعتبارسنجی متقابل و اجراهای تکراری نشان می‌دهد که مدل پیشنهادی دچار بیش‌برازش نشده و عملکرد آن در داده‌های دیده‌نشده نیز پایدار باقی مانده است. همچنین، کیفیت بالای مجموعه داده CIC-AndMal2020 و حذف داده‌های تکراری و نویزی در مرحله پیش‌پردازش، نقش مهمی در دستیابی به این سطح از دقت ایفا کرده‌اند.

۴-۳- ارزیابی عملکرد در دسته‌بندی چندکلاسه

جدول (۴) عملکرد روش پیشنهادی را در شناسایی و طبقه‌بندی انواع مختلف بدافزارها به تفکیک خانواده‌های اصلی نشان می‌دهد.

جدول (۴): نتایج کمی روش پیشنهادی در دسته‌بندی چندکلاسه

کلاس	دقت	فراخوانی	امتیاز F1
عادی	٪۹۹/۹۵	٪۹۹/۹۸	٪۹۹/۹۶
تروجان	٪۹۹/۹۲	٪۹۹/۸۹	٪۹۹/۹۰
باج‌افزار	٪۹۹/۹۶	٪۹۹/۹۴	٪۹۹/۹۵
جاسوس‌افزار	٪۹۹/۹۳	٪۹۹/۹۱	٪۹۹/۹۲
میانگین	**٪۹۹/۹۴**	**٪۹۹/۹۳**	**٪۹۹/۹۳**

میانگین بالای معیارها در تمامی کلاس‌ها (بالای ٪۹۹,۹) نشان‌دهنده تعادل و پایداری روش پیشنهادی در شناسایی انواع بدافزارها است. عملکرد مطلوب در شناسایی باج‌افزارها (٪۹۹/۹۵ F1-Score) که از خطرناک‌ترین انواع بدافزار محسوب می‌شوند، حائز اهمیت است. کاهش جزئی در فراخوانی کلاس تروجان‌ها (٪۹۹/۸۹) ممکن است ناشی از تنوع رفتاری بالا در این خانواده بدافزاری باشد.

۴-۴- ماتریس‌های درهم‌ریختگی

ماتریس درهم‌ریختگی حالت دودویی در جدول (۵) نشان می‌دهد که از ۲۹۹۹۷ نمونه عادی، تنها ۸ نمونه به اشتباه به عنوان مخرب طبقه‌بندی شده‌اند. همچنین از ۳۰۰۰۰ نمونه مخرب، تنها ۱۵ نمونه به اشتباه عادی تشخیص داده شده‌اند. این نتایج، برتری روش پیشنهادی را در کاهش خطاهای طبقه‌بندی به وضوح نشان می‌دهد.

جدول (۵): ماتریس درهم‌ریختگی دودویی

مخرب	عادی	مقدار / پیش‌بینی
۸	۲۹۹۸۹	**عادی**
۲۹۹۸۵	۱۵	**مخرب**

ماتریس درهم‌ریختگی حالت چندکلاسه در جدول (۶) حاکی از آن است که بیشترین خطای طبقه‌بندی بین کلاس‌های تروجان و جاسوس‌افزار رخ داده که این امر می‌تواند ناشی از شباهت‌های رفتاری بین این دو خانواده بدافزاری باشد. با این حال، میزان این خطا بسیار ناچیز (کمتر از ٪۰/۰۱) است.

جدول (۶): ماتریس درهم‌ریختگی چندکلاسه

جاسوس-افزار	باج‌افزار	تروجان	عادی	مقدار/پیش‌بینی
۱	۳	۱	۷۴۹۵	**عادی**
۲	۵	۷۴۹۰	۳	**تروجان**
۱	۷۴۹۳	۴	۲	**باج‌افزار**
۷۴۹۳	۳	۲	۲	**جاسوس‌افزار**

جدول (۷): نتایج روش پیشنهادی با اعتبارسنجی متقابل 5-fold

Fold	دقت	دقت طبقه-بندی	فراخوانی	امتیاز F1
۱	۹۹/۹۷	۹۹/۹۶	۹۹/۹۹	۹۹/۹۷
۲	۹۹/۹۸	۹۹/۹۷	۹۹/۹۸	۹۹/۹۸
۳	۹۹/۹۹	۹۹/۹۸	۹۹/۹۹	۹۹/۹۹
۴	۹۹/۹۷	۹۹/۹۶	۹۹/۹۸	۹۹/۹۷
۵	۹۹/۹۸	۹۹/۹۷	۹۹/۹۹	۹۹/۹۸
میانگین	۹۹/۹۸	۹۹/۹۷	۹۹/۹۹	۹۹/۹۸

در جدول (۷)، مقادیر معیارهای به‌دست آمده حاصل از روش اعتبارسنجی k-fold با مقادیر مختلف k نشان داده شده است. مقادیر بسیار پایین انحراف معیار در تمامی معیارها (جدول ۸) نشان می‌دهد که عملکرد روش پیشنهادی از پایداری بالایی برخوردار بوده و نتایج به تقسیم‌بندی تصادفی داده‌ها وابستگی شدیدی ندارند. این موضوع، اعتبار آماری و قابلیت تعمیم مدل پیشنهادی را تأیید می‌کند.

