iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61765

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Fortifying Network Security: Pioneering Hybrid Machine Learning for BotNet Attack Detection

M V Varun[1], Gurudeep R[2], Yogitha S[3], Yashila Raju[4], Renuka Patil[5]
Dept. of AIML, KS Institute of Technology

Abstract: ***With the rising sophistication of cyber threats, the detection of BotNet attacks on networks has become a perilous contest for cybersecurity. This paper proposes an innovative approach leveraging a hybrid machine learning (ML) framework for the detection of BotNet attacks in network atmosphere conquers a notable accuracy of 98.63%. By amalgamation of different strengths of ML algorithms such as KNN and Decision Trees, our approach aims to enhance the accuracy and efficiency of BotNet detection. The methodology comprises feature extraction from network traffic data, followed by training and testing of the hybrid model on labelled datasets to identify patterns revealing of BotNet activity. Experimental outcomes reveal the effectiveness of the proposed approach in precisely detecting BotNet attacks while diminishing false positives. The hybrid ML approach offers a forthcoming avenue for fortifying network security and mitigating the risks associated with BotNet threats.***

I. INTRODUCTION

A botnet is a grid of devices infected with malicious software and scrupulous remotely by an individual, known as the controller or botmaster. These compromised devices are often referred to as zombies or bots, it can include PCs, smartphones, tablets and IoT devices. Botnet attacks commonly involve data theft, large-scale DDoS attacks, spam or phishing campaigns. Attackers leverage the united power of these devices to carry out illegitimate actions, facilitated through a command-and-control model. Anomaly detection uses ML to establish a baseline of normal network activity and detect deviations such as sudden traffic increases or unusual communication patterns indicative of botnet activity. Behavioural analysis employs ML models to identify botnet-related behaviour patterns in system logs and network data, assisting in threat detection. Supervised learning trains the models using labelled datasets to classify network traffic, while unsupervised learning detects botnet activity without predefined attack signatures. Feature selection, data preprocessing, training dataset, model training and evaluation are crucial steps in development and accessing ML based botnet detection.

The consecutive segments of this paper are structured as follows: II conducts a literature survey reviewing the existing study in the arena of Botnet attack detection by ML techniques. III outlines our proposed method, enlightening the conceptual approach with architecture. IV describes the experimental steps including data collection, preprocessing, model training and evaluation measures. The experimental results and performance evaluation of our proposed method are presented in V. VI involves in a discussion of our findings and the implications of our approach. VII provides concluding remarks and highlights avenues for future research in this domain. Through this paper, our objective is to make a meaningful contribution towards the advancement of using Machine Learning to detect botnet attacks on networks in a reliable manner.

II. LITERATURE SURVEY

Sl. No	*Publication Year*	Title	Description
1.	2023	Botnet Attack Detection in IoT Networks using CNN and LSTM	To increase security of IoT devices against botnet threats, the utilization of dimensionality reduction methods like PCA and autoencoder on the Bot-IoT dataset are most efficient. This approach helps to reduce feature dimensions, allowing the application of memory-efficient deep learning algorithms such as LSTM and CNN to detect botnet attacks. Evaluating performance metrics such as accuracy, precision, recall and the confusion matrix will be crucial in determining the effectiveness of the security measures.

