

Intrusion Detection System Using Machine Learning

Shreyas Bairagi¹, Shubham Shelke², Kunal Ghumare³, Apurv Gaikwad⁴,
Ram Kumar Solanki⁵, Pawan R. Bhaladhare⁶

^{1,2,3,4}B.Tech (Scholar), School of Computer Science & Engineering , Sandip University ,
Nashik , India

⁵Assistant Professor , School of Computer Science & Engineering , Sandip University ,
Nashik , India

⁶ Professor , School of Computer Science & Engineering , Sandip University , Nashik , India

Email: ¹shreyasbairagi0519@gmail.com,
²shelkeshubham752@gmail.com,
³Kunalghumare5@gmail.com,
⁴apurvgaikwad03@gmail.com,
⁵ramkumar.solanki@sandipuniversity.edu.in,
⁶pawan.bhaladhare@sandipuniversity.edu.in

Abstract: For protecting and securing the network, Intrusion Detection Systems through hidden intrusion have become a popular and important issue in the network security sphere. The detection of attacks is the first step to securing any system. In this paper, the focus is on seven different attacks, including Brute Force attacks, Heartbleed/ Denial-of-service(dos), Web Attacks, Infiltration, Botnet, Port overlook, and Distributed Denial of Service (DDoS). We calculate features deduced from CICIDS- 2017 Dataset for these attacks. By using colorful subset-grounded point selection ways the performance of the attack has been linked to numerous features. Using these ways, the applicable group of attributes for changing every attack with affiliated bracket algorithms has been determined. Simulations of these ways present that unwanted points can be removed from attack detection ways and find the most precious set of attributes for a definite bracket algorithm with discretization and without discretization, which ameliorates the performance of the IDS preface.

1. INTRODUCTION

The Intrusion Detection Evaluation collection (CICIDS2017) is a collection of network businesses that includes both regular business and dissembled data brought on by deliberate attacks on a test network. For computer networks to be properly defended against outside threats, intrusion detection, and prevention systems are essential. For these systems to perform well and constantly distinguish between a secure network and one that has been compromised, current and material data must be used to educate them. With a variety of realistic typical network assaults used to pretend attacks on a controlled network, the IDS 2017 Dataset was developed with the mitigation of ultramodern attacks in mind.

The Intrusion Detection System (IDS) is a important tool for defending networks from numerous kinds of attacks. IDSs are primarily used to identify and stop unauthorized network access. Host-based intrusion detection systems (HIDS) and network-based intrusion detection systems (NIDS) are the two subcategories of IDS. While NIDS analyses network business,

HIDS focuses on a single host's conduct. NIDS will be the main content of this essay. Different Intrusion Detection Method Types of Intrusion Detection ways There are several ways used in intrusion detection systems. The most generally used are – Signature-Based Detection – it also known as a rule- based detection, which involves comparing the network business against a set of predefined rules or signatures. However, the IDS raises an alarm If the business matches a hand. Anomaly- Based Detection Anomaly- based detection involves covering the network business and relating diversions from normal gester. The IDS raises an alarm if the business deviates from the anticipated gester. Hybrid Detection combines signature-based and anomaly-based detection ways. This approach provides better detection delicacy than either of the individual ways.

Literature Survey

According to Vilhelm Gustavsson's [1] study proposal, the collection of attacks that were tested for this project are assaulted from the CICIDS2017 dataset. Since Zeek is unable to block malicious traffic, the emphasis of this thesis is on detecting it. No neural networks or deep learning are used; machine learning is restricted to statistical learning techniques. There is no user interface created for the alarms generated by machine learning categorization; they are just output to a text file or terminal. Both Luke Hsiao and Stephen Ibanez. [2] In this study, we use a standard approach to applying machine learning for NIDS, framing the issue as an unsupervised anomaly detection challenge, where we want to train a model to recognize regular, attack-free data and afterward recognize abnormal, potentially malicious traffic. To detect anomalies, we use two different techniques: a non-stationary model based on the Packet Header Anomaly Detection and a stationary model utilizing a combination of Gaussians without taking time into account.

Xilei Wang, Liru Long, and Xiaoxi Zhu. This study got its training and test data sets from the KDDCup99 website, which hosts an annual Data Mining and Knowledge Detection competition on behalf of the ACM Special Interest Group. To create network connection records, which each contain 41 feature fields and a class field, raw TCP dump data from a simulated LAN is analyzed.

A Hiten Patel. In this study, a classical machine learning (ML)-based anomaly detection system is used to identify and classify network traffic by examining manually derived network traffic attributes. Such methods continue to have a high false positive rate, which severely reduces the effectiveness of in-time detection, adds a significant amount of manual scrutinizing work, and is unable to identify any undiscovered and novel (0-day) assaults. On the other hand, Deep Learning-based systems have been demonstrated to automatically detect new features and attack patterns to uncover new attacks in this constantly changing environment. Deep Learning-based systems can not only analyze the manually extracted features but can also automatically extract the features from the original traffic.