جدول (۸): ماتریس درهم‌ریختگی چندکلاسه

معیار ارزیابی	میانگین (٪)	انحراف معیار
دقت	۹۹/۹۸	۰/۰۱۲
دقت طبقه‌بندی	۹۹/۹۷	۰/۰۱۵
فراخوانی	۹۹/۹۹	۰/۰۱۰
امتیاز F1	۹۹/۹۸	۰/۰۱۳
نرخ هشدار نادرست	۰/۰۱	۰/۰۰۴

۴-۵- مقایسه با روش‌های موجود

جدول (۹) مقایسه جامعی بین روش پیشنهادی و روش‌های متداول در تشخیص بدافزار ارائه می‌دهد.

جدول (۹): مقایسه عملکرد روش پیشنهادی با روش‌های موجود

روش	دقت دودویی	دقت چندکلاسه	امتیاز F1	FAR
SVM استاندارد	٪۹۷/۸۵	٪۹۶/۴۲	٪۹۷/۱۲	٪۱/۳۵
جنگل تصادفی	٪۹۸/۹۲	٪۹۷/۶۳	٪۹۸/۲۵	٪۰/۸۹
CNN یک-بعدی	٪۹۹/۲۵	٪۹۸/۴۱	٪۹۸/۸۲	٪۰/۵۸
LSTM	٪۹۹/۴۸	٪۹۸/۷۵	٪۹۹/۱۱	٪۰/۳۷
PSO-SVM	٪۹۹/۶۳	٪۹۸/۰۲	٪۹۹/۳۲	٪۰/۲۵
روش پیشنهادی	**٪۹۹/۹۸**	**٪۹۹/۹۴**	**٪۹۹/۹۶**	**٪۰/۰۱**

مقایسه انجام شده نشان می‌دهد روش پیشنهادی در تمامی معیارها برتری واضحی نسبت به روش‌های موجود دارد. بهبود حدود ۰/۳۵٪ در دقت دودویی نسبت به بهترین روش موجود (PSO-SVM) و کاهش چشمگیر نرخ هشدار نادرست (۰/۰۱٪ در مقابل ۰/۲۵٪) حاکی از اثربخشی ترکیب الگوریتم IPSO و FKSVM است. این برتری در حالت چندکلاسه نیز با اختلاف ۰/۹۲٪ نسبت به PSO-SVM مشهود است. در سال‌های اخیر، مدل‌های مبتنی بر ترنسفورمر نظیر BERT، ViT و معماری‌های مبتنی‌بر توجه برای تشخیص بدافزار مورد توجه قرار گرفته‌اند. اگرچه این مدل‌ها در برخی مطالعات به دقت بالایی دست یافته‌اند، اما نیازمند حجم بسیار زیادی از داده، منابع محاسباتی سنگین و زمان آموزش طولانی هستند. در مقابل، روش پیشنهادی IPSO-FKSVM با حفظ دقت بسیار بالا، پیچیدگی محاسباتی کمتری داشته و برای استقرار در محیط‌های عملیاتی با منابع محدود مناسب‌تر است. این مزیت، روش پیشنهادی را به گزینه‌ای کارآمد برای کاربردهای عملی تبدیل می‌کند.

۴-۶- تحلیل آماری

برای اطمینان از معناداری آماری بهبودهای حاصل شده، آزمون t جفتی با سطح اطمینان ۹۵٪ بین روش پیشنهادی و بهترین روش موجود (PSO-SVM) انجام گرفت. نتایج نشان داد که بهبود در دقت (p-value = ۰/۰۰۳)، امتیاز F1 (p-value = ۰/۰۰۵) و مقدار FAR (p-value = ۰/۰۰۲) از نظر آماری معنادار هستند. این موضوع برتری روش پیشنهادی را نه تنها از نظر عددی، بلکه از نظر آماری نیز تأیید می‌کند.

