International Journal of Computer Applications Technology and Research
Volume 15–Issue 06, 06 – 12, 2026, ISSN: -2319–8656
DOI:10.7753/IJCATR1506.1002

Intrusion Detection System Using Adaptive Machine Learning Approach

Chetan Negi
M.tech (CS)
Sushila Devi Bansal College of Technology
Indore, India

Pooja Hardiya
Assistant Professor
Sushila Devi Bansal College of Technology
Indore, India

Abstract: Each innovation brings with it a fresh set of problems to solve. The information stored on a system should be protected against access by anybody who has not been permitted to do so. Regarding system security, the first and most important task is to install and maintain a reliable and effective network intrusion detection system. If a structure has an interruption detection procedure in place, it may be able to estimate the total number of interruptions that will occur in the future and the present. The use of artificial intelligence approaches for organization-based interruption locations has been around for more than two decades already, and there are a variety of ways available. An effective interruption detection system will probably continue to be a topic of debate for a considerable amount of time. Companies must create better methods for keeping their systems and information secure from gatecrashers and hackers as the number of digital assaults and the bulk of system evidence continue to grow at an alarming pace. Due to the integration of more sophisticated security apparatuses into cutting-edge project designs, the volume of security events and ready information created continues to expand, making it more difficult to trace down the perpetrators of the assault as well as the gatecrashers. When it comes to managing the detection, reaction, and organization of security incidents and possible assaults on their systems, organizations are forced to depend on new ways to support and supplement human investigators for the first time. In particular, the emphasis of this Thesis is on differentiating between normal system traffic information and dangerous system traffic data. This study's objectives are to enhance the characteristics of generating system information using particle swarm optimization (PSO), and then to create a Network Intrusion Detection System using directed learning using a completely related Deep Neural Net (DNN) using directed learning (NIDS). It is feasible to develop sophisticated neural system models with the use of the NSL-KDD dataset that surpasses the constraints of the KDD Cup2009 interruption recognition datasets, which have been regularly utilized in the past. In experiments using the NSL-KDD datasets, it has been shown that deep neural networks with molecular swarm augmentation are very effective in terms of accuracy and recognition rates.

Keywords: Network Intrusion Detection System, NSL-KDD dataset, article Swarm Optimization, Deep Neural Networks, Artificial Intelligence.

1. INTRODUCTION

It is widely acknowledged that complete security is unattainable. Rather, the focus ought to be on risk mitigation and reducing the likelihood of an attack. Because it is impossible to foresee every scenario in which an attacker could circumvent the defences in place, security is an intractable problem. Attackers are always looking for new ways to get past defences; they change the way they attack and use methods they have never used before. Attacks of this kind can be extremely harmful and frustrating to the defence because the attacker has created an exploit that gets around the security of a particular system or software[1].

Zero-hour attacks, also referred to as zero-day attacks, present significant risks to both information and enterprise systems. Deploying Intrusion Detection Systems (IDS) is widely recognized as one of the most effective methods to safeguard the confidentiality, integrity, and availability of these systems after unauthorized access is gained. [2]IDS can be broadly categorized into two types: host-based and network-based. Host-based systems utilize tools such as host-based firewalls, antivirus/malware programs, data loss prevention agents, and monitoring system call trees to oversee and manage data originating from individual workstations. Conversely, network-based systems rely on various defenses like firewalls, proxy servers, intrusion detection systems, and antivirus software to monitor and regulate the flow of network traffic[3].

Among the pivotal technologies for ensuring computer network security are Network Intrusion Detection Systems (NIDSs), which play a vital role in enhancing the overall security posture of a network. However, to effectively mitigate threats, NIDSs should be integrated with other defense-in-depth strategies such as firewalls, antivirus software, access control mechanisms, and similar technologies. [4]This comprehensive approach enables cyber threat operations teams to promptly detect and respond to attacks, security incidents, and potential breaches.