2.	2023	A Hybrid Model for Botnet Detection using Machine Learning	A novel hybrid machine learning framework was developed for detection of botnet attacks in network traffic by joining k-means clustering, rule-based systems and decision trees. Using the CTU-13 dataset and features from Barnacles Mating Optimizer, the study demonstrated exceptional accuracy 99%. Notably, each individual component showed substantial accuracy rates: k-means (99.32%), decision tree (99.11%) and rule-based system (97.14%). Precision rates were also impressive, with k-means (98.93%), decision tree (98.37%) and rule-based system (95.93%) exhibiting high precision.
3.	2023	Botnet Dataset Overview Using Statistical Approach Based on Time Gap Activity Analysis	An innovative strategy for identification of botnet malware through the analysis of time intervals within botnet datasets using statistical analysis. Experimental findings reveal three distinct threshold values: the highest recorded at 4756 seconds, lowest at 28.69 seconds and average maximum time gap of 810.61 seconds. These thresholds demonstrate promising efficacy in distinguishing between botnet, normal and background activities, offering valuable insights into the detection of malicious botnet behaviour.
4.	2023	Detecting and Mitigating Botnet Attacks in Software-Defined Networks Using Deep Learning Techniques	Employing Software-Defined Networking (SDN) with external Deep Learning (DL) confronts security issues. This research adopts DL techniques to contest such threats in SDN environments. Leveraging a norm dataset and optimizing feature weights, a lightweight DL model is devised. The Convolutional Neural Networks outperforming achieving an accuracy 99% in detecting normal flows and 97% for identifying attacks.
5.	2022	Machine Learning based IoT BotNet Attack Detection Using Real-time Heterogeneous Data	This study delves into machine learning algorithms tailored for IoT device security, calculating Logistic Regression, Random Forest, K-Nearest Neighbours Gaussian NB, Decision Trees and Extreme Gradient Boosting for identification of malicious activities. Among these tree-based classifiers outshine LR and GNB achieving a binary classification accuracy of 99% and 0.99 F1-score. Multiclass results showcase DT, KNN, RF and XGB with accuracy of 98% and 0.98 F1-score. Furthermore, the research investigates the effects of feature reduction on accuracy and training duration.
6.	2022	AI-based botnet attack classification and detection in IoT devices.	Mounting use of domiciliary IoT devices carriages security perils, challenging traditional rule-based systems. AI offers a solution by retaining machine earning and deep learning to detect and classify IoT botnets. 6 ML and 3 DL models are evaluated, with top performers implemented as an API for higher security against threats.
7.	2021	Machine learning for misuse-based network intrusion detection: Overview, unified evaluation and feature choice comparison	Utilizing Network Intrusion Detection Systems (NIDS) empowered by artificial intelligence presents a dynamic shift from conventional hardcoding methods. Our study pioneers the integration of detection and identification scores, establishing universal benchmarks amidst diverse datasets and metrics. We devise an algorithmic workflow that converts raw packet data into machine learning inputs, enabling real-time, cost-effective

		framework	deep learning analysis of raw traffic attributes. Our findings demonstrate superior performance, surpassing existing methodologies and offering heightened network security measures.
8.	2021	Intrusion detection based on sequential information preserving log embedding methods and anomaly detection algorithms	Most traditional intrusion detection methods rely on pattern matching, drawing from expert knowledge, which can be easily bypassed by sophisticated strategies. An alternative proposed is an end-to-end method employing machine learning-based anomaly detection algorithms and sequential log embedding techniques. This approach aims to identify abnormal behaviour by preserving sequential information without solely relying on pre-defined patterns, offering potentially more robust detection capabilities against evolving intrusion tactics.
9.	2020	Systematic literature review on iot-based botnet attack	The swift proliferation of a specific technology owes much to its capacity to greatly enrich user experiences, as evidenced by its seamless integration across a myriad of devices. Yet, with the exponential rise of IoT devices, vulnerabilities have surfaced more prominently, paving the way for various types of attacks, notably the widespread IoT-based botnet attack. In investigation, the undertaken meticulous and methodical literature review on this topic, meticulously presenting and scrutinizing the existing state of the art. Through a systematic methodology, we aimed for thorough coverage of relevant studies, ensuring both replicability and rigor in our approach.
10.	2020	Iot-flock: An open-source framework for iot traffic generation	The Internet of Things (IoT) has emerged as a transformative force, captivating both researchers and businesses alike with its potential to revolutionize various aspects of human existence. From smart gadgets to advanced healthcare systems and interconnected industrial infrastructures, the IoT has reshaped daily life. However, recent cyberattacks exposing vulnerabilities in IoT security have raised significant concerns. Addressing these issues proves challenging due to the resource limitations of IoT devices and the distinct characteristics of IoT protocols, necessitating novel security approaches.
11.	2020	Botnet fingerprinting: a frequency distributions scheme for lightweight bot detection	Detecting bots efficiently is crucial for security, with new approaches shifting from flow-based to graph-based methods. Yet, scalability poses a hurdle. BotFP tackles this by streamlining communication graphs, focusing on frequency distributions of protocol attributes. It profiles host actions, refines through clustering or ML and distinguishes between benign users and bots. Tested on the CTU-13 dataset, BotFP demonstrates its lightweight nature, scalability and superior accuracy compared to conventional methods in managing extensive data volumes.

