Software Impacts 21 (2024) 100664

Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts

Original software publication

IDS-Anta: An open-source code with a defense mechanism to detect adversarial attacks for intrusion detection system

Kousik Barik [a], Sanjay Misra [b,c,*]

[a] Department of Computer Science, University of Alcala, Madrid, Spain
[b] Department of Computer Science and Communication, Østfold University College, Halden, Norway
[c] Department of Applied Data Science, Institute for Energy Technology, Halden, Norway

ARTICLE INFO

Keywords:
Adversarial attack
Intrusion detection system
Cybersecurity
Adversarial machine learning
Adversarial defense

ABSTRACT

An intrusion detection system (IDS) is critical in protecting organizations from cyber threats. The susceptibility of Machine Learning and Deep Learning-based IDSs against adversarial attacks arises from malicious actors' deliberate construction of adversarial samples. This study proposes a Python-based open-source code repository named IDS-Anta with a robust defense mechanism to identify adversarial attacks without compromising IDS performance. It uses Multi-Armed Bandits with Thomson Sampling, Ant Colony Optimization (ACO), and adversarial attack generation methods and is validated using three public benchmark datasets. This code repository can be readily applied and replicated on IDS datasets against adversarial attacks.

Code metadata

Current code version	*1.0*
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2024-102
Permanent link to reproducible capsule	
Legal code license	*MIT License*
Code versioning system used	*git*
Software code languages, tools and services used	*Python, Jupyter Notebook*
Compilation requirements, operating environments and dependencies	*Python, Scikit-learn, ZOO, FGSM, ART*
If available, link to developer documentation/manual	https://github.com/kousikbarik/lab-ids-anta/blob/main/README.md
Support email for questions	sanjay.misra@ife.no

1. Introduction to IDS and adversarial attacks

Due to the exponential growth and complexity of communication networks, adequate security control for protecting organizations is paramount. Intrusion Detection Systems (IDS) aim to monitor network activity and identify potential security vulnerabilities. However, the growing sophistication of attack types and the variety of network traffic have posed significant challenges in accurately categorizing through conventional rule-based IDS methods [1]. There are three broad categories of IDS: signature-based, anomaly-based, and hybrid [2]. Signature-based IDS detects established attacks by analyzing predetermined patterns or signatures within the system. However, these systems are unable to identify new attacks [3]. Anomaly-based IDS detects novel attacks by distinguishing between unfamiliar attacks and pre-existing normal activities [4]. Hybrid IDS are created by combining two methods.

Machine Learning (ML) and Deep Learning (DL) methods have been widely employed in Intrusion Detection Systems (IDS) to overcome those constraints and categorize network communication [5]. The techniques above utilize statistical measures and algorithms to dynamically acquire knowledge and detect abnormal network traffic patterns that indicate potential security vulnerabilities [6]. The primary objective of the IDS research based on the ML and DL techniques has been to improve the assessment of the model in terms of different evaluation parameters and minimize false positive and false negative issues [7,8]. Nevertheless, the model's resilience must be addressed, as most ML and DL techniques are susceptible to adversarial attacks [9,10]. Adversarial examples are typically constructed by introducing small amounts of noise into the original input during the training or testing stage. This

* Correspondence to: Institute for Energy Technology, Halden, Norway.
E-mail addresses: kousik.kousik@edu.uah.es (K. Barik), sanjay.misra@ife.no (S. Misra).

https://doi.org/10.1016/j.simpa.2024.100664
Received 22 April 2024; Received in revised form 10 May 2024; Accepted 12 May 2024