The goal of this research is to leverage the most recent developments in deep learning technology to further enhance the performance of NIDSs. Network intrusion detection systems can be broadly divided into two categories: those that rely on anomalies or unusual behavior and those that rely on signatures or misuse. Signature-based systems send out alerts to users to let them

know when known misuse or unwanted behavior happens. Such a system usually uses techniques to discriminate between input events and known harmful intrusion signatures[5].

The system classifies an input event as malicious and flags it for additional investigation whenever it exhibits patterns resembling those of known malicious intrusions. These systems have the potential to be effective in identifying known harmful attacks and flagging them for additional investigation due to their low false positive rate. However, one drawback of these systems is that they are unable to identify fresh attacks. [3]Alarms in anomaly-based systems are triggered when observed events behave significantly differently from previously established known good patterns; this is when the system is considered troublesome. One of these systems' advantages over signature-based systems is that these systems can recognize novel and evolving threats, whereas signature-based systems cannot. According to the definition, an anomaly is anything that deviates from what is thought to be typical, normalized, or predicted in a given scenario. [4]Anomalies are defined as rare departures from a system's expected behavior that are referred to as anomalous behavior. An anomaly detection system must first identify any event or sequence that deviates from a predetermined set of typical behaviors in order for the system to classify it as abnormal. Understanding that not all deviations in nature have malevolent intentions is crucial.[6] Anomalies are precisely what they are described as deviations from expected normal behavior, according to the definition of the term. When an anomalous event or pattern is identified, it can be categorized as either benign or malevolent based on the surrounding circumstances. One of the most difficult problems in computer science is producing both a high rate of false positives and false negatives, which is also a difficult problem when it comes to anomaly-based systems. It is crucial to understand that not every deviation in the natural world has malevolent intentions. Anomalies are exactly that—differences from expected normal behavior—according to the definition of the term. Depending on the surrounding circumstances, an anomalous event or pattern can be classified as either benign or malevolent as soon as it is identified. [7]The problem of producing a high rate of false positives and false negatives, which is also one of the most difficult problems in computer science, is one of the most difficult problems when it comes to anomaly-based systems.

2. LITERATURE REVIEW

This section conveys the state-of-the-art research in the field of "Interruption Detection Systems" and also includes a synopsis of some of the examination papers we have considered. Finding and preventing interruptions is becoming a major research project in the field of Internet security. Innovation has also led to an increase in hacking and device abuse, as well as the use of better disruption techniques, which increases the risk of intrusion and organization of security. [8], [9]The study used machine learning algorithms that were already in use to carry out a thorough attack detection process on the UNSW-NB15 and NSL-KDD datasets. Algorithms for machine learning were applied to the detection and classification of attacks. We assessed the machine learning algorithms' accuracy using the NSL-KDD and UNSW-NB15 datasets. Accuracy for two-class and multi-class on the UNSW-NB15 dataset was found to be 98.6% and 98.3%, respectively. Accuracy for the NSL-KDD dataset was 93.4% and 97.8%.[10]

Using three widely-used datasets (KDD 99, UNSW-NB15, and CSE-CIC-IDS 2018), the study assesses the effectiveness of different machine learning algorithms for binary and multi-class classification. The experimental results show promising overall classification performance and high binary classification accuracy rates.[11]

The study presents a novel ensemble approach that combines the XGBoost, decision tree, random forest, and extra tree algorithms for network intrusion detection. Python was used to implement the recommended strategy, which resulted in increased detection accuracy.[12]

The study presents a novel ensemble approach that combines the XGBoost, decision tree, random forest, and extra tree algorithms for intrusion detection in network systems. When tested on the CICIDS2017 dataset, the recommended approach, which is implemented using the Python programming language, demonstrated improved detection accuracy.[13]

The study uses a newly released SDN dataset (SDN Intrusion) that is accessible to the public to identify intrusions within SDN/NFV networks using machine learning techniques.[14]

The study offers a novel method for locating network system intrusions using machine learning techniques. The model's results show that it performs more accurately than other approaches, like Naive Bayes. The suggested approach produced a 1.26-minute performance time, a 97.38% accuracy rate, and a 0.25% error rate.[15]

