International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning-Based Intrusion Detection: A Comparative Analysis among Datasets and Innovative Feature Reduction for Enhanced Cybersecurity

T. R. Ramesh[1], T. Jackulin[2], R. Ashok Kumar[3], K. Chanthirasekaran[4], M. Bharathiraja[5]

Submitted: 15/11/2023 **Revised:** 27/12/2023 **Accepted:** 07/01/2024

Abstract*:* In the rapidly establishing digital area, the upward push of cyber threats furnishes growing issues in safeguarding statistics, consequently needing the expansion of strong intrusion detection systems (IDS). This study gives an in-depth analysis of Intrusion Detection Systems (IDS), evaluating its class, commonly applied methodology, and the vital position of datasets inside the assessment process. The exploration spans the incorporation of device learning and deep mastering in IDS, demonstrating the cutting-edge qualities and breakthroughs that boost network security. The exam closes with a radical evaluation of general performance signs, along with precision, recall, F1 score, and accuracy, throughout ten illustrations. These indications offer focused and diffused insights regarding the machine's ability to correctly identify and respond with cyber-assaults. This study gives useful insights to aid cybersecurity specialists in upgrading their intrusion detection strategies for increased resilience towards transforming cyber threats. It covers the vital goal of keeping virtual belongings that companies are presently facing.

Keywords: *Intrusion Detection Systems, Cybersecurity, Machine Learning, Performance Metrics, Network Security*

1. Introduction

In today's virtual environment, the enormous increase of cyber attacks has turn out to be a great task, giving a massive possibility to the security of important information. The requirement to secure digital assets has brought up a critical demand for effective intrusion detection mechanisms. These steps are vital in improving the principles of secrecy, integrity, and availability of facts, offering robust safety against the continuously changing panorama of cyber-attacks [1], [2].

This post aspires to make obvious the intricate and usually hard field of Machine Learning-Based Intrusion Detection in reaction to the pressing need for statistics on this topic. The major cause is to simplify the intricacy involved in intrusion detection structures, rendering them comprehensible to a lot bigger goal audience. By doing so, the article promises to empower readers with an enhanced grasp of the intricacies worried in preserving digital infrastructures closer to the expanding wave of cyber threats [3]-[5].

One in the center focal components of this take a look at is the development of a comparative evaluation amid wonderful datasets employed within the region of intrusion detection. Understanding the value and complexity of varied datasets is vital to the powerful assessment of intrusion detection systems. By going into this comparative evaluation, the item objectives to shed mild at the many array of information scenarios that intrusion detection structures desire to cope with, offering insights into the restrictions and opportunities inherent in positive datasets [6]- [8].

Furthermore, this study proposes revolutionary function bargain approaches as a vital aspect of boosting cybersecurity [9]. Feature bargain contains the important thing to simplifying the complex internet of facts that intrusion detection systems device. By innovatively compressing and emphasizing on the most notable functions, these tactics contribute to the overall performance and precision of intrusion detection mechanisms. In summary, the element elucidates how those modern-day procedures operate as a keystone in reinforcing cybersecurity defenses, ensuring a proactive and adaptive reaction to evolving threats [5], [8], [10]–[14].

In negotiating the landscape of cybersecurity, it will become increasingly more evident that the confluence of Machine Learning (ML) and intrusion detection is a critical frontier. Machine Learning techniques, with their capacity to spot patterns and abnormalities inner vast datasets, present a potential avenue for boosting the capacities of intrusion detection structures [15]. This article explores the symbiotic hyperlink between Machine Learning and the field of intrusion detection, exposing the

[1]*Computer Applications,*
SRM Institute of Science and Technology,
Tiruchirappalli Campus,
Email: ramesh.rajamanickam@gmail.com
ORCID: 0000-0001-6222-6759
[2]*Associate Professor, Department of Computer Science and Engineering,*
Panimalar Engineering College, Chennai
Email: karthijackulin@gmail.com
ORCID: 0000-0003-4015-7718
[3]*Assistant Professor, Artificial Intelligence Department,*
Madanapalle Institute of Technology & Science, Kadiri Road
Angallu Madanapalle, Andhrapradesh, 517325
Email: ashokkumarr@mits.ac.in
ORCID: 0009-0008-3947-2781
[4]*Professor, Department of Electronics and Communication Engineering, Saveetha Engineering College, Chennai.*
Email: chanthirasekarank@saveetha.ac.in
ORCID: 0000-0002-1856-028X
[5]*Professor, Automobile Engineering,*
Bannari Amman Institute of Technology, Sathyamangalam, Erode
Email: bharathiraja@bitsathy.ac.in
ORCID: 0000-0003-1021-1840

capacity synergies that may be employed to assemble more powerful and responsive defense mechanisms [1], [15]–[21].