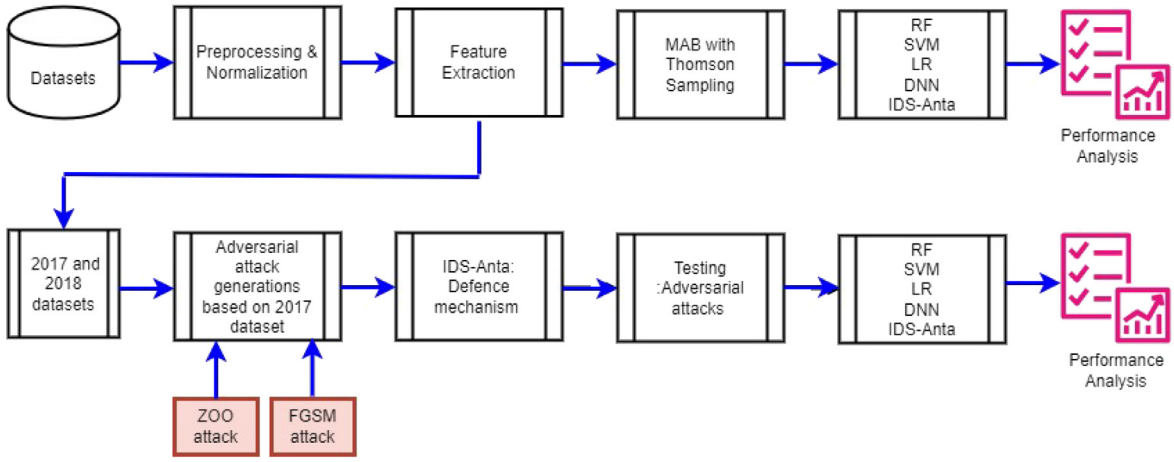

Fig. 1. A comprehensive outline of the IDS-Anta.

manipulation by the attacker leads to the model incorrectly predicting the outcome [11].

With the increasing prevalence of ML and DL-based IDS systems, adversarial attacks are becoming more important [12]. Hence, designing efficient protection methods to mitigate the adverse consequences and ensure trustworthiness and stability against adversarial attack scenarios is imperative. This proposed work analyzes the exposure of ML and DL-based IDS against two adversarial attack generation methods, i.e., Zeroth Order Optimization (ZOO) [13] and Fast Gradient Sign Attack (FGSM) [14]. The study proposes a novel, resilient defense mechanism named IDS-Anta to counter adversarial attacks on IDs. IDS-Anta uses preprocessing and feature extraction, employing z-score normalization and Singular Value Decomposition (SVD) [15]. It uses various techniques, i.e., Multi-Armed Bandits (MAB) [16] with Thompson sampling [17] and Ant Colony Optimization (ACO) [18] to dynamically choose the most effective classifiers or a combination of real-time classifiers for every input to identify an intrusion. This approach allows it to accomplish this goal while maintaining enactment on conventional traffic. This study used three ML classifiers, i.e., Random Forest (RF) [19], Support Vector Machine (SVM) [20], Logistic Regression (LR) [21], and one DL classifier, i.e., Deep Neural Network (DNN) [22].

2. IDS-Anta code function and adversarial attacks

IDS-Anta software repository empowers scientists to develop resilient defense strategies to counter adversarial attacks on machine learning and deep learning intrusion detection systems (IDS). IDS-Anta addresses the research questions listed below.

- What is the general procedure for ML- and DL-based IDS design, including the significance of preprocessing and feature extraction?
- How can MAB with Thomson Sampling dynamically choose an effective classifier, balance, and enhance the attack detection rate?
- How can ZOO and FGSM be used to generate adversarial samples?
- How can IDS performance be enhanced in detecting adversarial attacks with combined techniques (e.g., the proposed IDS-Anta)?

Fig. 1 depicts a high-level overview of the proposed IDS-Anta. The proposed code collection includes code implementations for detecting adversarial attacks in IDS: ML and DL-based IDS, MAB, adversarial attack generations using ZOO and FSGM, and IDS-Anta. The software is developed in Python, and different packages are used: Scikit-learn, NumPy, ZOO attack, and FGSM attack. This study uses three publicly available benchmark datasets, i.e., CIC-IDS-2017 [23], CEC-CIC-IDS-2018 [24], and CIC-DDoS-2019 [25]. These datasets possess different IDS attacks, including brute-force, XSS, SQL injection, port scan, botnets, Denial of Service, and infiltration attacks. The architecture of the IDS-Anta is illustrated in Fig. 2. The code repository contains the following specific code files:

1. Evaluation-2017.ipynb: The CIC-IDS-2017 dataset is processed and normalized using z-score. Features are extracted by utilizing the Singular Value Decomposition (SVC) and the MAB algorithm, which is employed to select the most optimal classifiers dynamically. Thomson Sampling balances and improves the overall attack detection rate. The proposed IDS-Anta combines four classifiers: RF, SVM, LR, and DNN. The classifiers are trained with the three selected datasets and the MAB algorithm with Thomson Sampling. The performance of the four selected classifiers and IDS-Anta is evaluated using the CIC-IDS-2017 dataset.
2. Evaluation-2018.ipynb: The CEC-CIC-IDS-2018 is considered in this scenario, and other procedures are followed, as stated previously. The performance of the four selected classifiers and IDS-Anta is evaluated using the CEC-CIC-IDS-2018 dataset.
3. Evaluation-2019.ipynb: The CIC-DDoS-2019 is considered in this scenario, and other procedures are followed, as stated previously. The performance of the four selected classifiers and IDS-Anta is evaluated employing the CIC-DDoS-2019 dataset.
4. ZOO-adversarial.ipynb: The code is implemented for the proposed IDS-Anta in an adversarial attack scenario. Based on the four chosen classifiers, the open-source enactment furnished by Adversarial Robustness Toolbox (ART) [26] and the classifier class have been created and used in both adversarial attack samples in IDS-Anta. The RF, SVM, LR, DNN, and IDs-Anta are trained with the CIC-IDS2017 and CEC-CIC-IDS-2018 datasets, the Multi-Armed Bandits algorithm, and Thomson sampling. The IDS-Anta is further optimized with Ant Colony Optimization. The proposed model is tested with adversarial samples generated using the ZOO method and CIC-IDS2017 dataset. The proposed IDS-Anta achieved better accuracy and other evaluation parameters than other classifiers. This symbolizes that the ML and DL-based IDS are susceptible to adversarial attacks, and the proposed IDS-Anta, with a robust defense mechanism, effectively reduces the impact of adversarial attacks.
5. FGSM-adversarial.ipynb: The code is accomplished for the proposed IDS-Anta in an adversarial attack scenario using the FGSM method. The same techniques are used as mentioned in ZOO-adversarial.ipynb. The outcomes indicate that the proposed IDS-Anta with defense method achieved more promising results than the selected classifiers in an adversarial attack scenario utilizing the FGSM method.

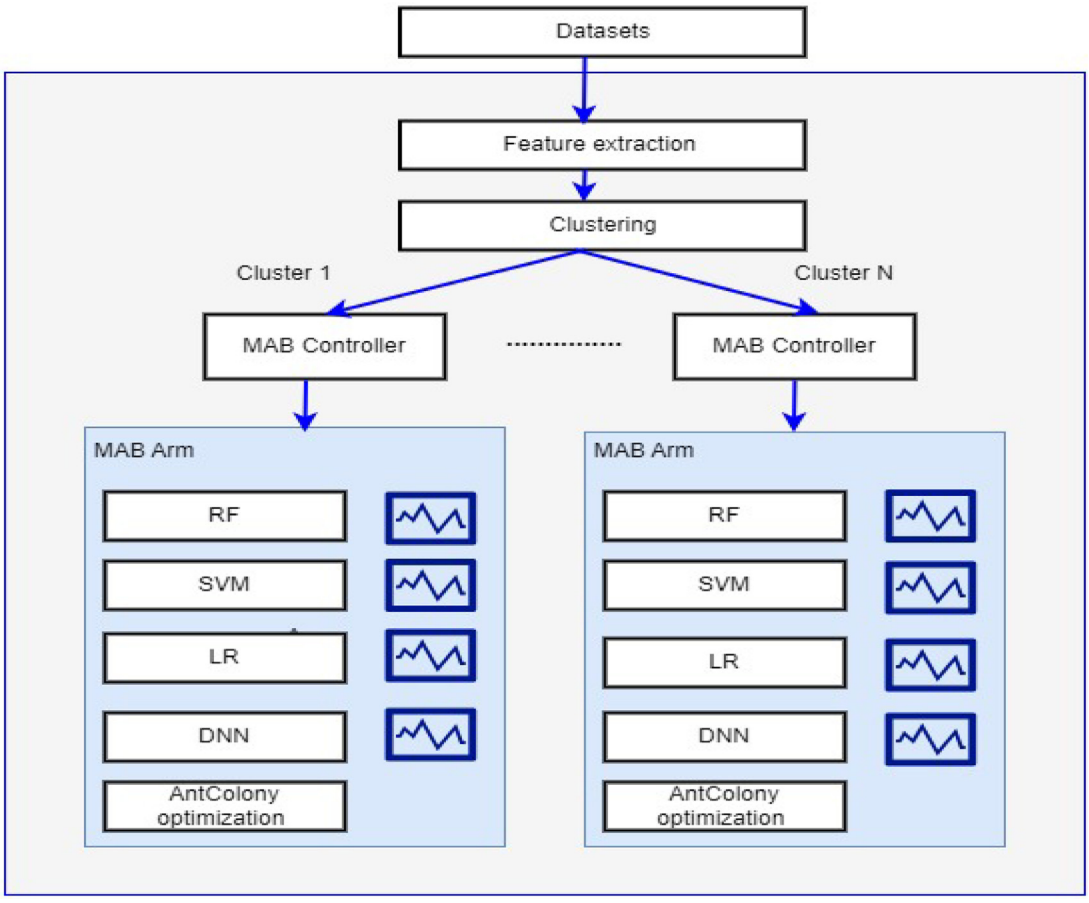

Fig. 2. IDS-Anta architecture.