This article's author talked about and considered the framework he was using at the time. Interruption Al recognition is one of the security innovations that protect the framework from malicious activity by screening it. This paragraph goes into great detail about the Intrusion Detection System and Internal Intrusion Detection System, which includes a diagram and requires multiple computations for the framework to function.[16]

Information mining techniques are used to develop strategies for digital inquiry based on the notion of interruption recognition. The overview states that the proposed work improves the exactness and identification rate by up to 94% when compared to prior IDS. When designing a new IDS system, this attribute set, according to its creator, may be used to distinguish between insiders and their malicious behaviors. A few businesses may use it to safeguard their sensitive data since it will be a legitimate intrusion detection system that reliably and continuously distinguishes between inner gatecrashers. [17]

There are many more ways to attack a system or framework because personal computers are widely used and have easy access to the internet. Interrupting someone is a crime that involves entering a building, breaking someone's rights, or claiming someone else's structure or

International Journal of Computer Applications Technology and Research
Volume 15–Issue 06, 06 – 12, 2026, ISSN: -2319–8656
DOI:10.7753/IJCATR1506.1002

resources. The separation of data infrastructure attacks from other forms of attacks is the main goal of creating an interruption location framework. It's a security technique that aims to distinguish between different kinds of attacks.[18]

Firewalls are useless for defending the system from any kind of attack because they can only identify intrusions that originate from outside the system. This study developed a framework for abuse-based interruption location evaluation. The review process for this framework was the same as that of ALAD, PHAD, LERAD, NETAD, and other quantifiable computations based on abnormalities. These days, PC organizations are developing at a faster rate than ever before, and arrangement security is the most amazing tool available. [19]

In this work, we select and identify relevant features from the NSL-KDD dataset [1] to improve the KDD dataset. The NSL-KDD dataset has fewer duplicate and null values than the KDD dataset. The primary objectives are outlined as follows: [20], [21]

1) To investigate the possible applications of machine learning in network intrusion detection.

2) Implementing the Random Forest and Support Vector Machine algorithms to create an intrusion detection system.

3) To select features using Random Forest and SVM using Recursive structure feature Elimination, which will speed up computation and improve accuracy.

4) To compare and illustrate these algorithms' recall, accuracy, and precision.

3. PROBLEM DOMAIN

Real-time attackers who operated without the administrator's knowledge or presence were beyond the detection range of earlier techniques like intrusion detection systems and firewalls.

• The current system employed basic machine learning techniques, which were insufficient to effectively secure the system and combat the new threats.

Real-time monitoring and response features can be added to the system, enabling quick action to be taken in the event of an intrusion.[22]

An investigation into the application of ensemble methods, which integrate several machine learning models, may be conducted to raise the intrusion detection system's overall accuracy and dependability.

By regularly updating the database of known attacks and modifying the machine learning models accordingly, the system can be made more capable of handling dynamic and evolving threats.

Both the training and testing modes are functional for the suggested system. Initially, the features of trusted and malicious network nodes are extracted.

4. MODEL CREATION

Every website, malicious or benign, produces a significant amount of data in the field of website analysis. In the past, limitations in computing power and analytical methods have made it difficult to conduct a thorough analysis of this data. However, managing and interpreting huge datasets has become much easier thanks to recent developments in computer science, data analysis, and machine learning. The focus of this paper is on analyzing data generated by websites across the internet using supervised machine learning techniques. The goal is to develop a predictive model that can distinguish between benign and malicious websites.