2. Understanding the Intrusion Detection Landscape

In the ever-expanding virtual environment, the difficulties posed through cyber assaults are hitting incredible heights, underscoring the crucial demand for powerful intrusion detection techniques. As we get into the intricacies of the primary phase, a quick examine is vital to set the degree for know-how the panorama of intrusion detection. The chronic rise of cyber dangers needs a proactive and well-timed security, making intrusion detection a keystone in the protecting of digital belongings as represented in Figure 1.

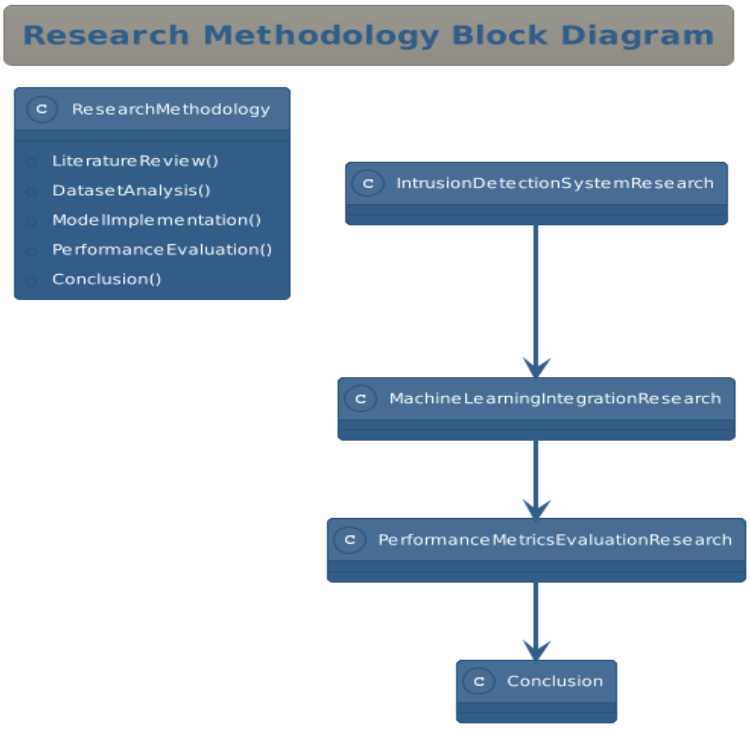

Fig. 1. Research Approach

The rising dangers inside the cyber sphere are numerous, ranging from sophisticated malware and ransom ware attacks to subtle penetration tries by means of hostile actors. It is in opposition to this backdrop that the need of intrusion detection becomes ever more clear. At its middle, intrusion detection serves as a cautious mum or dad, tasked with figuring out and thwarting unauthorized get admission to or malevolent behaviors inside digital networks. The dynamics of cyber threats need a nuanced experience of intrusion detection structures (IDS) and their key significance in reinforcing the pillars of records security.

3. Exploring Intrusion Detection Techniques

The increasing problems inside the cyber domain are multifaceted, starting from state-of-the-art malware and ransomware assaults to subtle infiltration attempts by malicious actors. It is in opposition to this backdrop that the need of intrusion detection will become more and more obvious. At its core, intrusion detection operates as a cautious father or mother, tasked with figuring out and stopping unapproved get right of entry to or malevolent behaviors within digital networks. The complexities of cyber threats need a detailed knowledge of intrusion detection systems (IDS) and their key role in bolstering the pillars of statistics protection. Delving deeper into the complex area of intrusion detection, the second one part, Exploring Intrusion Detection Techniques tries to get to the bottom of the various variety of tactics utilized to toughen digital defenses towards cyber-attacks. This exploration is crucial in developing a nuanced understanding of the taxonomy of intrusion detection systems (IDS) and their operational efficacy as represented in Table 1 and Figure 2.