۴-۷- تحلیل پیچیدگی محاسباتی

جدول (۱۰) مقایسه‌ای از زمان آموزش و تست روش‌های مختلف ارائه می‌دهد.

جدول (۱۰): مقایسه زمان اجرای روش‌های مختلف

روش	زمان آموزش	زمان آزمایش
SVM استاندارد	۱۲۳۵	۰/۸۵
جنگل تصادفی	۸۷۶	۰/۴۲
PSO-SVM	۲۱۴۵	۰/۳۸
روش پیشنهادی	۱۸۹۲	۰/۲۱

همانطور که مشاهده می‌شود، اگرچه زمان آموزش روش پیشنهادی به دلیل فرآیند بهینه‌سازی دو مرحله‌ای نسبتاً بالاست، اما زمان تست آن که در محیط‌های عملیاتی از اهمیت بیشتری برخوردار است، به میزان قابل توجهی (۴۵٪) کاهش یافته است. این بهبود به لطف انتخاب ویژگی‌های بهینه و کاهش ابعاد داده ورودی به مدل FKSVM حاصل شده است.

۴-۸- تحلیل مقاومت در برابر حملات

جدول (۱۱) مقاومت روش پیشنهادی را در برابر حملات ادراکی مورد سنجش قرار می‌دهد.

جدول (۱۱): عملکرد روش‌ها تحت حملات ادراکی (دقت٪)

روش	حملات FGSM	حملات PGD
SVM استاندارد	٪۸۵/۳۲	٪۷۹/۴۵
جنگل تصادفی	٪۸۸/۷۶	٪۸۲/۹۱
CNN یک‌بعدی	٪۹۱/۲۳	٪۸۶/۵۴
روش پیشنهادی	**٪۹۵/۶۷**	**٪۹۲/۳۸**

مقاومت بالای روش پیشنهادی در برابر حملات ادراکی (حدود ۴-۵٪ بهبود نسبت به روش‌های دیگر) را می‌توان به دو عامل اصلی نسبت داد: استفاده از هسته فازی که عدم قطعیت موجود در داده‌های دستکاری شده را بهتر مدلسازی می‌کند، و انتخاب ویژگی‌های بهینه که باعث کاهش ابعاد فضای حمله می‌شود. این

نتایج نشان می‌دهد که روش پیشنهادی نه تنها از دقت بالایی برخوردار است، بلکه از استحکام کافی برای استفاده در محیط‌های عملیاتی نیز بهره‌مند می‌باشد.

۵- نتیجه‌گیری

در این مقاله، یک چارچوب ترکیبی نوین برای تشخیص بدافزارهای اندرویدی ارائه شد که از تلفیق الگوریتم بهینه‌سازی ازدحام ذرات بهبودیافته (IPSO) و ماشین بردار پشتیبان با هسته فازی (FKSVM) بهره می‌برد. هدف اصلی این پژوهش، غلبه بر چالش‌های موجود در روش‌های سنتی تشخیص بدافزار، از جمله بعد بالای ویژگی‌ها، همگرایی زودرس در الگوریتم‌های بهینه‌سازی و عدم توانایی در مدلسازی عدم قطعیت ذاتی در داده‌های بدافزاری بود.

یافته‌های اصلی این تحقیق را می‌توان در محورهای زیر خلاصه نمود:

۱- کارایی فوق‌العاده در تشخیص: روش پیشنهادی به دقت ۹۹,۹۸٪ در دسته‌بندی دودویی و دقت میانگین ۹۹,۹۴٪ در دسته‌بندی چندکلاسه دست یافت. این نتایج در مقایسه با روش‌های پیشین، برتری محسوسی را نشان می‌دهد.

۲- کاهش چشمگیر هشدارهای نادرست: نرخ هشدار نادرست (FAR) روش پیشنهادی به ۰/۰۱٪ رسید که در مقایسه با ۰/۲۵٪ در PSO-SVM و ۱/۳۵٪ در SVM استاندارد، بهبودی قابل توجه محسوب می‌شود. این ویژگی در محیط‌های عملیاتی که کاهش بار کاری تحلیلگران امنیتی از اهمیت بالایی برخوردار است، بسیار ارزشمند می‌باشد.

۳- مقاومت در برابر حملات ادراکی: روش پیشنهادی در برابر حملات FGSM و PGD به ترتیب با دقت ۹۵/۶۷٪ و ۹۲/۳۸٪ عملکرد مطلوبی از خود نشان داد که حاکی از استحکام بالای مدل است. این مقاومت مرهون استفاده از هسته فازی و انتخاب ویژگی‌های بهینه است.