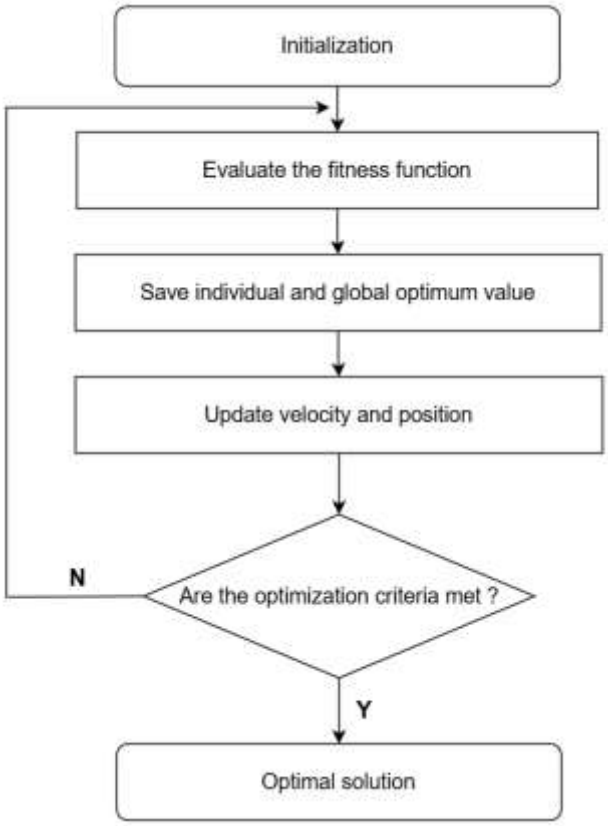


Figure 1. Particle Swarm Optimization Flowchart

Confusion matrix: A confusion matrix can be used to demonstrate the efficacy of a supervised learning classification method. This kind of matrix is depicted in Figure 1. The projected class is represented by the columns, and the ground truth is represented by the rows. The following definition of network intrusion detection metrics applies to network intrusion detection [8].

True Positive Rate (TPR) or Recall: TPR =

$Recall$ = TP/TP+FN

International Journal of Computer Applications Technology and Research
Volume 15–Issue 06, 06 – 12, 2026, ISSN: -2319–8656
DOI:10.7753/IJCATR1506.1002

		PREDICTED	
		Benign	**Mallcious**
ACTUAL	**Benign**	**True Negative** Predicted benign, target was benign	**False Positive** Predicted mallcious, target was mallcious
	Mallcious	**False Negative** Predicted benign, target was mallcious	**True Positive** Predicted mallcious, target was mallcious

Note that in other studies, the terms "Detection Rate" and "True Positive Rate," also referred to as "Recall," are used interchangeably. Consequently, when comparing the findings of this work to those of these other studies, the term "Detection Rate," which is interchangeable with "True Positive Rate," is used.

Data understanding

NSL-KDD It includes information about network connections that was gathered from a fictitious network setting. This data contains a variety of network traffic-related characteristics, including source and destination addresses, lengths of connections, protocol and service types, and more. An enhancement of the KDD'99 dataset is the NSL-KDD.

Intrusion Detection Systems (IDS) are systems that are trained on internet traffic record data to identify malicious traffic inputs. The standard for contemporary internet traffic is the NSL-KDD. indicative of actual networks that are in place.

Data preparation

We will clean up and get the data ready for additional analysis in this section. The main focus will be on handling outliers, incomplete data, and missing values. We need to prepare the data we receive for the project because it is unfit for analysis.

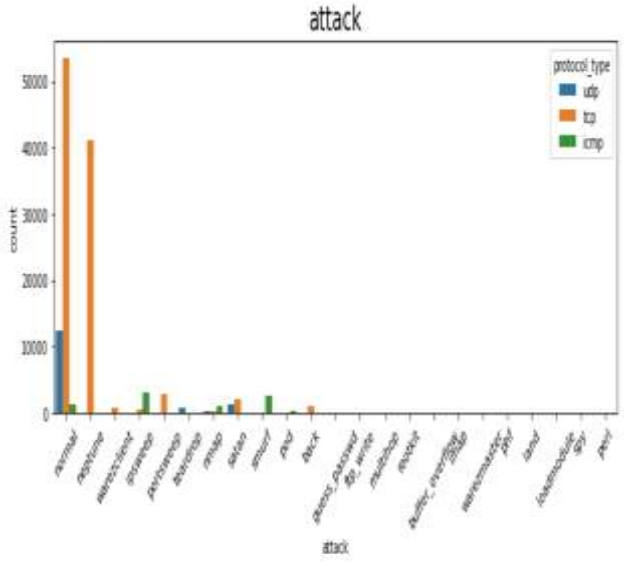
But plotting a correlation plot is actually clearer

Understanding a correlation plot involves creating a relationship between two variables and determining whether it is inversely proportional (less than -0.5), proportional (greater than 0.5), or not related at all (near zero).