III. PROPOSED METHOD

BorutaPy plays a pivotal role in selecting crucial features from the dataset, thereby improving model efficiency. These selected features are then subjected to classification by individual models: KNN and Decision Tree, each possessing distinctive features.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Utilizing their respective features, these models make predictions regarding outcomes. Subsequently, through a Voting Classifier, predictions from all models are combined, drawing on diverse perspectives to generate a more accurate final prediction.

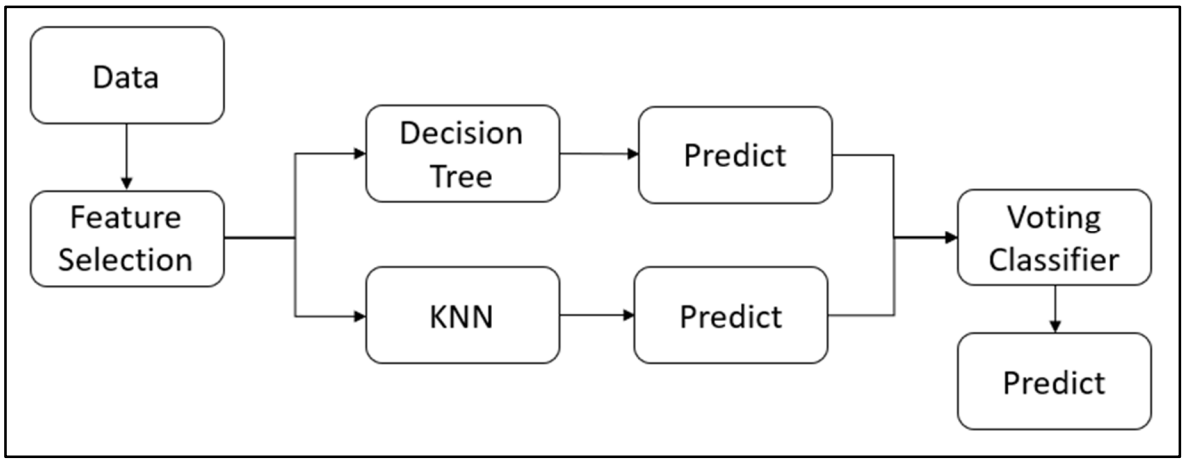


Fig 1: Architecture of Proposed model

This hybrid approach optimizes feature selection and utilizes the strengths of different classifiers, resulting in a robust model capable of handling varied data characteristics and enhancing overall prediction performance.

IV. EXPERIMENTAL STEPS

A. Data Collection

The dataset encompasses a large volume of diverse data entries collected from various sources. It consists of 104345 rows and 23 columns, predominantly comprising numeric data types with 20 columns. Additionally, 3 columns contain object data types, potentially representing textual or categorical information.

B. Preprocessing

Preprocessing involves cleaning and transforming datasets to enhance quality. Feature selection techniques are employed to retain relevant features, reducing dimensionality and enhancing model performance. This includes handling missing values, encoding categorical variables, scaling numerical features and removing outliers to ensure data integrity. Additionally, evaluating feature importance and selecting subsets based on statistical metrics or machine learning algorithms optimizes the dataset for analysis. These steps collectively refine the dataset, preparing it for accurate and reliable analysis, thus facilitating meaningful insights in research papers.

C. Model Training

The model training phase encompasses developing and optimizing machine learning algorithms for effective classification of instances in the dataset. This iterative process involves careful model selection, hyperparameter tuning, and cross-validation to ensure the model achieves optimal performance and generalizability across diverse network environments. Employing techniques such as supervised learning and ensemble methods, the model is trained on labelled data, iteratively refining its predictive capabilities to accurately detect and classify BotNet attacks in network traffic data. The trained model forms a foundational component of the research paper, demonstrating its efficacy in bolstering network security against sophisticated cyber threats.