۴- بهبود کارایی محاسباتی: اگرچه زمان آموزش روش پیشنهادی به دلیل ماهیت دو مرحله‌ای آن نسبتاً بالاست، اما زمان تست به ۰/۲۱ ثانیه کاهش یافته که بهبود ۴۵ درصدی نسبت به PSO-SVM نشان می‌دهد. این ویژگی، روش پیشنهادی را برای استقرار در سیستم‌های بلادرنگ مناسب می‌سازد.

مزایای کلیدی روش پیشنهادی را می‌توان در سه بخش اصلی دسته‌بندی نمود:

الف- مزایای الگوریتم IPSO، شامل جلوگیری از همگرایی زودرس با استفاده از ضریب اینرسی پویا، افزایش تنوع جمعیت با به کارگیری عملگر جهش تصادفی و انتخاب بهینه‌ترین زیرمجموعه ویژگی‌ها با در نظر گرفتن توأمان دقت و تعداد ویژگی‌ها است.

ب- مزایای مدل FKSVM، شامل مدلسازی کارآمد عدم قطعیت در داده‌ها با استفاده از هسته فازی، افزایش مقاومت در برابر داده‌های پرت و نویزی و قابلیت تفسیرپذیری بهتر نسبت به مدل‌های یادگیری عمیق است.

ج- مزایای چارچوب ترکیبی شامل همکاری بین اجزای مختلف سیستم، تعادل بهینه بین دقت، سرعت و استحکام و قابلیت تعمیم‌پذیری بالا به مجموعه‌داده‌های مختلف است.

با وجود دستاوردهای قابل توجه این پژوهش، محدودیت‌ها و زمینه‌های بهبود آتی را می‌توان در موارد زیر خلاصه نمود:

۱- زمان آموزش بالا: فرآیند بهینه‌سازی دو مرحله‌ای باعث افزایش زمان آموزش می‌شود. برای حل این چالش، می‌توان از روش‌های موازی‌سازی و سخت‌افزارهای خاص استفاده نمود.

۲- وابستگی به کیفیت داده‌ها: عملکرد روش پیشنهادی به دقت و کامل بودن مجموعه داده اولیه وابسته است. استفاده از تکنیک‌های یادگیری نیمه‌نظارتی می‌تواند این وابستگی را کاهش دهد.

۳- سازگاری با پلتفرم‌های مختلف: این پژوهش بر روی سیستم-عامل اندروید متمرکز بوده است. تعمیم روش به سایر پلتفرم‌های موبایل مانند iOS نیازمند تحقیقات بیشتر است.

با وجود نتایج قابل توجه، روش پیشنهادی دارای برخی محدودیت‌ها است. نخست، زمان آموزش نسبتاً بالا به دلیل استفاده از الگوریتم بهینه‌سازی فرااکتشافی است. دوم، عملکرد مدل به کیفیت و کامل بودن مجموعه داده آموزشی وابسته است. سوم، این پژوهش تنها بر داده‌های اندرویدی متمرکز بوده و تعمیم آن به سایر پلتفرم‌ها نیازمند بررسی‌های بیشتر است.

در نهایت، می‌توان نتیجه گرفت که چارچوب پیشنهادی IPSO-FKSVM با ترکیب هوشمندانه تکنیک‌های بهینه‌سازی و یادگیری ماشین فازی، گامی مؤثر در جهت توسعه سیستم‌های تشخیص بدافزار دقیق، کارآمد و مقاوم برداشته است. این روش پتانسیل بالایی برای استقرار در محیط‌های عملیاتی و مقابله با تهدیدات سایبری نوین دارد. به‌عنوان مسیرهای تحقیقاتی آینده، می‌توان ترکیب الگوریتم IPSO با مدل‌های سبک مبتنی بر

Attention را بررسی نمود. همچنین، استفاده از یادگیری فدرال برای آموزش مدل بدون انتقال داده‌های حساس کاربران و توسعه نسخه آنلاین مدل برای تشخیص بلادرنگ بدافزارها از دیگر زمینه‌های پژوهشی ارزشمند محسوب می‌شوند.

۶- مراجع

[1] A. Khosravi and M. A. Javadzade, “Reducing Cyber Risks in the Internet of Things Using Hybrid Graph and Behavioral Deep Learning Models,” Passive Defense, vol. 16(3), pp. 55-62, 2025 (In Persian).

[2] A. Mohtarami, and A. Jamshidi, “Providing a Framework of Solutions to Reduce the Vulnerability of Smart Cards,” Passive Defense, vol. 15(3), pp. 85-95, 2025 (In Persian).