As adding irrelevant columns will skew or distort the model, knowing how the columns correlate will help you create a better one.

This is what we discovered: Features linked to connections have a strong correlation (rates).

Additionally, highly correlated are host-based traffic features (dsts).



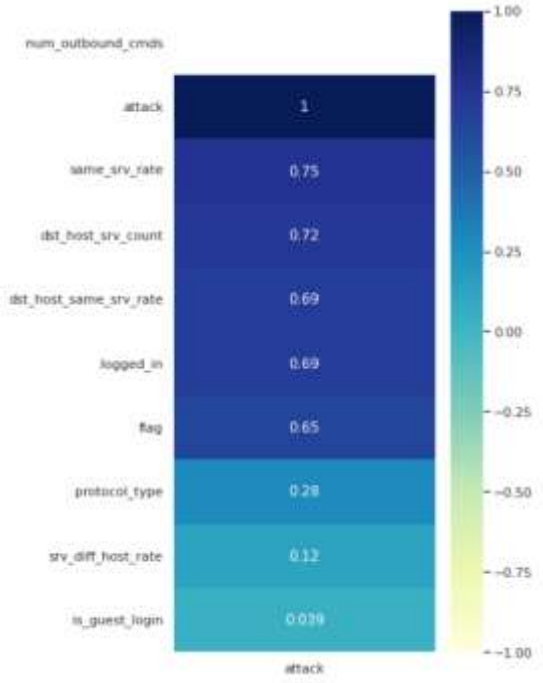
Missing Data Analysis

Duration is a crucial column because its indications are somewhat significant, but it contains a lot of zeros, and it isn't acceptable for the duration to be zero. As a result of these outliers, the model runs very slowly.

We have three options for resolving this: either remove every zero row from the duration column, which will result in a significant loss of data.

or, to ensure that the data is not lost, we substitute the median or a value that is nearly equal to zero for the zero.

International Journal of Computer Applications Technology and Research
Volume 15–Issue 06, 06 – 12, 2026, ISSN: -2319–8656
DOI:10.7753/IJCATR1506.1002



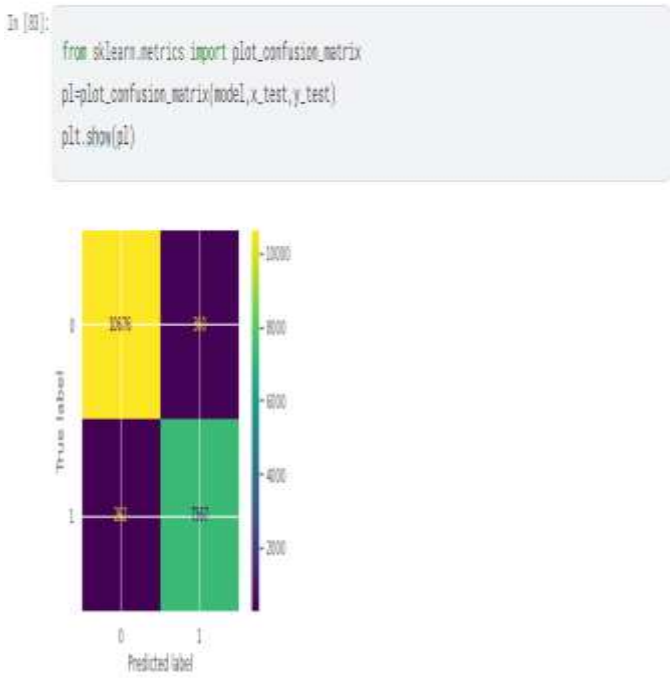


Figure : Missing value proportions across the dataset

Modeling

We will begin the modeling phase by doing an exploratory analysis of the data. This means that to determine which columns are related to the target column, we will utilize graphs and charts. We will be able to select the pertinent columns for our model as a result. Following the completion of the exploratory analysis, a classifier will be trained using a range of machine-learning techniques. Here are a few algorithms we can attempt:

Logistic Regression

Decision Trees

Random Forest

Support Vector Machines (SVM)

Once all models have been trained, we will choose the best one using the evaluation method.

Among the five models we trained, the Random Forest algorithm emerged as the most effective. Several reasons support this conclusion: Firstly, the logistic regression model's reliance on the linear nature of predictor variables can be limited by their distributions, potentially overlooking non-linear relationships between predictors and the target (Sperandei, 2022). Secondly, decision trees, the building blocks of random forests, are inherently simplistic (Rokach and Maimon, 2018). Additionally, support vector machines, another linear model, can struggle to establish a single decision boundary in high-dimensional spaces (Shihong, Ping, and Peiyi, 2021). Furthermore, ensemble methods like random forests have consistently exhibited robust performance across various classification tasks (Fawagreh, Gaber, and Elyan, 2020). Our best F1 score on the test set reached 0.78, indicating strong performance but leaving room for improvement. Given that only 12% of the websites in our dataset were malicious, addressing class imbalance through sampling techniques could enhance model performance. By oversampling malicious websites, the model can better learn patterns indicative of malignancy (Guo et al., 2008). Employing advanced training algorithms such as gradient boosting could also yield further improvements."

5. CONCLUSION

To construct a classification model aimed at predicting whether a website is malicious, we systematically prepared and cleaned a dataset comprising both benign and malicious websites. Our analysis indicates that machine learning models possess a notable capability to discern malicious websites, as demonstrated by the detection of ninety out of seventy-five malicious websites within the test set. Furthermore, we've identified essential characteristics crucial for

forecasting the probability of a website's malicious nature. Notably, several of these critical features—including URL length, frequency of special characters, geographical origin, and age of the website—are readily accessible. We anticipate further improvements in our findings by leveraging more intricate models and employing sampling strategies to address class imbalance.

6. REFERENCES

[1] [1] P. Verma *et al.*, "A novel intrusion detection approach using machine learning ensemble for iot environments," *Applied Sciences (Switzerland)*, vol. 11, no. 21, Nov. 2021, doi: 10.3390/app112110268.

[2] [2] J. Asharf, N. Moustafa, H. Khurshid, E. Debie, W. Haider, and A. Wahab, "A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions," *Electronics (Switzerland)*, vol. 9, no. 7. MDPI AG, Jul. 01, 2020. doi: 10.3390/electronics9071177.

[3] [3] J. Zuniga-Mejia, R. Villalpando-Hernandez, C. Vargas-Rosales, and A. Spanias, "A Linear Systems Perspective on Intrusion Detection for Routing in Reconfigurable Wireless Networks," *IEEE Access*, vol. 7, pp. 60486–60500, 2019, doi: 10.1109/ACCESS.2019.2915936.

[4] [4] A. Almotairi, S. Atawneh, O. A. Khashan, and N. M. Khafajah, "Enhancing intrusion detection in IoT networks using machine learning-based feature selection and ensemble models," Systems Science & Control Engineering, vol. 12, no. 1, Dec. 2024, doi: 10.1080/21642583.2024.2321381.

[5] [5] S. Ahmad and A. Tamimi, "Detecting Malicious Websites Using Machine Learning." [Online]. Available: https://scholarworks.rit.edu/theses

[6] [6] P. Dini, A. Elhanashi, A. Begni, S. Saponara, Q. Zheng, and K. Gasmi, "Overview on Intrusion Detection Systems Design Exploiting Machine Learning for Networking Cybersecurity," *Applied Sciences (Switzerland)*, vol. 13, no. 13. Multidisciplinary Digital Publishing Institute (MDPI), Jul. 01, 2023. doi: 10.3390/app13137507.