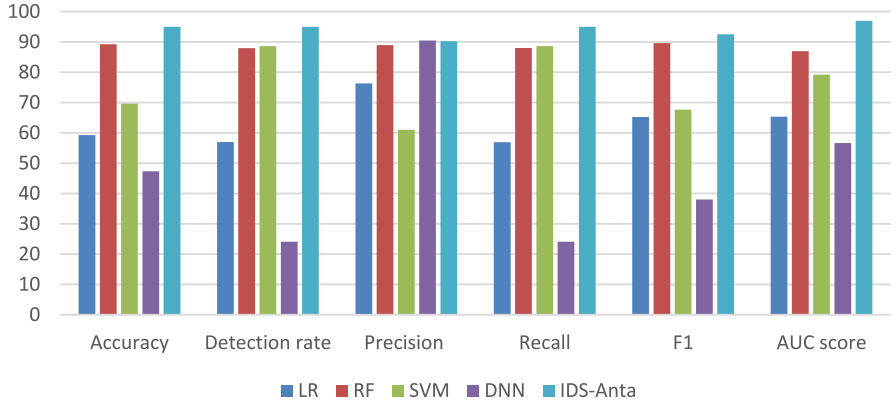

Fig. 3. Comparative outcome using ZOO attack.

Figs. 3 and 4 outline the comparative outcomes of the IDS-Anta, along with other classifiers, against ZOO and FGSM adversarial samples.

However, the IDS-Anta does not entirely negate them. This indicates the possibility of enhancing the model's resilience to strengthen its ability to defend against adversarial attacks. It is necessary to mention that the proposed model has certain limitations. Although achieving an impassable defense is challenging, the proposed model seeks to substantially enhance a potential attacker's time and computing resources to carry out an attack effectively. In practical situations, the increased duration and computing expenses can render an attack ineffective or economically impractical, serving as a deterrent and enhancing security measures.

3. Software impacts

The development of IDS-Anta addresses the practical implementation of defense mechanisms against adversarial attacks. Several existing studies are available based on ML and DL IDS, but publicly available code with defense strategies against adversarial attacks is limited. Our

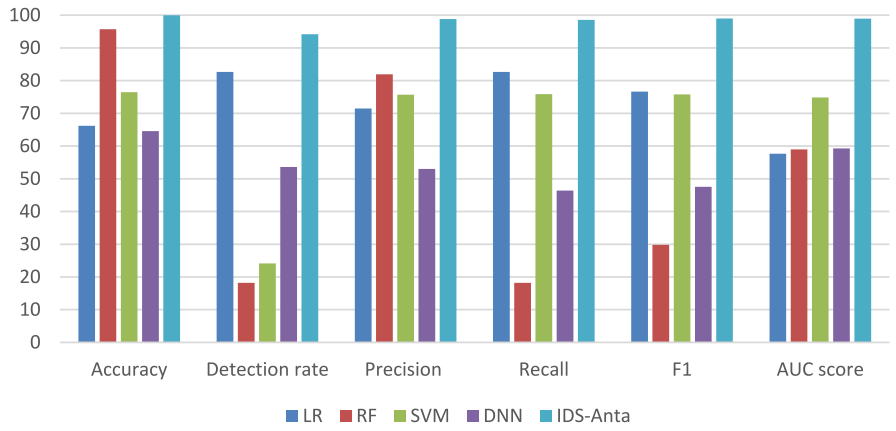

Fig. 4. Comparative outcome using FGSM attack.

source code is publicly accessible under the MIT license to extend the existing research on adversarial machine learning. This software is easy to use and has been implemented in Python, extensively used in ML and DL-based projects.