Ziaddon Salama A. Mostafa, Robiah Yusof, Nazrulzhar Bahaman, and Kamil Maseer. presented in [5] Both supervised and unsupervised learning techniques can be used to train an ML or DL system. Classification is carried out through supervised learning using data instances that have been marked during the training stage. The ANN, DT (both versions c4.5, ID3), k-NN, NB, RF, SVM, and CNN are examples of supervised learning algorithms. Unsupervised learning is used to identify unlabeled data instances, with clustering as the main learning strategy. The techniques for unsupervised learning include SOM, EM clustering, and k-means clustering.

2. METHODOLOGY

The following are important modules that were implemented in the project:

1. Dataset preparation
2. Multiple-class categorization
3. assessing ML models
4. Predicting network packet outcomes

1. Dataset preparation on the CI-CIDS 2017 Dataset, exploratory data analysis was carried out, and feature extraction was completed using Python libraries and domain knowledge research.

2. Different attacks are identified in the dataset using a multi-class classification method by joining the dataset's files together. SMOTE-TOMEK function data balancing for multi-class classification.

3. Importing multiple machine learning models from the Sklearn package for evaluation. dividing the data into training and testing, fitting the training and test sets of data to the models, and analyzing the confusion metrics derived from those models.

4. Network packet outcome prediction- The flask-web framework is made to examine real-time network packets using our most accurate machine learning model, and network sniffing is carried out using the sniffer tool Cyc flowmeter.