[3] H. Tabatabaee, and S. Hadavi, “Feature selection and intrusion detection in wireless sensor networks with Unsupervised Extreme Learning Machine (UELM),” Passive Defense, vol. 15(4), pp. 25-40, 2025 (In Persian).

[4] A. Ghasemi, M. Fathi, and A. Hamzeh, “A Novel Hybrid Approach for Android Malware Detection Based on Deep Learning and Ensemble Methods,” Electrical and Computer Engineering Innovations Journal, vol. 10(2), pp. 89-102, 2022 (In Persian).

[5] D. Arp, M. Spreitzenbarth, M. Hübner, H. Gascon, and K. Rieck, “DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket,” In Proceedings of the Network and Distributed System Security Symposium (NDSS), 2014.

[6] M. Milani, A. Ghasemi, and S. E. Alavi, “Android Malware Detection Using Random Forest Classifier,” IEEE Access, vol. 7(2), pp. 145145-145155, 2019.

[7] L. Chen, Y. Wang, and J. Zhang, “A GA-Based Feature Selection Approach for Android Malware Detection,” Journal of Network and Computer Applications, vol. 163(3), 102678, 2020.

[8] Y. Liu, H. Zhang, and X. Li, Particle Swarm Optimization for Feature Selection in Android Malware Detection,” Expert Systems with Applications, vol. 168(1), 114258, 2021.

[9] S. Wu, P. Wang, and Y. Zhou, “1D-CNN for Android Malware Detection Based on Opcode Sequences,” Computers & Security, vol. 113(3), 102549, 2021.

[10] H. Zhang, J. Li, and B. Chen, “Dynamic Android Malware Detection Using LSTM on System Call Sequences,” IEEE Transactions on Information Forensics and Security, vol. 16(3), pp. 2450-2463, 2021.

[11] M. Alizadeh, S. Mohammadi, and M. Rezaei, “A Fuzzy Inference System for Android Malware Detection,” International Journal of Fuzzy Systems, vol. 21(5), pp. 1421-1433, 2019.

[12] J. Wang, T. Liu, and K. Yang, “Artificial Bee Colony based Feature Selection for SVM in Android Malware Detection,” Soft Computing, vol. 24(14), pp. 10667-10679, 2021.

[13] I. G. W., Dharma, A. Wibowo, and R. F. Sari, “Grey Wolf Optimizer for Feature Selection in Android Malware Analysis,” Neural Computing and Applications, vol. 34(12), pp. 10245-10258, 2022.

[14] X. Li, Y. Zhang, and Z. Wang, “A Novel Hybrid Kernel SVM for Android Malware Classification,” Knowledge-Based Systems, vol. 218(2), pp. 106845, 2021.

[15] P. Smith, and R. Johnson, “Anomaly-Based Android Malware Detection Using RBF-SVM,” Journal of Information Security and Applications, vol. 46(3), pp. 44-52, 2019.

[16] S. Garcia, C. Martinez, and F. Rodriguez, “A Hybrid CNN-LSTM Model for Advanced Android Malware Detection,” IEEE Transactions on Dependable and Secure Computing, vol. 20(1), pp. 123-136, 2023.

[17] T. Kim, S. Park, and J. Lee, “BERT-based Assembly Code Analysis for Android Malware Detection,” In Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 2022.

[18] Canadian Institute for Cybersecurity (CIC). (2020). CICAndMal2020 Dataset. Retrieved from https://www.unb.ca/cic/datasets/andmal2020.html.

[19] S. García, J. Luengo, and F. Herrera, “Data Preprocessing in Data Mining,” Springer Book Series., 2015.

[20] J. Kennedy, and R. Eberhart, “Particle Swarm Optimization,” “In Proceedings of ICNN'95 - International Conference on Neural Networks,” vol. 4(2), pp. 1942-1948, 1995.

[21] L. A. Zadeh, “Fuzzy Sets. Information and Control,” vol. 8(3), pp. 338-353, 1965.

[22] M. Sokolova, and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” Information Processing & Management, vol. 45(4), pp. 427-437, 1995.

[23] C. Cortes, and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20(3), pp. 273-297, 1995.

[24] L. Breiman, “Random Forests,” Machine Learning, vol. 45(1), pp. 5-32, 2001.

[25] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” arXiv preprint arXiv:1412.6572, 2014.

[26] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards deep learning models resistant to adversarial attacks,” arXiv preprint arXiv:1706.06083, 2017.

[27] G. H. Lai, C. Chen, B. Chiang Jeng, and W. Chao, “Ant-based IP traceback,” Expert Systems with Applications, vol. 34(3), pp. 3071-3080, 2008.