D. Evaluation Measures

Evaluation measures play a pivotal role in assessing the performance of machine learning models across various tasks, including classification, regression and clustering. Common metrics include accuracy, precision, recall and F1-score, which provide insights into the model's effectiveness in correctly predicting outcomes. Additionally, confusion matrix metrics such as true positive, false positive, true negative and false negative rates offer a detailed understanding of the model's classification performance.

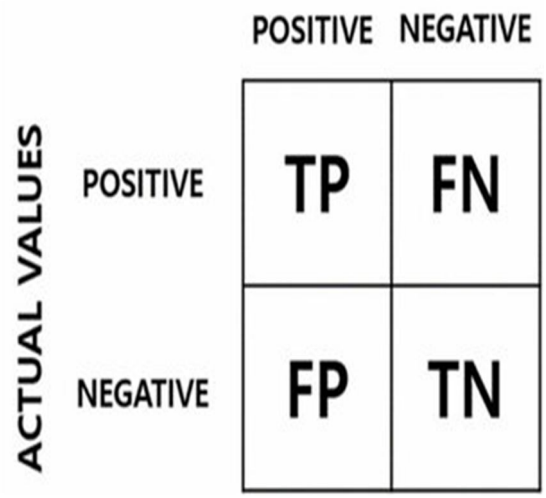


Fig 2: Confusion Matrix

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} X\ 100 \text{ ----(1)}$$

$$Precision = \frac{TP}{TP+FP} X\ 100 \text{ ----(2)}$$

$$Recall = \frac{TP}{TP+FN} X\ 100 \text{ ----(3)}$$

$$F1-score = \frac{2\times Precision\times Recall}{Precision+Recall} \text{ ----(4)}$$

By selecting appropriate evaluation measures tailored to the task at hand, researchers and practitioners can effectively gauge the model's performance and make informed decisions regarding its deployment and optimization. Accuracy (1): Represents the proportion of correctly classified instances, calculated as the ratio of true positives (TP) and true negatives (TN) to the total number of instances. Precision (2): Indicates the percentage of true positive instances among all instances projected as positive, hence assessing the accuracy of positive forecasts. Recall (Sensitivity) (3): Recall is the ratio of real positive cases to the total number of actual positive examples and it measures how well the model can recognize positive instances. F1-Score (4): Represents the harmonic mean of precision and recall providing a balanced measure of a classifier's performance.

V. EXPERIMENTAL RESULT

Our approach, driven by machine learning, attains exceptional metrics— accuracy, precision, recall, and F1-score in IoT botnet attack detection. It proactively spots anomalies, mitigates vulnerabilities and fortifies device security, bolstering the resilience of interconnected IoT systems.

Table 1: Classification Report

Parameters	Accuracy	Precision	Recall	F1-Score
Values (%)	98.63	99.02	97.87	98.63

Our system determines an accuracy of 98.63%, signifying its effectiveness in suitably classifying instances. Furthermore, with a precision of 99.02%, it exhibits a high level of correctness in identifying botnet attacks, minimizing false positives. The recall of 97.87% highlights its ability to capture majority of authentic botnet instances, reducing false negatives. The F1-score, which combines precision and recall, underscores the overall effectiveness of our approach in detecting botnet attacks with a harmonized metric. These exceptional results validate the robustness and reliability of our machine learning-based solution for IoT botnet attack detection.

Fig 3: Home Page

Fig 4 presents a demonstration of benign network behaviours, showcasing typical patterns and communication protocols associated with normal user activity. The portrayed traffic exhibits regularity and conformity to expected norms, reflecting activities devoid of malicious intent.

Contrarily, Fig 5 depicts a demonstration of malicious network behaviours, categorized by anomalous patterns, atypical protocol usage and activities revealing of cyber threats. The observed traffic displays irregularities and potentially harmful actions, signifying the incidence of cyberattacks or unauthorized access attempts.