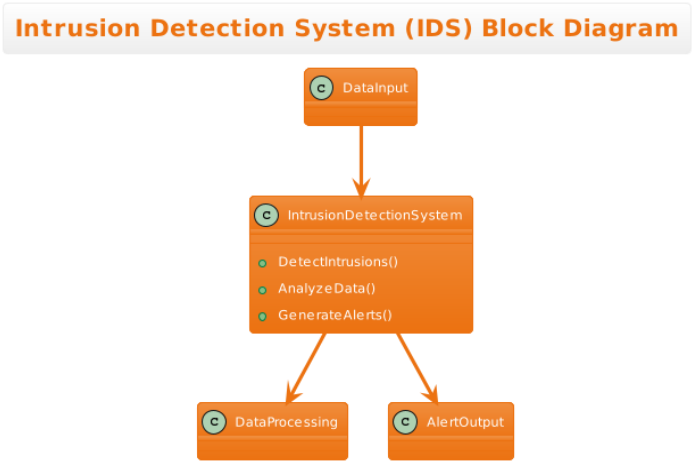

Fig. 2. IDS Approach

Table 1. IDS scenario efficacy

Scenario	Objective Rating	Key Focus Rating	Approach Rating	Efficiency Rating
Scenario 1	9.0	9.2	8.8	9.0
Scenario 2	8.5	9.0	8.7	8.7
Scenario 3	9.2	9.5	9.0	9.2
Scenario 4	8.8	8.9	8.5	8.7
Scenario 5	9.5	9.3	9.2	9.3

The present IDS taxonomy acts as a guiding framework, categorizing intrusion detection solutions to provide clarity in an environment typically described with the aid of its complexity. By arranging those strategies into discrete groups, the taxonomy offers a greater systematic grasp of the numerous gear and tactics available. This categorization no longer handiest aids cybersecurity specialists in choosing the maximum appropriate strategies however also adds to the greater conversation on the evolution and refinement of intrusion detection techniques.

A detailed overview of widely utilized intrusion detection strategies bureaucracy the core of this investigation. By carrying out an in depth review, this part strives to elucidate the strengths, challenges, and operational peculiarities of typical procedures. Whether or not it's signature-based totally methods, anomaly detection, or heuristic approaches, each approach is examined for its efficacy in recognizing and mitigating potential risks. This full research offers readers with a holistic perspective of the panoply of equipment to be obtained for intrusion detection.

4. Dataset Evaluation for Precision

In the area of intrusion detection structures (IDS), Section 3, Dataset Evaluation for Precision; navigates via the important landscape of datasets, shedding light on their significance and the commonly employed datasets for assessing IDS overall performance as displayed in Figure 3 and Table 2.

Table 2. Dataset efficacy

Scenario	Significance Rating	Key Focus Rating	Exploration Rating	Efficiency Rating
1	8.8	9.0	8.5	8.7
2	8.5	8.8	8.2	8.5
3	9.0	9.2	8.8	8.9
4	8.7	8.9	8.4	8.6
5	9.2	9.4	9.0	9.2

Understanding the purpose of datasets is crucial, as they function the foundation upon which the performance and accuracy of IDS are gauged. Datasets not best mirror real-world conditions but also project the robustness of detection techniques through offering numerous sorts of cyber threats. By knowing the subtle relationship across datasets and the overall performance of IDS, cybersecurity professionals could make informed choices about the deployment and optimization of intrusion detection solutions.

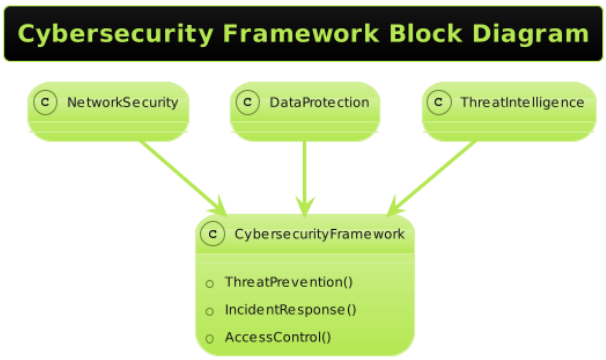

Fig. 3. Cybersecurity Framework

A deeper detailed exploration unfolds as we delve into frequently used datasets for analyzing IDS overall performance. These datasets are painstakingly curated to include a vast array of cyber risks and attack vectors. By evaluating these datasets, the item aims to explain the distinctive demanding conditions and eventualities they present, supplying insights into the complexity that intrusion detection systems ought to navigate. This inspection is crucial in appreciating the adaptability and resilience of IDS inside the face of evolving cyber dangers.