3. Diagram

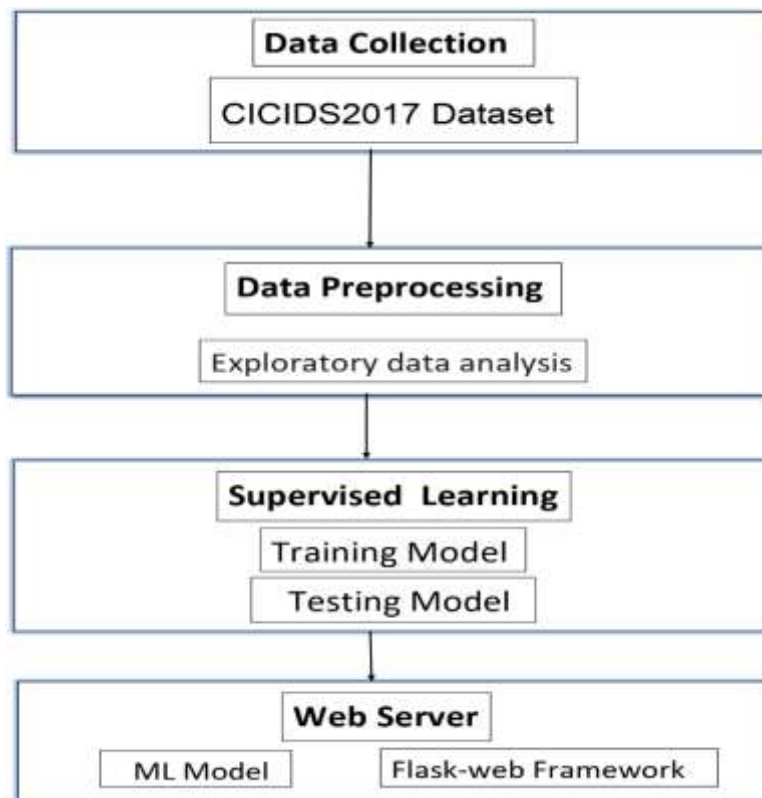


Figure 1.1: Flowchart Diagram

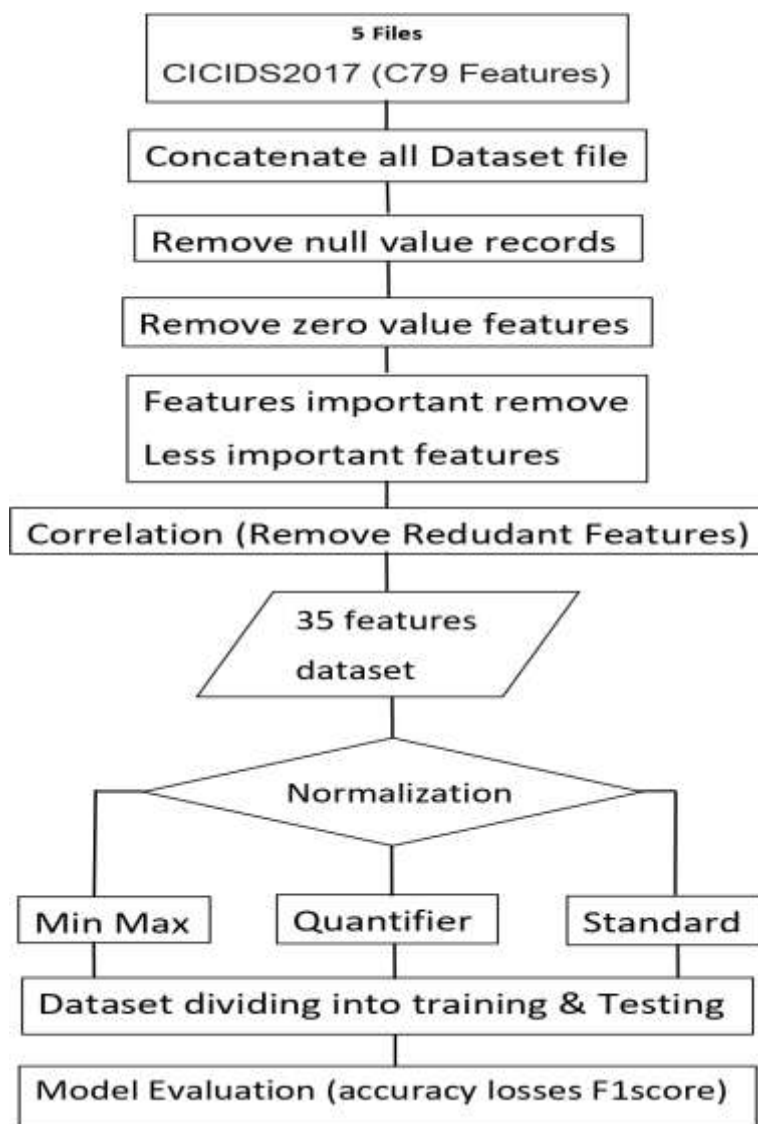


Figure 1.2: Data Flow Diagram

3. RESULTS

The purpose is to validate that each unit of the software performs as designed. The Proposed System works in following manner: In Network Sniffing Module, we select Interface, easily track and analyze network traffic & also get information about IP Address. In Packets storing module, all the sniffed packages are gathered. In Packets labeling module stage, all the Packets are Identified and Labelled. In Result storing module, we collect information about of total flows and also it stores total number of data flows and data attacks. UI (User Interface) displays Machine Learning Model Information, attacks, Data Flows Information, & How many packets transferred.