The presented software is not just a tool but a practical solution that helps researchers and system administrators understand the impact of adversarial attacks on IDS. Our software demonstrates two techniques, i.e., FGSM and ZOO, to generate adversarial attacks. It also uses two defense mechanisms, i.e., the MAB technique, to optimally choose the most suitable-fit classifier with Thomson sampling to balance and enhance attack detection. The exploratory analysis of three public benchmark datasets demonstrates that IDS-Anta can prevent and efficiently identify adversarial attacks. Hence, researchers can significantly benefit from utilizing the IDS-Anta software to develop proficiency in state-of-the-art approaches to enhance their IDS against adversarial attacks.

IDS-Anta has significantly supported our research on adversarial attacks and mitigating cyber threats using different approaches. This software has been utilized in cases that contribute to extending the study of adversarial machine learning, specifically in generating and developing defense mechanisms to protect against adversarial attacks on IDS. Two published research papers, which were supported in this approach, include:

1. Research to create a cybersecurity dataset, extensive feature engineering for data processing, and propose a framework using deep learning-based techniques to detect attacks, published in [8].
2. Research to develop adversarial attacks using different techniques and propose an optimized model with a defense mechanism to analyze the performance of IDS against adversarial attacks was published in [11].

This tool made a comparable contribution to the existing studies [12,14,25]. These studies presented a defense mechanism for IDS against adversarial attacks. The defense mechanism of the presented model's flexibility facilitates the possibility for researchers to explore other domains, but it is not limited to image security and monitoring systems.

4. Conclusions and future work

The severity and complexity of cyber-attacks are increasing. It is crucial to identify different types of intrusions and understand their techniques. The IDS-Anta software collection offers a user-friendly model to protect against adversarial attacks. Because of its straightforward implementation and concise explanation, cybersecurity researchers can utilize this code against adversarial samples in IDS.

Different prospective research and development possibilities arise based on this study's insights and can be extended and enhanced in three primary research areas. First, prospective studies can analyze the influence of materializing adversarial attack methods with different types of datasets on IDS. Second, further research should undertake comprehensive studies into black-box attacks to address the constraints associated with normal attack scenarios. Third, by integrating the heuristic and signature-based approaches, one can thoroughly examine potential threats and minimize the occurrence of false alarms.

CRediT authorship contribution statement

Kousik Barik: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Sanjay Misra:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank their departments for providing research environments and facilities to perform this research. The open access funding is provided by Østfold University College, Halden, Norway.

References

[1] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, Survey of intrusion detection systems: techniques, datasets, and challenges, Cybersecurity 2 (1) (2019) 1–22, http://dx.doi.org/10.1186/s42400-019-0038-7.

[2] K. Samunnisa, G.S.V. Kumar, K. Madhavi, Intrusion detection system in distributed cloud computing: Hybrid clustering and classification methods, Meas. Sens. 25 (2023) http://dx.doi.org/10.1016/j.measen.2022.100612.

[3] Y. Guo, A review of machine learning-based zero-day attack detection: Challenges and future directions, Comput. Commun. 198 (2023) 175–185, http://dx.doi.org/10.1016/j.comcom.2022.11.001.

[4] M.H.L. Louk, B.A. Tama, Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system, Expert Syst. Appl. 213 (2023) 119030, http://dx.doi.org/10.1016/j.eswa.2022.119030.

[5] O.H. Abdulganiyu, T.A. Tchakoucht, Y.K. Saheed, Towards an efficient model for network intrusion detection system (IDS): systematic literature review, Wirel. Netw. 30 (1) (2024) 453–482, http://dx.doi.org/10.1007/s11276-023-03495-2.

[6] N.S. Musa, N.M. Mirza, S.H. Rafique, A. Abdallah, T. Murugan, Machine learning and deep learning techniques for distributed denial of service anomaly detection in software defined networks-current research solutions, IEEE Access (2024) http://dx.doi.org/10.1109/ACCESS.2024.3360868.

[7] L. Yang, A. Shami, IDS-ML: An open source code for Intrusion Detection System development using Machine Learning, Softw. Impacts 14 (2022) 100446, http://dx.doi.org/10.1016/j.simpa.2022.100446.

[8] K. Barik, S. Misra, K. Konar, L. Fernandez-Sanz, M. Koyuncu, Cybersecurity deep: Approaches, attacks dataset, and comparative study, Appl. Artif. Intell. 36 (1) (2022) http://dx.doi.org/10.1080/08839514.2022.2055399.

[9] P. Bountakas, A. Zarras, A. Lekidis, C. Xenakis, Defense strategies for adversarial machine learning: A survey, Comp. Sci. Rev. 49 (2023) http://dx.doi.org/10.1016/j.cosrev.2023.100573.