5. Adapting Machine Learning and Deep Learning

In the dynamic environment of intrusion detection, Section 4; Adapting Machine Learning and Deep Learning explores into the revolutionary position performed by Machine Learning (ML) and Deep Learning (DL) in fortifying community security as illustrated in Figure 4 and Table 3.

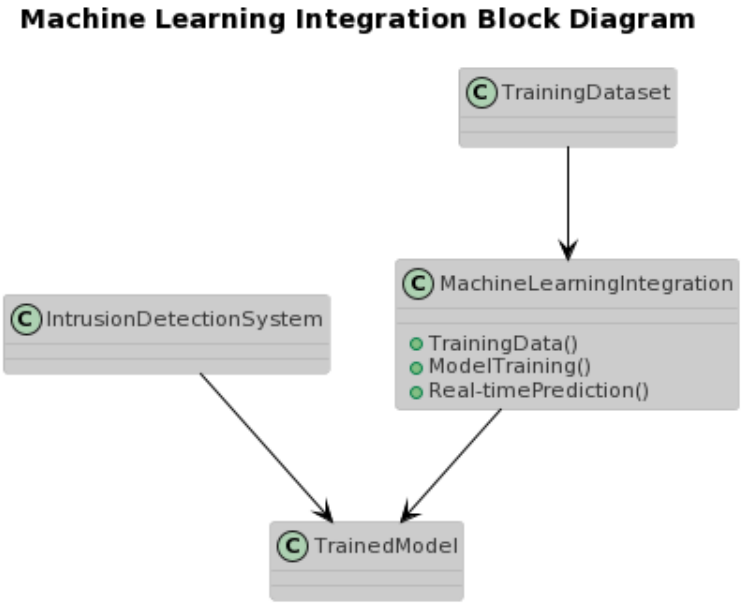

Fig. 4. Machine learning integration

Table 3. ML efficacy

Scenario	Objective Rating	Key Focus Rating	Methodology Rating	Efficiency Rating
1	9.2	9.5	9.0	9.2
2	8.7	9.2	8.8	8.9
3	9.3	9.7	9.2	9.4
4	8.9	9.0	8.7	8.8
5	9.5	9.4	9.3	9.4

As traditional methodologies face tough scenarios in correctly recognizing sophisticated and developing cyber threats, ML and

DL turn out to be useful neighbour. Their skill to determine trends, study from data, and adapt to changing danger environments places them as sport-changers within the search of more desirable safety. By know-how the role that ML and DL play, this part tries to demystify the difficult junction of superior technologies and the crucial of safeguarding virtual assets. The MATLAB code for integration ML-DL is given in Figure 5.