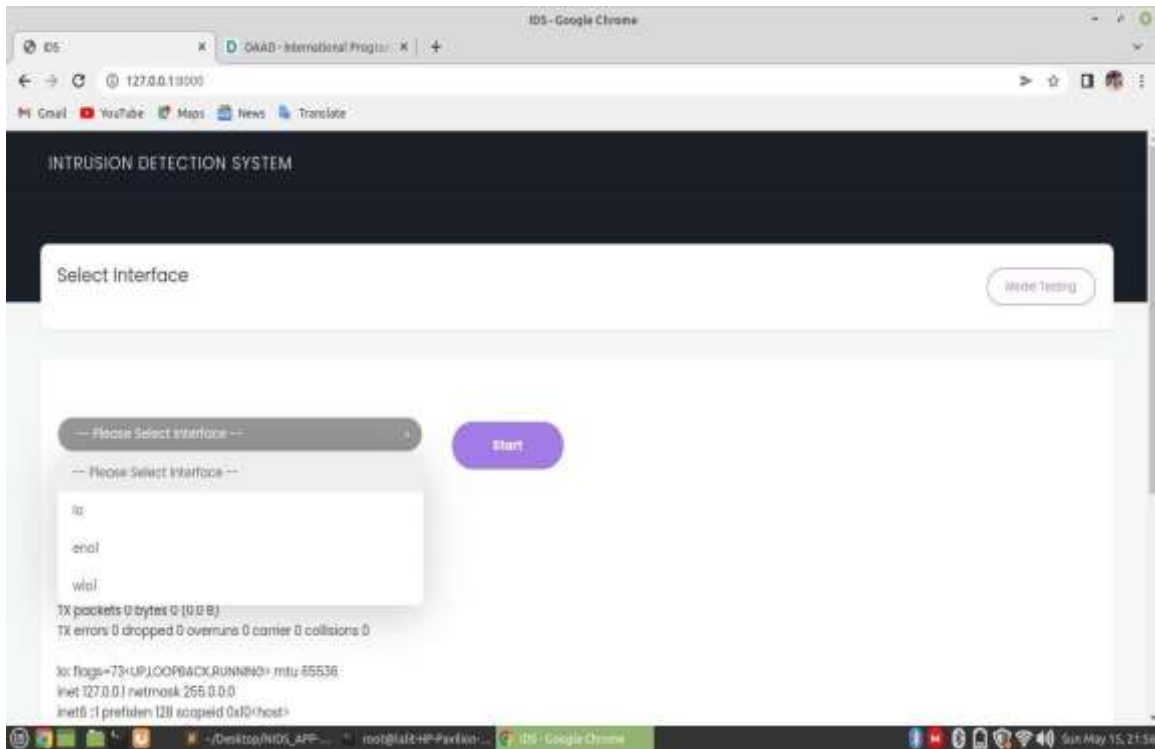


Fig 2.1 Home Page For Selecting Interface

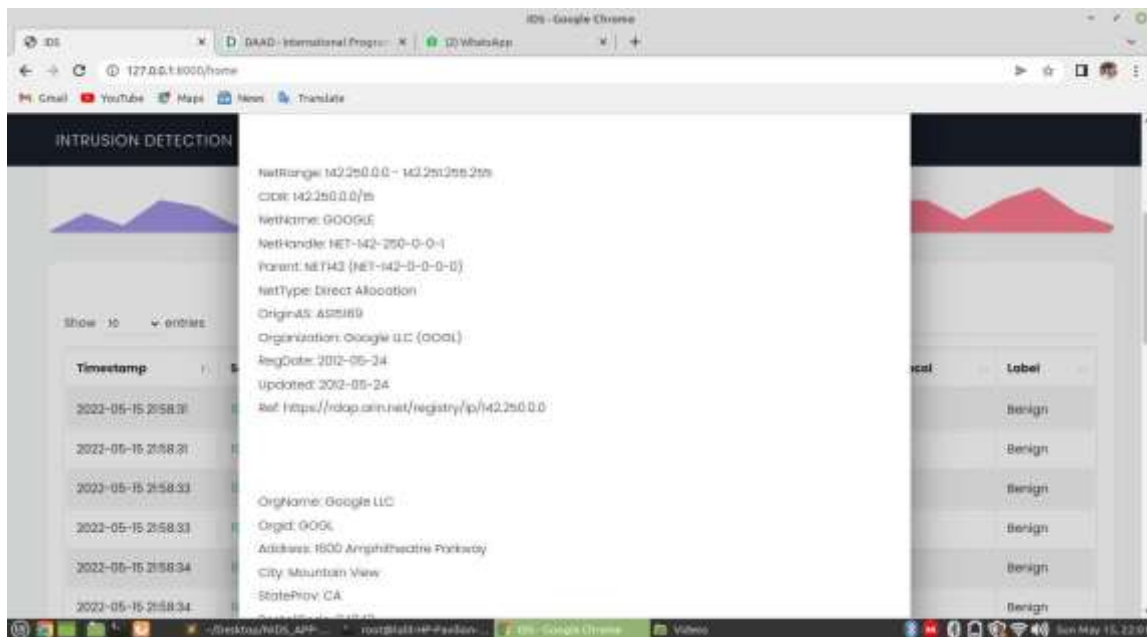


Fig 2.2 IP Information

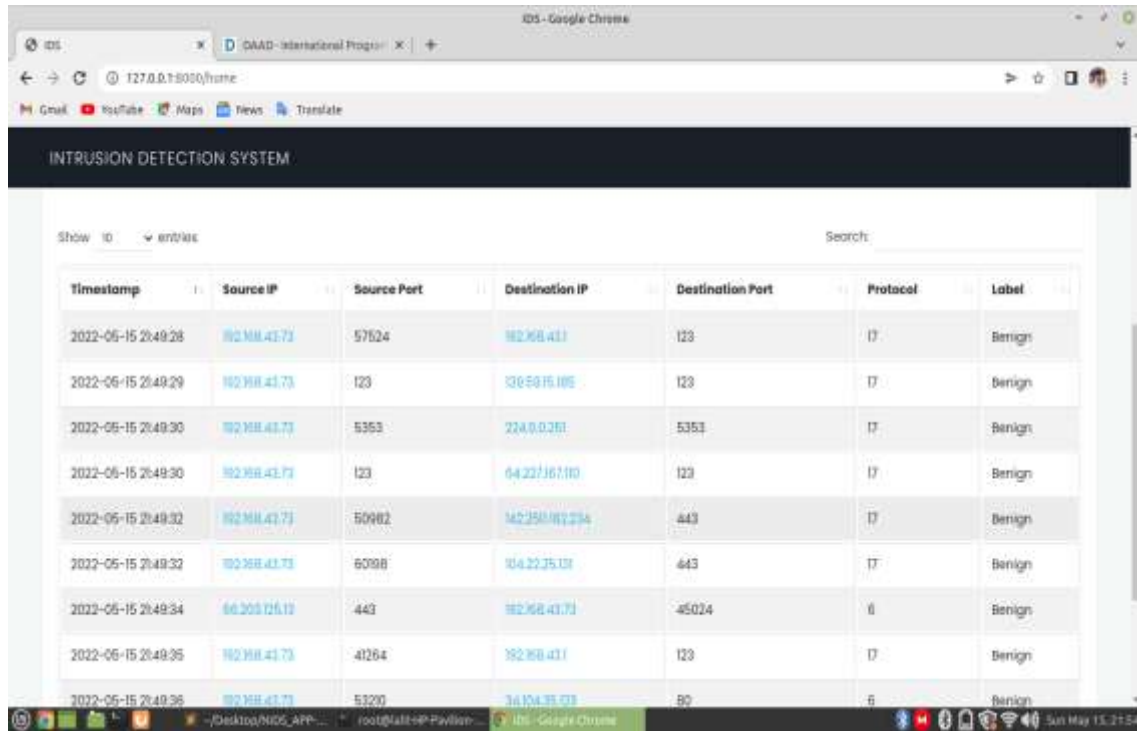


Fig 2.3 Gathered Sniffed Packets

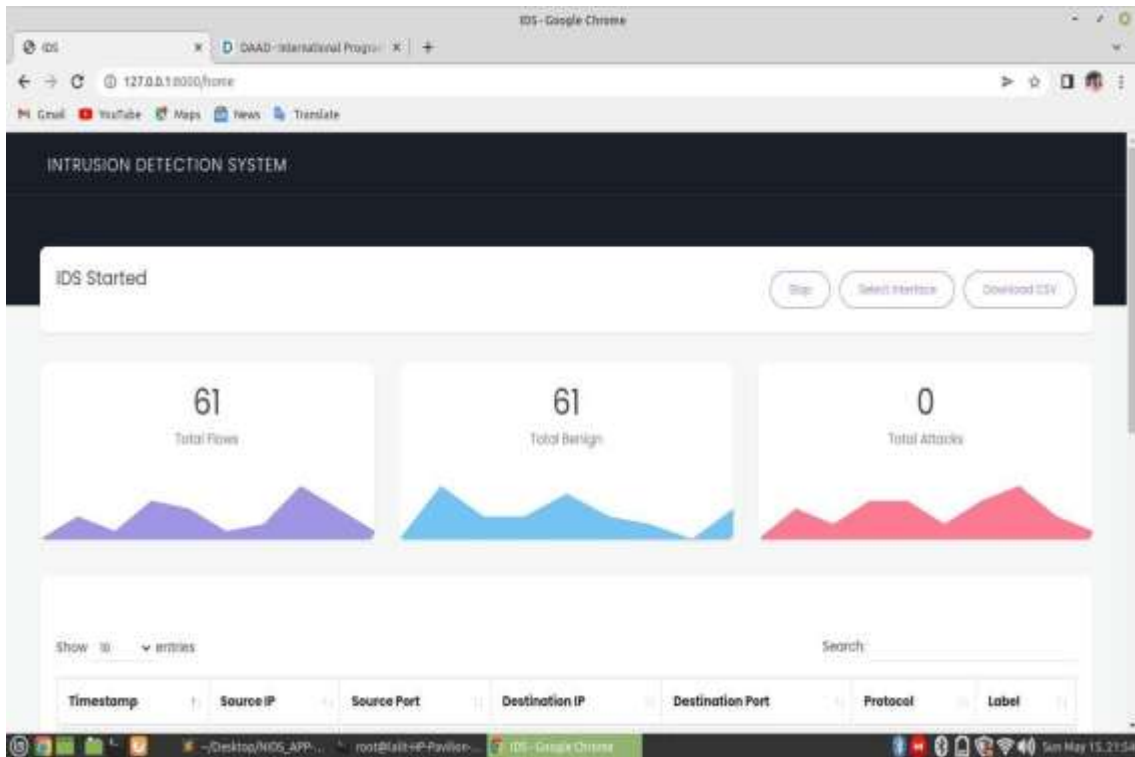


Fig 2.4 Total Number of Network Packets Gathered and Total Attacks

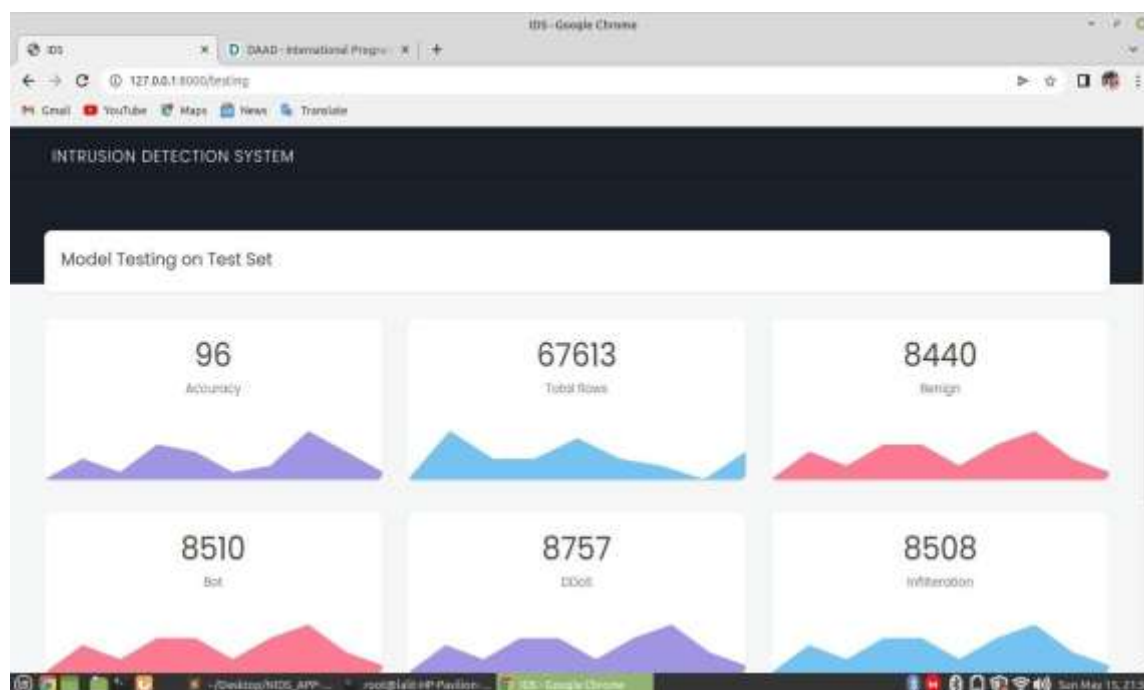


Fig 2.5 Machine Learning Model Information

4. CONCLUSION

When using signature-based detection, sometimes referred to as rule-based detection, network data is compared to a list of predetermined rules or signatures. The IDS triggers an alarm if the traffic matches a signature. This publication analyses earlier studies on AIDS that make use of various datasets. Additionally, testing is how well similar ML-AIDS models perform while trying to identify attacks on a binary dataset. These models demonstrate limits in the multi-classification detection of novel attack types. Additionally, most of these studies employ accuracy as a primary evaluation criterion, which prevents them from fairly comparing and evaluating different ML-AIDS. This work offers a benchmarking approach that leverages real data and several procedures to address this issue, ensuring an accurate assessment of AIDS performance based on ML Algorithms. The evaluation considers several distinct factors, different types of raw network datasets, and suggested performance indicators. Benchmarking experiments are also run utilizing supervised and unsupervised ML techniques (such as ANN, DT, k-NN, NB, RF, SVM, CNN, EM, K-means, and SOM) to evaluate the development of successful ML-AIDS. By executing web assaults against the CICIDS2017 datasets, the tests are carried out. The findings of the experiment demonstrate the lack of a single ML algorithm capable of identifying all kinds of web attacks. Due to their high FP and FN