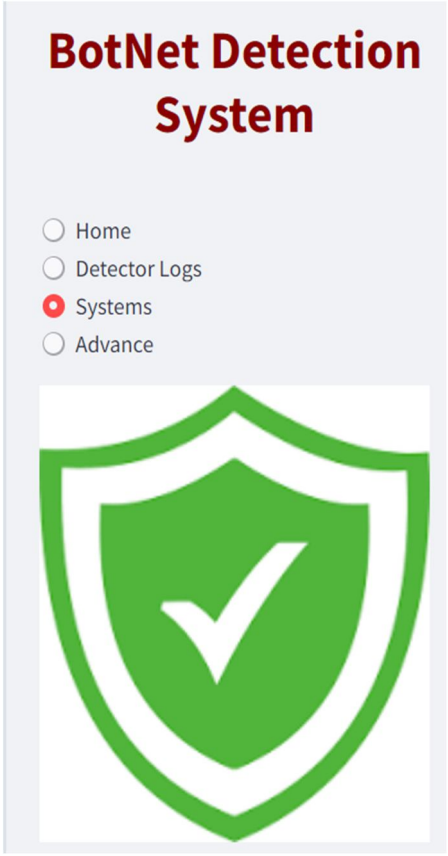


Fig 4: Benign Traffic Illustration

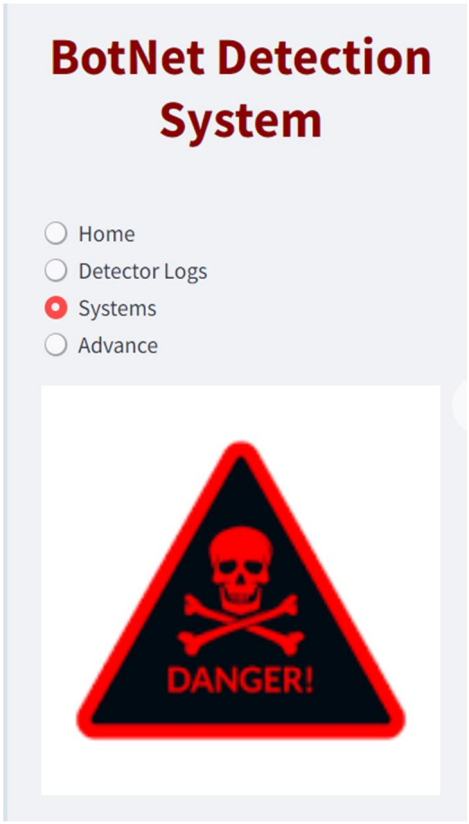


Fig 5: Malicious Traffic Illustration

VI. DISCUSSION

The achievement in detection of BotNet attacks through a hybrid ML approach represents a significant advancement in network security. This high accuracy underscores the effectiveness of combining multiple ML algorithms, leveraging their respective strengths to enhance detection capabilities. The hybrid approach likely enables comprehensive analysis of network traffic patterns, enabling the system to discern subtle anomalies indicative of BotNet activity with greater precision. However, while the results are promising, it is essential to consider potential limitations such as the generalizability of the model across diverse network environments and the need for ongoing refinement to adapt to evolving threat landscapes. Future research could focus on further optimizing the hybrid approach, exploring additional features or algorithms to expand detection accurateness and robustness in real-world network scenarios. Overall, this study pays valuable insights to the field of cybersecurity and emphasizes the potential of hybrid ML approaches in contending sophisticated cyber threats.