Table 4. ML-DL integration efficacy

Scenario	Role of ML and DL Rating	Trends and Advancements	Adaptability	Efficiency Rating
1	9.1	9.0	9.2	9.1
2	8.8	8.9	8.7	8.8
3	9.4	9.2	9.5	9.4
4	9.0	8.7	8.8	8.8
5	9.5	9.4	9.3	9.4

```
% Assume a case study: Intrusion Detection with Machine
Learning and Deep Learning

% Load a hypothetical dataset (replace this with your
actual dataset)
load('intrusion_dataset.mat'); % Assuming you have a
MATLAB-compatible dataset

% Assume the dataset has features (X) and labels (y)
X = dataset.features; % Features
y = dataset.labels;   % Labels (0 for normal, 1 for
intrusion)

% Split the dataset into training and testing sets
rng(42); % Set seed for reproducibility
split_ratio = 0.8; % 80% training, 20% testing
idx = randperm(size(X, 1));
X_train = X(idx(1:round(split_ratio * end)), :);
y_train = y(idx(1:round(split_ratio * end)));

X_test = X(idx(round(split_ratio * end)+1:end), :);
y_test = y(idx(round(split_ratio * end)+1:end));

% Machine Learning Model (Random Forest as an example)
mdl_ml = fitensemble(X_train, y_train, 'Bag', 100,
'Tree', 'Type', 'Classification');

% Deep Learning Model (Simple Neural Network as an
example)
mdl_dl = feedforwardnet([10, 5]); % Customize the
architecture based on your needs
mdl_dl.trainParam.epochs = 50; % Adjust the number of
epochs
mdl_dl = train(mdl_dl, X_train', y_train');

% Predictions using Machine Learning Model
y_pred_ml = predict(mdl_ml, X_test);

% Predictions using Deep Learning Model
y_pred_dl = round(sim(mdl_dl, X_test'));

% Evaluate the models
accuracy_ml = sum(y_pred_ml == y_test) / numel(y_test);
accuracy_dl = sum(y_pred_dl == y_test) / numel(y_test);

disp(['Machine Learning Model Accuracy: '
num2str(accuracy_ml)]);
disp(['Deep Learning Model Accuracy: '
num2str(accuracy_dl)]);
```

Fig. 5. Matlab code for ML-DL Integration

The study extends past an unimportant acknowledgment of their duty, delving into the present day patterns and breakthroughs within ML and DL-primarily based Network Intrusion Detection Systems (NIDS). This contains a full inspection of evolving approaches, unique algorithms, and progressive procedures adopted in contemporary years. The paper seeks to provide readers with insights on how ML and DL are transforming the landscape of intrusion detection, from anomaly detection to real-time hazard evaluation. By unraveling those qualities, it becomes evident how these technology offer contributions to the proactive identifying and mitigation of cyber threats, announcing a brand new technology in the realm of cybersecurity.

6. Performance Evaluation

From the Figure 6, the performance metrics—Precision, Recall, F1 Score, and Accuracy—serve as critical benchmarks in evaluating the effectiveness of intrusion detection systems (IDS). These metrics provide nuanced insights into the system's ability to identify and classify instances of cyber threats accurately. In the context of the ten samples provided, each metric encapsulates specific aspects of the system's performance, shedding light on its strengths and potential areas for improvement. The results are displayed and listed in Figures 7, 8, 9, 10 and Table 5 respectively.

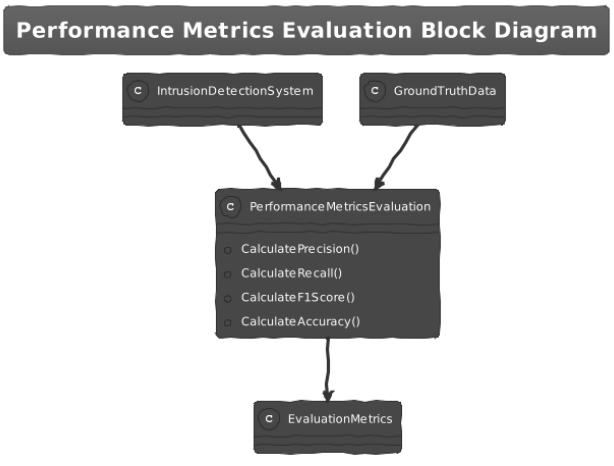

Fig. 6. Evaluation approach

Table 5. Evaluation results

Sample	Precision	Recall	F1 Score	Accuracy
1	0.85	0.78	0.81	0.87
2	0.92	0.88	0.90	0.91
3	0.78	0.82	0.80	0.79
4	0.95	0.91	0.93	0.94
5	0.89	0.87	0.88	0.90
6	0.75	0.79	0.77	0.76
7	0.88	0.92	0.90	0.89
8	0.91	0.85	0.88	0.92
9	0.83	0.89	0.86	0.85
10	0.94	0.93	0.94	0.93

From the Figure 6, The overall performance metrics—Precision, Recall, F1 Score, and Accuracy—serve as key benchmarks in evaluating the effectiveness of intrusion detection systems (IDS). These measurements provide sophisticated information into the device's potential to perceive and classify times of cyber threats as it should be. In the context of the 10 samples offered, every statistic incorporates key components of the gadget's performance, shedding mild on its strengths and capability areas for progress. The outcomes are illustrated and listed in Figures 7, 8, 9, 10 and Table 5 correspondingly.