alarms, the SOM-AIDS and EM-AIDS models perform poorly compared to the K-NN-AIDS, DT-AIDS, and NB-AIDS models, which all achieve great performance. Researchers can build a better AIDS model and compare their findings to those of this study with the assistance of the benchmarking approach that has been suggested. Future studies should focus on measuring the impact of feature selection and consider new methodological steps for developing a deep-learning CNN-AIDS model.

5. Acknowledgement

I would like to express my sincere gratitude to Prof. Dr. Solanki, Assistant Professor and Prof.

Dr. Bhaladhare, Head of Department, SOCSE, Sandip University, Nashik, India, for giving us the much-needed encouragement to translate my in-depth research into a survey paper.

5. REFERENCES

- [1] Ziadoon Kamil Maseer, Robiah Yusof, Nazrulazhar Bahaman, Salama Mostafa- "Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset."
- [2] Zeeshan Ahmad, Adnan Shahid Khan, Cheah Wai Shiang, Johari Abdullah, Farhan Ahmad- "Network intrusion detection system: A systematic study of machine learning and deep learning approaches ©2020 IEEE
- [3] Aman Juneja- Evaluation of ML Algorithms for Intrusion Detection Systems
- [4] Tan, J.; Goyal, S.B.; Singh Rajawat, A.; Jan, T.; Azizi, N.; Prasad, M. Anti-Counterfeiting and Traceability Consensus Algorithm Based on Weightage to Contributors in a Food Supply Chain of Industry 4.0. *Sustainability* 2023, 15, 7855. <https://doi.org/10.3390/su15107855>
- [5] Rajawat, A.S. et al. (2023). Real-Time Driver Sleepiness Detection and Classification Using Fusion Deep Learning Algorithm. In: Singh, Y., Singh, P.K., Kolekar, M.H., Kar, A.K., Gonçalves, P.J.S. (eds) *Proceedings of International Conference on Recent Innovations in Computing. Lecture Notes in Electrical Engineering*, vol 1001. Springer, Singapore. https://doi.org/10.1007/978-981-19-9876-8_34.