[10] Sicong Han, Chenhao Lin, Chao Shen, Qian Wang, Xiaohong Guan, Interpreting adversarial examples in deep learning: A review, ACM Comput. Surv. 55 (14s) (2023) 1–38, http://dx.doi.org/10.1145/3594869.

[11] K. Barik, S. Misra, L. Fernandez-Sanz, Adversarial attack detection framework based on optimized weighted conditional stepwise adversarial network, Int. J. Inf. Secur. (2024) http://dx.doi.org/10.1007/s10207-024-00844-w.

[12] A. Paya, S. Arroni, V. García-Díaz, A. Gómez, Apollon: A robust defense system against Adversarial Machine Learning attacks in Intrusion Detection Systems, Comput. Secur. 136 (2024) http://dx.doi.org/10.1016/j.cose.2023.103546.

[13] M. Macas, C. Wu, W. Fuertes, Adversarial examples: A survey of attacks and defenses in deep learning-enabled cybersecurity systems, Expert Syst. Appl. (2023) http://dx.doi.org/10.1016/j.eswa.2023.122223.

[14] H. Mohammadian, A.A. Ghorbani, A.H. Lashkari, A gradient-based approach for adversarial attack on deep learning-based network intrusion detection systems, Appl. Soft Comput. 137 (2023) http://dx.doi.org/10.1016/j.asoc.2023.110173.

[15] A.V. Turukmane, R. Devendiran, M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning, Comput. Secur. 137 (2024) http://dx.doi.org/10.1016/j.cose.2023.103587.

[16] L. Dekel, I. Leybovich, P. Zilberman, R. Puzis, MABAT: A multi-armed bandit approach for threat-hunting, IEEE Trans. Inf. Forensics Secur. 18 (2022) http://dx.doi.org/10.1109/TIFS.2022.3215010.

[17] C. Kalkanlı, A. Özgür, Asymptotic performance of Thompson sampling for batched multi-armed bandits, IEEE Trans. Inform. Theory (2023) http://dx.doi.org/10.1109/TIT.2023.3274678.

[18] A. Alsarhan, M. Alauthman, E.A. Alshdaifat, A.R. Al-Ghuwairi, A. Al-Dubai, Machine learning-driven optimization for SVM-based intrusion detection system in vehicular ad hoc networks, J. Ambient Intell. Humaniz. Comput. 14 (5) (2023) http://dx.doi.org/10.1007/s12652-021-02963-x.

[19] I.H. Hassan, M. Abdullahi, M.M. Aliyu, S.A. Yusuf, A. Abdulrahim, An improved binary manta ray foraging optimization algorithm based feature selection and random forest classifier for network intrusion detection, Intell. Syst. Appl. 16 (2022) http://dx.doi.org/10.1016/j.iswa.2022.200114.

[20] A.A. Alqarni, Toward support-vector machine-based ant colony optimization algorithms for intrusion detection, Soft Comput. 27 (10) (2023) http://dx.doi.org/10.1007/s00500-023-07906-6.

[21] J. Zhu, X. Liu, An integrated intrusion detection framework based on subspace clustering and ensemble learning, Comput. Electr. Eng. 115 (2024) http://dx.doi.org/10.1016/j.compeleceng.2024.109113.

[22] A. Thakkar, R. Lohiya, Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System, Inf. Fusion 90 (2023) http://dx.doi.org/10.1016/j.inffus.2022.09.026.

[23] A. Yulianto, P. Sukarno, N.A. Suwastika, Improving Adaboost-Based Intrusion Detection System (IDS) Performance on CIC IDS 2017 Dataset, in: Journal of Physics: Conference Series, vol. 1192, IOP Publishing, 2019, 012018, http://dx.doi.org/10.1088/1742-6596/1192/1/012018.

[24] I.F. Kilincer, F. Ertam, A. Sengur, Machine learning methods for cyber security intrusion detection: Datasets and comparative study, Comput. Netw. 188 (2021) http://dx.doi.org/10.1016/j.comnet.2021.107840.

[25] M. Mittal, K. Kumar, S. Behal, Deep learning approaches for detecting DDoS attacks: A systematic review, Soft Comput. 27 (18) (2023) http://dx.doi.org/10.1007/s00500-021-06608-1.

[26] C. Eleftheriadis, A. Symeonidis, P. Katsaros, Adversarial robustness improvement for deep neural networks, Mach. Vis. Appl. 35 (3) (2024) http://dx.doi.org/10.1007/s00138-024-01519-1.