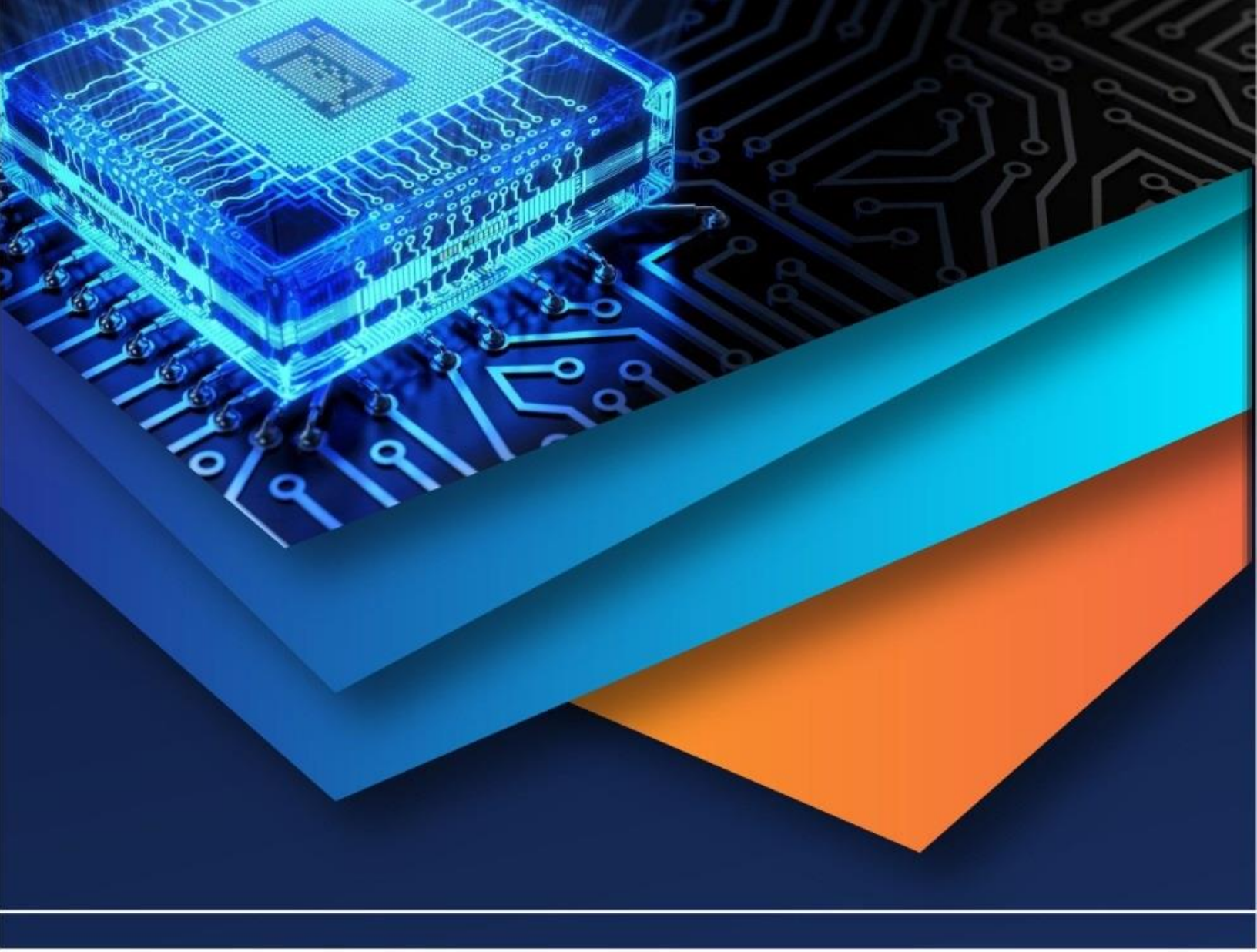
VII. CONCLUSION

In conclusion, the hybrid machine learning approach demonstrated remarkable efficacy in detecting BotNet attacks, achieving an impressive accuracy of 98.63%. This underscores the importance of leveraging diverse algorithms to boost network security. While the results are promising, enduring fine-tuning and revision are essential to discourse evolving threats. The study's findings contribute significantly to advancing cybersecurity measures, providing a framework for robust BotNet detection. Future research should focus on refining the hybrid approach and evaluating its applicability in diverse network environments. Overall, this study marks a substantial step forward in safeguarding networks against sophisticated cyber threats.