- [6] Rajawat, A.S.; Goyal, S.B.; Bedi, P.; Verma, C.; Ionete, E.I.; Raboaca, M.S. 5G-Enabled Cyber-Physical Systems for Smart Transportation Using Blockchain Technology. *Mathematics* 2023, 11, 679. <https://doi.org/10.3390/math11030679>
- [7] Rajawat, A.S.; Goyal, S.B.; Chauhan, C.; Bedi, P.; Prasad, M.; Jan, T. Cognitive Adaptive Systems for Industrial Internet of Things Using Reinforcement Algorithm. *Electronics* 2023, 12, 217. <https://doi.org/10.3390/electronics12010217>.
- [8] Nagaraj, S.; Kathole, A.B.; Arya, L.; Tyagi, N.; Goyal, S.B.; Rajawat, A.S.; Raboaca, M.S.; Mihaltan, T.C.; Verma, C.; Suci, G. Improved Secure Encryption with Energy Optimization Using Random Permutation Pseudo Algorithm Based on Internet of Thing in Wireless Sensor Networks. *Energies* 2023, 16, 8. <https://doi.org/10.3390/en16010008>.
- [9] R. S. Chouhan et al., "Experimental Analysis for Position Estimation using Trilateration and RSSI in Industry 4.0," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 904-908, doi: 10.1109/SMART55829.2022.10047276.
- [10] Rajawat, A.S. et al. (2023). Real-Time Driver Sleepiness Detection and Classification Using Fusion Deep Learning Algorithm. In: Singh, Y., Singh, P.K., Kolekar, M.H., Kar, A.K., Gonçalves, P.J.S. (eds) *Proceedings of International Conference on Recent Innovations in Computing. Lecture Notes in Electrical Engineering*, vol 1001. Springer, Singapore. https://doi.org/10.1007/978-981-19-9876-8_34
- [11] S. Rajawat, S. B. Goyal, P. Bedi, N. B. Constantin, M. S. Raboaca and C. Verma, "Cyber-Physical System for Industrial Automation Using Quantum Deep Learning," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 897-903, doi: 10.1109/SMART55829.2022.10047730.
- [12] S. Rajawat et al., "Security Analysis for Threats to Patient Data in the Medical Internet of Things," 2022 11th International Conference on System Modeling &

- Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 248-253, doi: 10.1109/SMART55829.2022.10047322.
- [13] P. Pant et al., "Using Machine Learning for Industry 5.0 Efficiency Prediction Based on Security and Proposing Models to Enhance Efficiency," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 909-914, doi: 10.1109/SMART55829.2022.10047387.
- [14] P. Pant et al., "AI based Technologies for International Space Station and Space Data," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 19-25, doi: 10.1109/SMART55829.2022.10046956
- [15] Rajawat, A.S.; Goyal, S.B.; Bedi, P.; Simoff, S.; Jan, T.; Prasad, M. Smart Scalable ML-Blockchain Framework for Large-Scale Clinical Information Sharing. Appl. Sci. 2022, 12, 10795. <https://doi.org/10.3390/app122110795>.
- [16] S. Rajawat et al., "Visual Cryptography and Blockchain for Protecting Against Phishing Attacks on Electronic Voting Systems," 2022 International Conference and Exposition on Electrical And Power Engineering (EPE), Iasi, Romania, 2022, pp. 663-666, doi: 10.1109/EPE56121.2022.9959765.
- [17] S. Rajawat et al., "Electrical Fault Detection for Industry 4.0 using Fusion deep Learning Algorithm," 2022 International Conference and Exposition on Electrical And Power Engineering (EPE), Iasi, Romania, 2022, pp. 658-662, doi: 10.1109/EPE56121.2022.9959762.
- [18] Rajawat, Anand Singh and Chauhan, Chetan and Goyal, S B and Bhaladhare, Pawan R and Rout, Dillip and Gaidhani, Abhay R, Utilization Of Renewable Energy For Industrial Applications Using Quantum Computing (August 11, 2022). Available at SSRN: <https://ssrn.com/abstract=4187814> or <http://dx.doi.org/10.2139/ssrn.4187814>
- [19] Anand Singh Rajawat, Pradeep Bedi, S. B. Goyal, Sandeep Kautish, Zhang Xihua, Hanan Aljuaid, Ali Wagdy Mohamed, "Dark Web Data Classification Using Neural Network", Computational Intelligence and Neuroscience, vol. 2022, Article ID 8393318, 11 pages, 2022. <https://doi.org/10.1155/2022/8393318>.
- [20] Piyush Pant, Anand Singh Rajawat, S.B. Goyal, Pradeep Bedi, Chaman Verma, Maria Simona Raboaca, Florentina Magda Enescu, Authentication and Authorization in Modern Web Apps for Data Security Using Nodejs and Role of Dark Web, Procedia Computer Science, Volume 215, 2022, Pages 781-790, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2022.12.080>.
- [21] Robin Singh Chouhan, Anand Singh Rajawat, SB Goyal, Pradeep Bedi , AI-Enabled Augmented Reality-Based Shared Collaborative Experience, Book AI-Enabled Multiple-Criteria Decision-Making Approaches for Healthcare Management Pages 85-96 Publisher IGI Global.
- [22] Anand Singh Rajawat, Pradeep Bedi, S. B. Goyal, Piyush Kumar Shukla, Atef Zaguia, Aakriti Jain, Mohammad Monirujjaman Khan, "Reformist Framework for Improving Human Security for Mobile Robots in Industry 4.0", Mobile Information Systems, vol. 2021, Article ID 4744220, 10 pages, 2021. <https://doi.org/10.1155/2021/4744220>
- [23] S. Srivastava and R. Kumar, "Indirect method to measure software quality using CK-OO suite," 2013 International Conference on Intelligent Systems and Signal Processing (ISSP), 2013, pp. 47-51, doi: 10.1109/ISSP.2013.6526872.
- [24] Ram Kumar, Gunja Varshney , Tourism Crisis Evaluation Using Fuzzy Artificial Neural network, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, June 2011
- [25] Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, "A Survey Paper on Altered Fingerprint Identification & Classification" International Journal of Electronics