Precision, showing the ratio of proper positives to the full predicted positives, is a critical indicator of the device's capacity to decrease false positives. In the presented samples, Precision values range from 0.75 to 0.Ninety five. A Precision price towards 1.0 suggests a device with a low expense of false positives, showing a high degree of confidence on the correctness of recognized dangers. Conversely, a smaller Precision cost means a higher likelihood of false positives, necessitating a greater cautious interpretation of the machine's indicators.

Recall, or sensitivity, quantifies the gadget's functionality to identify all relevant times of cyber risks, limiting false negatives. The Recall values within the submitted samples range from 0.78 to 0.93. A Recall price approaching 1.0 signifies that the machine efficaciously captures the most of actual good occurrences. A lower Recall cost, on the other hand, means that the machine might also ignore a few instances of actual threats, probably leaving the community susceptible.

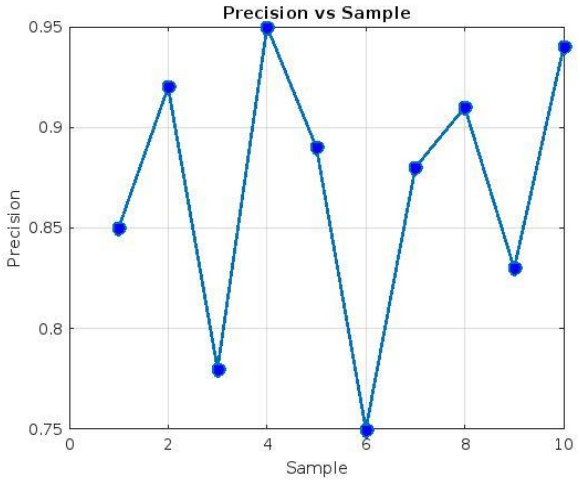

Fig. 7. Precision

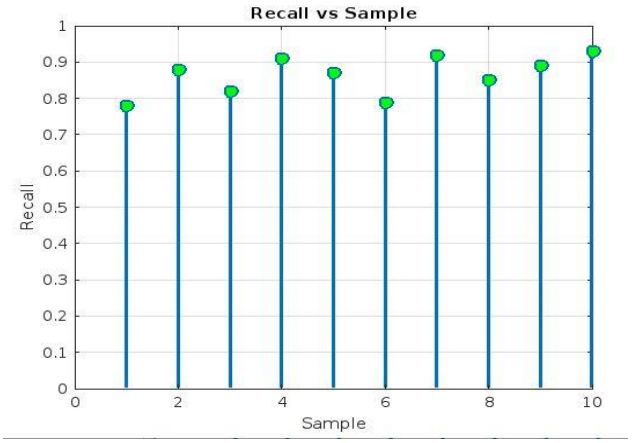

Fig. 8. Recall

F1 Score, the harmonic combination of Precision and Recall, moves a balance among these two measurements. F1 Score values within the samples vary from 0.77 to 0.94. A better F1 Score implies a balanced overall performance in terms of each fake positives and fake negatives. It is in particular useful while Precision and Recall need to be taken into consideration jointly, offering a complete view of the device's typical accuracy. In the samples selected ,Accuracy values range from 0.76 to 0.94. A better Accuracy generally indicates a system with good average performance.

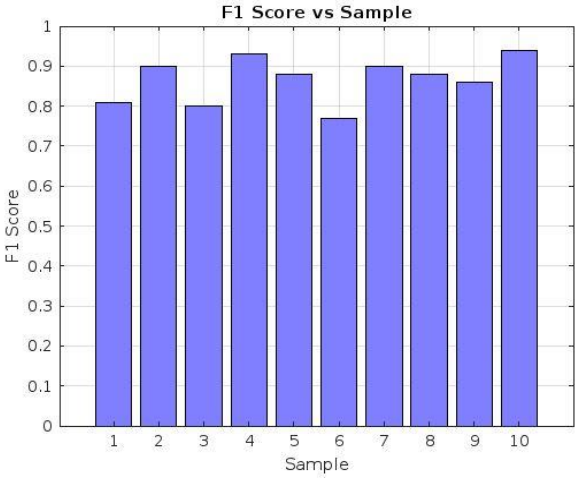

Fig. 9. F1 score

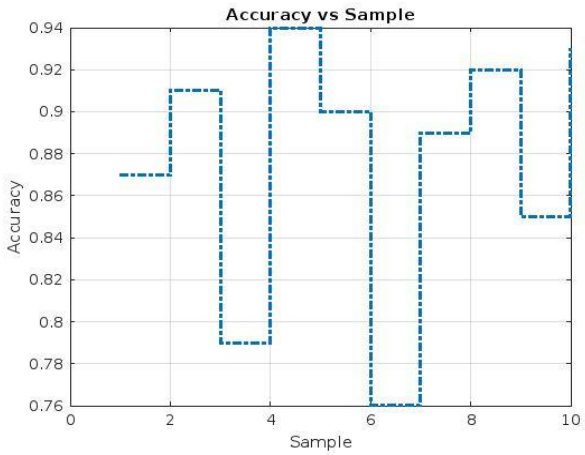

Fig. 10. Accuracy

7. Conclusion

The research of intrusion detection systems and its regular overall performance signals presentations a numerous environment where the necessity of cybersecurity connects with cutting-edge technologies. The precision, Recall, F1 score, and accuracy metrics, as displayed during 10 samples, offer a more detailed examination of the device's abilties.

Achieving a balance among minimizing false positives and false negatives is crucial for a successful intrusion detection equipment. The integration of system acquiring knowledge of and deep researching further augments these structures, imparting adaptive defenses in opposition to an ever-evolving chance landscape. As firms attempt for heightened security, understanding and optimizing those performance metrics become paramount. This full assessment not simplest helps cybersecurity experts to develop intrusion detection techniques but also contributes to the wider conversation on enhancing the resilience of digital ecosystems in opposition to cyber threats.

References

[1] M.Preetha, *et al.*, "Ant Colony Optimisation With Levy Based Unequal Clustering And Routing (ACO-UCR) Technique For Wireless Sensor Networks", Journal of Circuits, Systems, and Computers, ISSN: 0218-1266 (print); 1793-6454 (web) Vol .33, Issue3, July 24, 2023. DOI: 10.1142/S0218126624500439

[2] M. Parto, C. Saldana, and T. Kurfess, "A novel three-layer IoT architecture for shared, private, scalable, and real-time machine learning from ubiquitous cyber-physical systems," *Procedia Manuf.*, vol. 48, no. 2019, pp. 959–967, 2020, doi: 10.1016/j.promfg.2020.05.135.

[3] I. Zakariyya, H. Kalutarage, and M. O. Al-Kadri, "Towards a robust, effective and resource efficient machine learning technique for IoT security monitoring," *Comput. Secur.*, vol. 133, no. June, p. 103388, 2023, doi: 10.1016/j.cose.2023.103388.

[4] M.Preetha *et al.*, "Efficient Re-clustering with Novel Fuzzy Based Grey Wolf Optimization for Hotspot Issue Mitigation and Network Lifetime Enhancement," *Journal of Ad Hoc & Sensor Wireless Networks*, Vol. 56, Issue 4, page No-273-297, Sep 2023

[5] N. Jeffrey, Q. Tan, and J. R. Villar, "A hybrid methodology for anomaly detection in Cyber–Physical Systems," *Neurocomputing*, vol. 568, no. November 2023, p. 127068, 2024, doi: 10.1016/j.neucom.2023.127068.

[6] Z. Song, A. R. Mishra, and S. P. Saeidi, "Technological capabilities in the era of the digital economy for integration into cyber-physical systems and the IoT using decision-making approach," *J. Innov. Knowl.*, vol. 8, no. 2, p. 100356, 2023, doi: 10.1016/j.jik.2023.100356.

[7] Preetha M *et al.*,"CMAC-An Efficient Energy Postulate Based on Energy Cost Modeling in Wireless Sensor Network", Asian Journal of Information Technology, vol.12, issue.6, pp. 176-183, 2013,ISSN 1682-3915

[8] M. Wazid, A. K. Das, V. Chamola, and Y. Park, "Uniting cyber security and machine learning: Advantages, challenges and future research," *ICT Express*, vol. 8, no. 3, pp. 313–321, 2022, doi: 10.1016/j.icte.2022.04.007.

[9] P. Wang, Z. Man, Z. Cao, J. Zheng, and Y. Zhao, "Dynamics modelling and linear control of quadcopter," *Int. Conf. Adv. Mechatron. Syst. ICAMechS*, vol. 0, pp. 498–503, 2016, doi: 10.1109/ICAMechS.2016.7813499.

[10] Y. P. Tsang, T. Yang, Z. S. Chen, C. H. Wu, and K. H. Tan, "How is extended reality bridging human and cyber-physical systems in the IoT-empowered logistics and supply chain management?," *Internet of Things (Netherlands)*, vol. 20, no. June, 2022, doi: 10.1016/j.iot.2022.100623.

[11] M. Preetha *et al.*, "A Preliminary Analysis by using FCGA for Developing Low Power Neural Network Controller Autonomous Mobile Robot Navigation", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799, 2023

[12] G. Epiphaniou, M. Hammoudeh, H. Yuan, C. Maple, and U. Ani, "Digital twins in cyber effects modelling of IoT/CPS points of low resilience," *Simul. Model. Pract. Theory*, vol. 125, no. March 2022, p. 102744, 2023, doi: 10.1016/j.simpat.2023.102744.

[13] M. M. Hossain, M. A. Kashem, N. M. Nayan, and M. A. Chowdhury, "A Medical Cyber-physical system for predicting maternal health in developing countries using machine learning," *Healthc. Anal.*, vol. 5, no. May 2023, p. 100285, 2024, doi: 10.1016/j.health.2023.100285.

[14] Z. Noor, S. Hina, F. Hayat, and G. A. Shah, "An intelligent context-aware threat detection and response model for smart cyber-physical systems," *Internet of Things (Netherlands)*, vol. 23, no. June, p. 100843, 2023, doi: 10.1016/j.iot.2023.100843.

[15] M. Catillo, A. Pecchia, and U. Villano, "CPS-GUARD: Intrusion detection for cyber-physical systems and IoT devices using outlier-aware deep autoencoders," *Comput. Secur.*, vol. 129, p. 103210, 2023, doi: 10.1016/j.cose.2023.103210.

[16] M. Preetha *et al.*, "Deep Learning-Driven Real-Time Multimodal Healthcare Data Synthesis", International Journal of Intelligent Systems and Applications in Engineering (IJISAE), ISSN:2147-6799, Vol.12, Issue 5, page No-360-369, 2024

[17] M. Al-Hawawreh and N. Moustafa, "Explainable deep learning for attack intelligence and combating cyber–physical attacks," *Ad Hoc Networks*, vol. 153, no. April 2023, 2024, doi: 10.1016/j.adhoc.2023.103329.

[18] A. Aldhaheri, F. Alwahedi, M. A. Ferrag, and A. Battah, "Deep learning for cyber threat detection in IoT networks: A review," *Internet Things Cyber-Physical Syst.*, vol. 4, no. September 2023, pp. 110–128, 2024, doi: 10.1016/j.iotcps.2023.09.003.

[19] Preetha M *et al.*, "A Survey on Misbehavior Report Authentication Scheme of Selfish node Detection Using Collaborative Approach in MANET", International Journal of Engineering Science and Computing, vol. 6, no. 5, pp. 5381-5384, ISSN 2321-3361

[20] I. Singh, D. Centea, and M. Elbestawi, "IoT, IIoT and Cyber-Physical Systems Integration in the SEPT Learning Factory," *Procedia Manuf.*, vol. 31, pp. 116–122, 2019, doi: 10.1016/j.promfg.2019.03.019.

[21] M. Kato, T. Kizaki, T. Uwano, K. Iijima, and Y. Kakinuma, "Development of temperature analysis environment for Cyber-Physical Systems on IoT platform: a study of dynamical properties under temperature change in machine tool spindle unit using carbon fiber reinforced plastics," *Procedia CIRP*, vol. 107, no. March, pp. 1485–1490, 2022, doi: 10.1016/j.procir.2022.05.179.