[7] [7] S. Soner, R. Litoriya, R. Khatri, A. A. Hussain, S. Pagare, and S. K. Kushwaha, "Real-Time Face Mask Detection in Mass Gathering to reduce COVID-19 Spread," *Journal of Automation, Mobile Robotics and Intelligent Systems*, pp. 51–58, Dec. 2023, doi: 10.14313/jamris/1-2023/7.

[8] [8] A. Bhaskar Abhale and S. S. Manivannan, "Deep Learning Algorithmic Approach for Operational Anomaly Based Intrusion Detection System in Wireless Sensor Networks," 2021, doi: 10.21203/rs.3.rs-777010/v1.

[9] [9] F. TÜRK, "Analysis of Intrusion Detection Systems in UNSW-NB15 and NSL-KDD Datasets with Machine Learning Algorithms," *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, vol. 12, no. 2, pp. 465–477, Jun. 2023, doi: 10.17798/bitlisfen.1240469.

[10] K.-A. Tait *et al.*, "Intrusion Detection using Machine Learning Techniques: An Experimental Comparison," May 2021, [Online]. Available: http://arxiv.org/abs/2105.13435

[11] M. Verma, A. Kumar Choudhary, S. Pagare, R. Shukla, and M. Shrivastava, "System to Design a Baseless and Wireless Mouse using Different Sensors," 2022. [Online]. Available: www.ijfans.org

[12] J. Ren, J. Guo, W. Qian, H. Yuan, X. Hao, and H. Jingjing, "Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms," *Security and Communication Networks*, vol. 2019, 2019, doi: 10.1155/2019/7130868.

[13] V. Saikushwanth and G. R. Rao, "Intrusion Detecction System Using Machine Learning," 2021.

[14] A. Shankar, R. Shetty, and B. Nath, "A Review on Phishing Attacks," 2019. [Online]. Available: http://www.ripublication.com

[15] J. Chhikara, R. Dahiya, N. Garg, and M. Rani, "Phishing & Anti-Phishing Techniques: Case Study," 2013. [Online]. Available: https://www.researchgate.net/publication/263773425

[16] Dr. A. Yadav, "PHISHING IN INDIA – ANALYTICAL STUDY," *IARJSET*, vol. 8, no. 8, Aug. 2021, doi: 10.17148/iarjset.2021.88110.

[17] G. Karatas, O. Demir, and O. K. Sahingoz, "Deep Learning in Intrusion Detection Systems," in *International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 113–116. doi: 10.1109/IBIGDELFT.2018.8625278.

[18] R. Patgiri, U. Varshney, T. Akutota, and R. Kunde, "An Investigation on Intrusion Detection System Using Machine Learning," in *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence, SSCI 2018*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 1684–1691. doi: 10.1109/SSCI.2018.8628676.

[19] E. E. Abdallah, W. Eleisah, and A. F. Otoom, "Intrusion Detection Systems using Supervised Machine Learning Techniques: A survey," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 205–212. doi: 10.1016/j.procs.2022.03.029.

[20] H. Sadreazami, A. Mohammadi, A. Asif, and K. N. Plataniotis, "Distributed-Graph-Based Statistical Approach for Intrusion Detection in Cyber-Physical Systems," *IEEE Trans Signal Inf Process Netw*, vol. 4, no. 1, pp. 137–147, Mar. 2018, doi: 10.1109/TSIPN.2017.2749976.

[21] [21] G. Palaniappan, S. Sangeetha, B. Rajendran, Sanjay, S. Goyal, and B. S. Bindhumadhava, "Malicious Domain Detection Using Machine Learning on Domain Name Features, Host-Based Features and Web-Based Features," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 654–661. doi: 10.1016/j.procs.2020.04.071.

[22] [22] H. Xuan and M. Manohar, "Intrusion Detection System with Machine Learning and Multiple Datasets." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html