- Communication and Computer Engineering Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209
- [26] Kumar, R., Singh, J.P., Srivastava, G. (2014). Altered Fingerprint Identification and Classification Using SP Detection and Fuzzy Classification. In: , et al. Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012), December 28-30, 2012. Advances in Intelligent Systems and Computing, vol 236. Springer, New Delhi. https://doi.org/10.1007/978-81-322-1602-5_139
- [27] Gite S.N, Dharmadhikari D.D, Ram Kumar, ” Educational Decision Making Based On GIS” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-1, Issue-1, April 2012.
- [28] Ram Kumar, Sarvesh Kumar, Kolte V. S.,” A Model for Intrusion Detection Based on Undefined Distance”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1 Issue-5, November 2011
- [29] Vibhor Mahajan, Ashutosh Dwivedi, Sairaj Kulkarni, Md Abdullah Ali, Ram Kumar Solanki, ” Face Mask Detection Using Machine Learning”, International Research Journal of Modernization in Engineering Technology and Science, Volume:04/Issue:05/May-2022
- [30] Kumar, Ram and Sonaje, Vaibhav P and Jadhav, Vandana and Kolpyakwar, Anirudha Anil and Ranjan, Mritunjay K and Solunke, Hiralal and Ghonge, Mangesh and Ghonge, Mangesh, Internet Of Things Security For Industrial Applications Using Computational Intelligence (August 11, 2022). Available at SSRN: <https://ssrn.com/abstract=4187998> or <http://dx.doi.org/10.2139/ssrn.4187998>
- [31] Kumar, Ram and Aher, Pushpalata and Zope, Sharmila and Patil, Nisha and Taskar, Avinash and Kale, Sunil M and Gadekar, Amit R, Intelligent Chat-Bot Using AI for Medical Care (August 11, 2022). Available at SSRN: <https://ssrn.com/abstract=4187948> or <http://dx.doi.org/10.2139/ssrn.4187948>
- [32] Kumar, Ram and Patil, Manoj, Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies (July 22, 2022). Available at SSRN: <https://ssrn.com/abstract=4182372>
- [33] Ram Kumar, Manoj Eknath Patil ,” Improved the Image Enhancement Using Filtering and Wavelet Transformation Methodologies”, Turkish Journal of Computer and Mathematics Education ,Vol.13 No.3(2022), 987-993.
- [34] Ram Kumar, Jasvinder Pal Singh, Gaurav Srivastava, “A Survey Paper on Altered Fingerprint Identification & Classification” International Journal of Electronics Communication and Computer Engineering ,Volume 3, Issue 5, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209.
- [35] Chetna kwatra, Bukya Mohan Babu, M.Praveen, Dr T.Sampath Kumar, Ram Kumar Solanki ,Dr A V R Mayuri. (2023). Modified Cnn Based Heart Disease Detection Integrated With Iot. Journal of Pharmaceutical Negative Results, 993–1001. <https://doi.org/10.47750/pnr.2023.14.S02.120>
- [36] Solanki, R. K., Rajawat, A. S., Gadekar, A. R., & Patil, M. E. (2023). Building a Conversational Chatbot Using Machine Learning: Towards a More Intelligent Healthcare Application. In M. Garcia, M. Lopez Cabrera, & R. de Almeida (Eds.), Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines (pp. 285-309). IGI Global. <https://doi.org/10.4018/978-1-6684-7164-7.ch013>
- [37] S. B. Goyal, A. S. Rajawat, R. K. Solanki, M. A. Majmi Zaaba and Z. A. Long, "Integrating AI With Cyber Security for Smart Industry 4.0 Application," 2023

- International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2023, pp. 1223-1232, doi: 10.1109/ICICT57646.2023.10134374.
- [38] Pardeshi, D., Rawat, P., Raj, A., Gadail, P., Solanki, R. K., & Bhaladhare, D. P. R. (2023). Efficient Approach for Detecting Cardiovascular Disease Using Machine Learning. *Int. J. of Aquatic Science*, 14(1), 308-321
- [39] Patle, S., Pal, S., Patil, S., Negi, S., Rout, D. D., & Solanki, D. R. K. (2023). Sun-Link Web Portal for Management for Sun Transportation. *Int. J. of Aquatic Science*, 14(1), 299-307.
- [40] Sayyed, T., Kodwani, S., Dodake, K., Adhayage, M., Solanki, R. K., & Bhaladhare, P. R. B. (2023). Intrusion Detection System. *Int. J. of Aquatic Science*, 14(1), 288-298.
- [41] Gupta, A., Sevak, H., Gupta, H., & Solanki, R. K. (2023). Swiggy Genie Clone Application. *Int. J. of Aquatic Science*, 14(1), 280-287.
- [42] Khode, K., Buwa, A., Borole, A., Gajbhiye, H., Gadekar, D. A., & Solanki, D. R. K. (2023). Live Stock Market Prediction Model Using Artificial Neural Network. *Int. J. of Aquatic Science*, 14(1), 333-340.
- [43] hire, S., Gorhe, S., Palod, T., Khalkar, A., Chauhan, D., & Solanki, D. K. (2023). First Copy Logo Detection System. *Int. J. of Aquatic Science*, 14(1), 322-332.