REFERENCES

[1] Sharma, P. B. Mishra and G. Geetha, "Botnet Attack Detection in IoT Networks using CNN and LSTM," 2023 2nd International Conference on Edge Computing and Applications (ICECAA), Namakkal, India, 2023, pp. 1270-1275, Doi: 10.1109/ICECAA58104.2023.10212330.
[2] A. Zaheer, S. Tahir, M. F. Almufareh and B. Hamid, "A Hybrid Model for Botnet Detection using Machine Learning," 2023 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2023, pp. 1-8, Doi: 10.1109/ICBATS57792.2023.10111161.
[3] M. A. Rachman Putra, T. Ahmad and D. P. Hostiadi, "Botnet Dataset Overview Using Statistical Approach Based on Time Gap Activity Analysis," 2023 11th International Symposium on Digital Forensics and Security (ISDFS), Chattanooga, TN, USA, 2023, pp. 1-6, Doi: 10.1109/ISDFS58141.2023.10131832.
[4] M. W. Nadeem, H. G. Goh, Y. Aun and V. Ponnusamy, "Detecting and Mitigating Botnet Attacks in Software-Defined Networks Using Deep Learning Techniques," in IEEE Access, vol. 11, pp. 49153-49171, 2023, Doi: 10.1109/ACCESS.2023.3277397.
[5] A. Ahmed and C. Tjortjis, "Machine Learning based IoT-BotNet Attack Detection Using Real-time Heterogeneous Data," 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 2022, pp. 1-6, Doi: 10.1109/ICECET55527.2022.9872817.
[6] V. Puri, A. Kataria, V. K. Solanki and S. Rani, "AI-based botnet attack classification and detection in IoT devices," 2022 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT), Soyapango, El Salvador, 2022, pp. 1-5, Doi: 10.1109/ICMLANT56191.2022.9996464.
[7] L. Le Jeune, T. Goedemé, and N. Mentens, “Machine learning for misuse-based network intrusion detection: Overview, unified evaluation and feature choice comparison framework,” Ieee Access, 2021.
[8] C. Kim, M. Jang, S. Seo, K. Park, and P. Kang, “Intrusion detection based on sequential information preserving log embedding methods and anomaly detection algorithms,” IEEE Access, vol. 9, pp. 58 088–58 101, 2021.
[9] I. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, and A. Gani, “Systematic literature review on iot-based botnet attack,” IEEE Access, 2020.
[10] S. Ghazanfar, F. Hussain, A. U. Rehman, U. U. Fayyaz, F. Shahzad, and G. A. Shah, “Iot-flock: An open-source framework for iot traffic generation,” in 2020 International Conference on Emerging Trends in Smart Technologies (ICETST). IEEE, 2020, pp. 1–6.
[11] A. Blaise, M. Bouet, V. Conan, and S. Secci, “Botnet fingerprinting: a frequency distributions scheme for lightweight bot detection,” IEEE Transactions on Network and Service Management, 2020.
[12] S. Sriram, R. Vinayakumar, M. Alazab, and K. Soman, “Network flow based iot botnet attack detection using deep learning,” in IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2020, pp. 189–194.
[13] B. Nugraha, A. Nambiar, and T. Bauschert, “Performance evaluation of botnet detection using deep learning techniques,” in 2020 11th International Conference on Network of the Future (NoF). IEEE, 2020, pp. 141–149.
[14] F. Hussain, S. G. Abbas, M. Husnain, U. U. Fayyaz, F. Shahzad, and G. A. Shah, “Iot dos and ddos attack detection using resnet,” in 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1–6.
[15] G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, “A hierarchical hybrid intrusion detection approach in iot scenarios,” in GLOBECOM 2020-2020 IEEE Global Communications Conference. IEEE, 2020, pp. 1–7.
[16] S. Gaonkar, N. F. Dessai, J. Costa, A. Borkar, S. Aswale, and P. Shetgaonkar, “A survey on botnet detection techniques,” in 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020, pp. 1–6.
[17] M. u Nisa and K. Kifayat, “Detection of slow port scanning attacks,” in 2020 International Conference on Cyber Warfare and Security (ICCWS). IEEE, 2020, pp. 1–7.
[18] F. Hussain, S. G. Abbas, U. U. Fayyaz, G. A. Shah, A. Toqeer, and A. Ali, “Towards a universal features set for iot botnet attacks detection,” in 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1–6.

[19] S. Maeda, A. Kanai, S. Tanimoto, T. Hatashima, and K. Ohkubo, "A botnet detection method on sdn using deep learning," in 2019 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2019, pp. 1–6.
[20] F. Tang, Y. Kawamoto, N. Kato, K. Yano, and Y. Suzuki, "Probe delay based adaptive port scanning for iot devices with private ip address behind net," IEEE Network, vol. 34, no. 2, pp. 195–201, 2019.
[21] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," in 2019 International Carnahan Conference on Security Technology (ICCST). IEEE, 2019, pp. 1–8.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY