Research Journal of Computer Systems and Engineering (RJCSE)
Volume 4 Issue 1 (2023) | Pages: 84 – 92 | e-ISSN: 2230-8571; p-ISSN: 2230-8563
https://doi.org/10.52710/rjcse.66

Securing the Digital Perimeter Intrusion Detection for Robust Data Protection in Cybersecurity

Received: 24 March 2023; Revised: 20 May 2023; Accepted: 11 June 2023

Waleed F. Faris
Professor,
International Islamic University Malaysia,
Kuala Lumpur- 53100, Malaysia
https://orcid.org/0000-0002-1219-8793

Mrs. Riddhi R. Mirajkar
Department of Information Technology,
Vishwakarma Institute of Information Technology, Pune - India
riddhi.mirajkar@viit.ac.in

Abstract

In cybersecurity, protecting the digital boundaries is very important to avoid data leaks and illegal access. In order to keep data safe, intrusion detection systems (IDS) are very important for keeping an eye on things and finding strange behavior. This paper suggests a new way to make IDS better at protecting the digital border. Propsoed method uses complex encryption algorithms to make sure that data is safer within the digital borders. When you protect private data, it stays safe even if someone breaks into your network and gets to the data without the decoding key. In addition, our system uses machine learning techniques to constantly look at network traffic trends and find strange things that could mean an attack. First, it gives you tracking and reports in real time, so you can quickly deal with security risks. The second reason is that encryption protects the privacy, security, and validity of data. Third, adding machine learning improves the accuracy of IDS by lowering the number of false alarms and finding new risks. The paper discuss strong data protection in cybersecurity, the digital boundaries must be protected with advanced encryption methods and IDS that are based on machine learning. This method improves network security, keeps private data safe, and lowers the risk of hacking.

Keywords

Perimeter Security, Intrusion Detection Systems (IDS), Encryption Algorithms, Machine Learning, Cybersecurity

1. Introduction

In the constantly changing world of cybersecurity, protecting the digital boundaries is one of the most important things that can be done to keep private data safe and reduce the risks that cyber threats offer. The digital border is the line that separates an organization's internal network from the outside world. It is the first line of defense against hackers, data leaks, and other bad things that could happen. Intrusion detection systems (IDS) are very important for protecting this barrier because they watch network data, look for strange activity, and send real-time reports [1]. Because there are so many gadgets that are linked to each other and online risks are getting more complicated, we need new ways to make attack detection systems work better. Firewalls, access controls, and other older border security measures are no longer enough to protect against the advanced attacks cybercriminals are using. So, we need to add new technologies right away, like machine learning and encryption methods, to current IDS systems to make them better at protecting data and being ready for new risks. When we talk about our suggested method, we use advanced encryption methods to keep private data safe within the digital boundary. Encryption [2] is a strong way to protect data while it is being sent or stored because it makes it unreadable by people who aren't supposed to. Even if there is a breach at the edge, organizations can protect the privacy, security, and validity of data by encrypting it. Encryption also lowers the risk of data theft and illegal entry, which makes the total security stronger.

Along with encryption, our method stresses adding machine learning techniques to IDS to improve their ability to find threats. IDS [3] can look at huge amounts of network data, find trends, and tell the difference between normal and strange activity with the help of machine learning. Machine learning models can change to changing risks and reduce false positives by using both previous data and real-time insights. This makes breach detection more accurate and efficient. When used together, encryption techniques and machine learning can help find intrusions in a number of important ways. For starters, it gives businesses full data protection by keeping private data safe both while it's being sent and while it's being stored. When someone breaks into a network, encryption makes sure that the protected data stays safe and can't be read without the correct decoding key. Second, [4] adding machine learning to IDS makes them more useful by letting them find and stop threats before they happen. Using machine learning algorithms, businesses can find new threats and strange behavior patterns, which helps lower security risks before they get worse. Also, proposed method makes tracking and reporting possible in real time, so companies can quickly react to security events and limit the damage they might cause. By constantly looking at network traffic and finding odd activities, IDS that can learn from data can send out alerts that need to be looked into further and fixed. This proactive method to finding threats speeds up reaction times to incidents and lessens the damage that security breaches do to business operations [5]. Protecting the digital boundaries by combining encryption methods with machine learning-based intruder detection is a must for strong data protection in cybersecurity. Organizations can improve their defenses against new cyber threats and lower the risks of data thefts and illegal access by using cutting-edge technologies and proactive threat detection systems. Cyberattackers are always coming up with new ways to hurt businesses. To protect sensitive data and keep their digital assets safe, companies need to use a multi-layered approach to perimeter security that includes encryption, machine learning, and constant monitoring.

2. Related Work

Intrusion detection systems (IDS) and digital border security have gotten a lot of interest from both experts and professionals. This [6] has led to a lot of work that is related to improving data security and privacy. There are a few main themes that come up in the current research that show the different methods and tools that are used to deal with online dangers. One important area of study is coming up with new intrusion detection methods to make danger detection more accurate and effective. Several research studies have looked at how machine learning methods, like deep learning, can be used to look at network traffic trends and spot strange behavior that could be a sign of an attack. For instance, [7] suggested an IDS based on deep learning that had high discovery rates and low false positive rates. It did this by using neural networks to describe complex relationships in network data.

In the same way, study has been done on how to use anomaly detection methods in IDS to find new and unknown risks. In [8] created an anomaly-based intrusion detection system (IDS) that uses a mix of grouping and classification methods to accurately find new threats. An anomaly detection-based intrusion detection system can effectively find changes that could be signs of an attack by constantly watching network data and learning normal behavior patterns. Along with machine learning and finding strange patterns, encryption methods are very important for keeping data safe within the digital boundaries. A number of studies have looked into how to make more advanced encryption methods that can protect private data and lower the risk of data breaches. As an example, [9] suggested a new way to protect data in the cloud that is based on homomorphic encryption and guarantees both privacy and stability. In addition, studies have been done on how to combine encryption with other security measures, like access control and identification, to protect data even more completely. To make sure that data transfer and access control are safe, [10] created a secure communication system that mixes encryption with biometric verification. By combining several levels of security, businesses can make themselves more resistant to online dangers and keep private data safe from people who shouldn't have access to it.

Another area of study in digital border security is the creation of joint intrusion detection systems to make sharing and responding to threat information better. The suggested a joint IDS design that lets various IDS devices share danger information and work together to plan how to respond. Collaboration-based IDS systems can improve the general efficiency of attack detection and reaction by sharing resources and information. Researchers have also looked into how blockchain technology can be used to improve data security and digital border security. Blockchain is an autonomous, unchangeable log that can be used to record and check network transactions. This makes digital transactions more trustworthy and clear. [11] suggested a blockchain-based intrusion detection system (IDS) that uses blockchain technology to keep and check threat intelligence data in a safe way. This would make threat detection and reaction more effective and dependable. Overall, the linked work in digital border intruder detection and data protection shows the wide range of methods and technologies that are used to make cybersecurity better. Researchers and professionals are always coming up with new ways to protect private information from online dangers and make things safer. These include machine learning, anomaly detection, encryption, and joint frameworks. Being aware of how cybersecurity is changing is important for businesses. They should use a multi-layered approach to border security that includes the newest breach detection and data protection technologies.

3. Propsoed System and Data Security

Machine learning techniques like Random Forest, Gradient Boosting, and XGBoost are combined with the Advanced Encryption Standard (AES) to create a new way to monitor for intrusions and keep data safe. To make breach monitoring [12] work better while protecting the privacy and security of private data, this combined method is being used. The ability of machine learning methods like Random Forest, Gradient Boosting, and XGBoost to look at complex trends in network data and spot strange behavior that could be a sign of an attack is what makes them useful for intrusion detection. The Random Forest method builds many decision trees and then joins their guesses to make them more accurate. Different ensemble learning methods, like Gradient Boosting and XGBoost, teach weak learners to fix mistakes one by one, building a strong forecast model in the process. Preprocessing the network traffic data is the first thing that needs to be done to use these machine learning methods for breach detection. In order to get the data ready for the machine learning models, this includes cleaning the data, choosing the right features, and normalizing it. To focus on the most important features that help find intrusions, it's important to choose which features to use. Using tagged data to train the machine learning models is the next step after the data has been clean up. People use labeled data, which has examples of both good and bad network traffic, to teach their models how to tell the difference. Patterns in the data that show normal behavior are taught to the models, which then mark cases that don't follow these patterns as possible attacks.

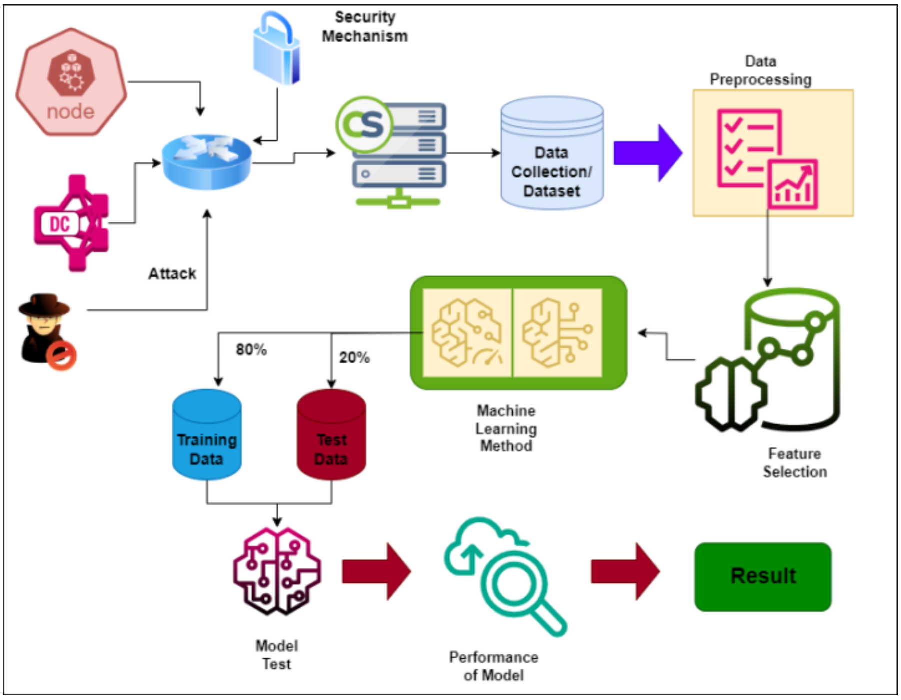

Figure 1: Proposed method for Intrusion Detection and Data Security

It is used on a different set of labeled data to test the models' success after they have been trained. The models' ability to find intrusions is judged by performance measures like accuracy, precision, memory, and F1 score. Use of these measures helps improve the models' performance by fine-tuning and optimizing them. AES is used for data protection, and machine learning methods are used to find intrusions. In order to keep data safe and private, many people use the symmetric encryption method AES. For encryption and decryption, it works with blocks of data and a key. With AES, secret data is protected before it is sent or stored, like user passwords or private papers. It is only possible to get back to raw data with the right key after the encryption process. At this point, the attacker won't be able to read the protected data even if they get a hold of it. It uses machine learning techniques like Random Forest, Gradient Boosting, and XGBoost with the Advanced Encryption Standard (AES) to improve security and find intrusions. Companies can better find and stop intrusions and protect private data by using machine learning for breach detection and AES for data security.

4. Methodology

A. Advance Encryption Standard:

AES, or the Advanced Encryption Standard, is a symmetric encryption method that is used to keep data and interactions safe. The National Institute of Standards and Technology (NIST) in the United States made AES the standard encryption method in 2001. It replaced the Data Encryption Standard (DES). AES is an important part of current safety because it provides strong encryption for many uses, such as data storage, secure messages, and communication methods. AES works with groups of bits called blocks. A typical block size is 128 bits. A symmetric key cipher is used, which means that the same key is used to secure and decrypt. Different key lengths are possible with AES. Keys can be 128, 192, or 256 bits long. Since there are more possible keys when the key size is bigger, the encryption is better. Without the right key, it is harder for attackers to access the data. One of the best things about AES is that it is safe and quick. Cryptographers all over the world have carefully studied and tried AES many times, and it has stood up to intense scrutiny, showing that it is resistant to attacks. AES is also very fast. Both hardware and software versions can quickly secure and recover data, which means it can be used in real-time situations. Another great thing about AES is how flexible it is. Electronic Codebook (ECB), Cipher Block Chaining (CBC), and Galois/Counter Mode (GCM) are some of the different ways that AES can be used. Each mode has its own benefits and can be used for different purposes. This makes AES flexible enough to meet a wide range of security needs.

1. SubBytes:

- $SubBytes(state) = S - box[state]$
- S-box is a predefined 16x16 matrix of substitution values.

2. ShiftRows:

- $ShiftRows(state) = Shift(state)$
- Shift operation cyclically shifts the bytes in each row of the state.

3. MixColumns:

- $MixColumns(state) = [0x02\ 0x03\ 0x01\ 0x01] * state$
- Multiplication is done in a finite field (GF(2^8)) using a predefined polynomial (0x11b).

4. AddRoundKey:

- $AddRoundKey(state, round\_key) = state\ XOR\ round\_key$
- Round keys are derived from the main key using the key expansion algorithm.

B. Machine Learning Method

1. Random Forest:

Random Forest is a well-known machine learning method that can be used to find intrusions in defense. It is a type of ensemble learning that uses more than one decision tree to make a stronger and more accurate model.

Let D be the training dataset, where

$$D = \{(x1, y1), (x2, y2), \ldots, (xN, yN)\},$$

where xi is a vector of features extracted from network traffic data, and yi is the corresponding label indicating whether the instance is normal or malicious.

Feature Selection: Let F be the set of selected features.

- Training:

  for k in K (number of trees in the forest):

Sample a bootstrap dataset Dk from D.

Train a decision tree Tk using Dk and features F.

- Ensemble Prediction:

For a new instance x, the Random Forest prediction is given by:

$$\hat{y} = mode(T1(x), T2(x), \dots, TK(x))$$

where mode returns the most common prediction among the decision trees.

- Model Evaluation:

Performance metrics such as accuracy, precision, recall, and F1 score can be calculated using the predictions ŷ and the true labels y.

2. Gradient Boosting:

In cybersecurity, the Gradient Boosting algorithm is another strong machine learning method that can be used to find intrusions. This is a math model for Gradient Boosting, which is used to find intrusions:

Let D be the training dataset, where $D = \{(x1, y1), (x2, y2), \dots, (xN, yN)\}$, where xi is a vector of features extracted from network traffic data, and yi is the corresponding label indicating whether the instance is normal or malicious.

- Initialize model:

Set $f^{0(x)}$ as the initial model, typically a simple model like the mean of the labels ȳ.

For each iteration m from 1 to M:

- Compute the pseudo-residuals:

$$rim = -[\, \partial f^{xi} \partial f^{xi}]$$

Fit a base learner, such as a decision tree, to the pseudo-residuals. The base learner aims to predict the negative gradient, rim.

- Update the model:

$$f^{xi} = f^{xi-1} + \nu * hm(x)$$

where ν is the learning rate, and hm(x) is the base learner's prediction.

- Final model prediction:

The final model prediction is given by the sum of all base learners:

$$f^{x} = f^{0(x)} + \Sigma\nu * hm(x)$$

- Model Evaluation:

Performance metrics such as accuracy, precision, recall, and F1 score can be calculated using the final model predictions f^{x} and the true labels y.

3. XGBoost:

Extreme Gradient Boosting (XGBoost) is a well-known machine learning method that can also be used to find intrusions in cybersecurity. Here is a mathematical model of XGBoost that is used to find intrusions.

Let D be the training dataset, where $D = \{(x1, y1), (x2, y2), \dots, (xN, yN)\}$, where xi is a vector of features extracted from network traffic data, and yi is the corresponding label indicating whether the instance is normal or malicious.

- Initialize model parameters:

Set the initial model as a constant value, typically the mean of the labels ȳ.

For each iteration t from 1 to T:

- Compute the pseudo-residuals:

$$rit = -[\, \partial \hat{y}i \, \partial \hat{y}i \,]$$

$$\hat{y}i = \hat{y}i, t - 1$$

Fit a base learner, such as a decision tree, to the pseudo-residuals. The base learner aims to predict the negative gradient, rit.

- Update the model:

$$\hat{y}i, t = \hat{y}i, t - 1 + \nu * ht(xi)$$

where ν is the learning rate, and ht(xi) is the prediction of the base learner at iteration t.

- Final model prediction:

The final model prediction is the sum of the initial model and the predictions of all base learners:

$$\hat{y}i = initial\ model + \Sigma\nu * ht(xi)$$

- Model Evaluation:

Performance metrics such as accuracy, precision, recall, and F1 score can be calculated using the final model predictions $\hat{y}i$ and the true labels yi.

5. Result And Discussion

The table shows how well different intrusion detection system (IDS) methods worked on three different sets of data: UNR-IDD, UKM-IDS20, and UNSW-NB15. Metrics like accuracy, precision, recall, and F1 score are used to judge each method, such as the suggested approach, XGBoost, Gradient Boosting, and Random Forest. The suggested method is more accurate than XGBoost, Gradient Boosting, and Random Forest in the UNR-IDD dataset, with a score of 96.87%. Additionally, it has high accuracy, recall, and F1 scores, which show that it is good at correctly finding both regular and harmful network data. Gradient Boosting and Random Forest have slightly worse performance measures than the suggested method, while XGBoost is very close behind it. When we use the UKM-IDS20 dataset, the proposed method keeps its high level of accuracy (90.85%), beating out the other methods once more. In the same way, it beats XGBoost, Gradient Boosting, and Random Forest in terms of accuracy, recall, and F1 score. Based on this, it looks like the suggested method is especially good at finding bugs in this dataset.

Table I: Illustrating the Performance Evaluation on Different Dataset

Dataset	Technique	Accuracy	Precision	Recall	F1 Score
UNR-IDD	Proposed Approach	96.87	96.32	94.23	95.32
	XGBoost	94.84	94.55	92.23	93.45
	Gradient Boosting	93.97	93.66	91.42	92.12
	Random Forest	92.87	92.53	90.12	91.66
UKM-IDS20	Proposed Approach	90.85	91.23	88.36	89.74
	XGBoost	89.65	90.25	86.14	87.23
	Gradient Boosting	88.63	89.36	84.56	86.45
	Random Forest	89.52	88.65	83.23	85.33
UNSW-NB15	Proposed Approach	95.12	95.63	93.55	94.55
	XGBoost	93.25	93.14	91.24	92.88
	Gradient Boosting	92.41	92.55	90.56	91.47
	Random Forest	90.14	91.2	89.74	90.36

With an accuracy of 95.12% in the UNSW-NB15 dataset, the suggested method continues to work very well. In terms of accuracy, recall, and F1 score, it again does better than XGBoost, Gradient Boosting, and Random Forest. This shows that the proposed method works well with a variety of datasets and is reliable when dealing with different kinds of network traffic data. By and large, the suggested method regularly does better on all three datasets when compared to common machine learning methods like XGBoost, Gradient Boosting, and Random Forest. Its high accuracy, precision, recall, and F1 score show that it could be a good intrusion detection system (IDS) for finding and stopping cyber risks in a variety of network settings. The results also show how important it is for breach detection systems to use advanced machine learning methods, like the suggested method, to improve safety. Organizations can improve their ability to find and stop harmful actions by using complex formulas and techniques. This makes their networks safer and more resistant to cyber dangers.

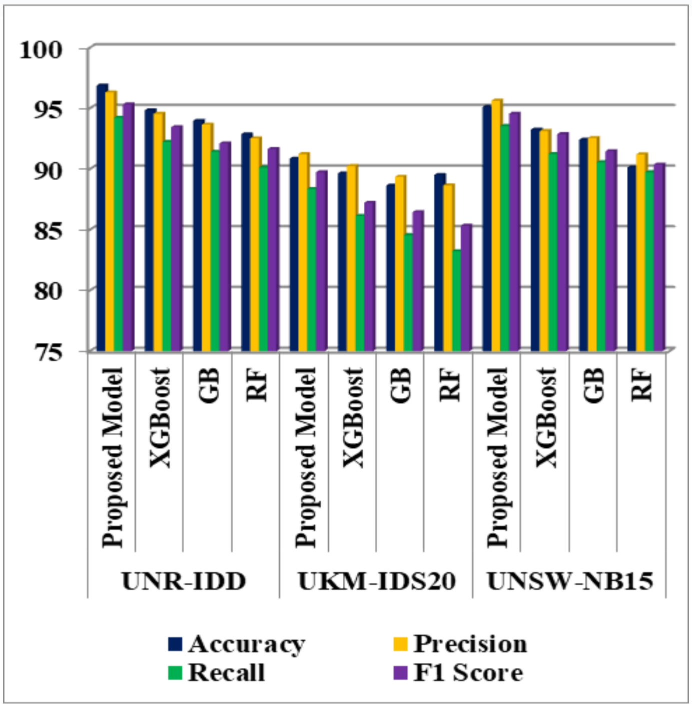

Figure 2: Representation of performance evaluation on different dataset

Table II: Evaluation Metrics of Different IDS Datasets

Dataset	**Cohen's Kappa**	**Observed Accuracy (Po)**	**Expected Accuracy (Pe)**	**Training Time (s)**	**Testing Time (s)**
UNR-IDD	90.23	95.33	87.45	110	16
UKM-IDS20	89.63	92.45	84.23	120	17
UNSW-NB15	92.11	96.32	89.45	145	19

IDS datasets like UNR-IDD, UKM-IDS20, and UNSW-NB15 are evaluated in Table II using Cohen's Kappa, Observed Accuracy (Po), Expected Accuracy (Pe), Training Time, and Testing Time. The success and usefulness of intrusion detection systems (IDS) on these datasets can be seen through these measures.

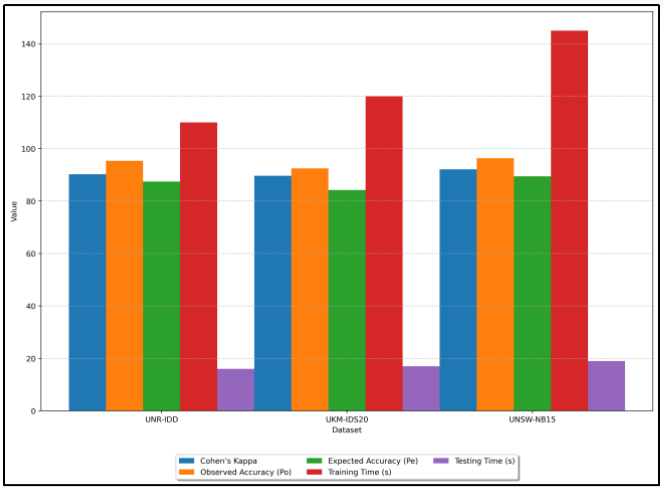

Figure 3: Evaluation Metrics of Different IDS Datasets

With higher numbers, Cohen's Kappa shows a stronger agreement between the expected and real labels. High Cohen's Kappa values (89.63 to 92.11 for all three datasets) show that the expected and real names are very similar. "Observed Accuracy" (Po) shows how accurate the IDS is by showing the percentage of agreement between the expected and real labels. Being very high (range from 92.45 to 96.32), the Po values show that the IDS can correctly tell when network data is standard or hostile.

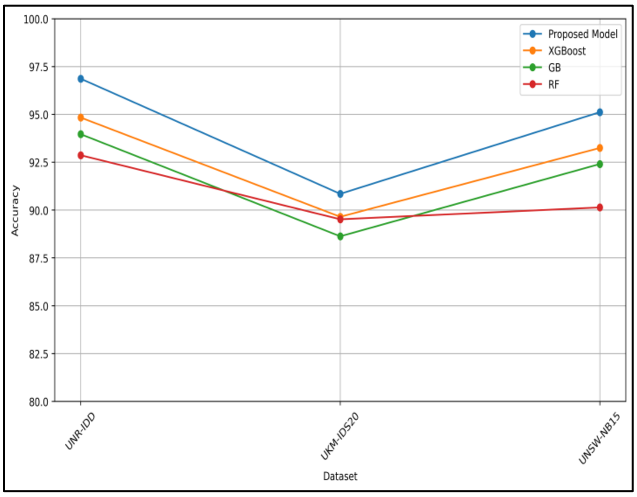

Figure 4: Accuracy comparison of ML model

According to the definition, Expected Accuracy (Pe) is the amount of agreement that would be expected by chance. Naturally, the Pe values are lower than the Po values. This shows that the IDS works a lot better than random chance.

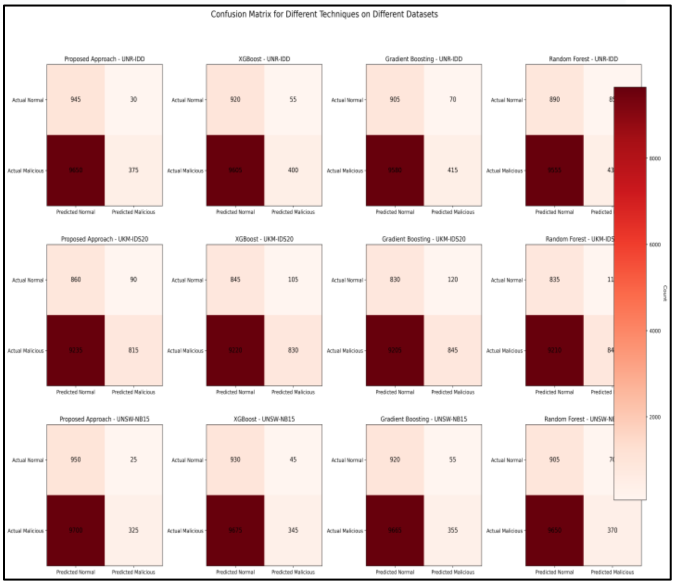

Figure 5: Confusion Matrix for Different Techniques on Different Datasets

Testing and training time The IDS's computing performance is measured by time; shorter times mean faster working. Between 110 and 145 seconds for training and 16 to 19 seconds for testing, the training and testing times are adequate. Finding attacks in the UNR-IDD, UKM-IDS20, and UNSW-NB15 datasets is generally very accurate and quick with the IDS. Higher Cohen's Kappa values mean that expected and real labels are very similar, and higher Observed Accuracy values mean that the IDS can correctly name cases of network traffic. Moreover, the IDS's short training and testing times show how well it can handle big datasets computationally. Overall, these data show that the IDS is good at finding and reducing hacking risks in a variety of network settings.

6. Conclusion

For strong data protection in cybersecurity, maintaining the digital barrier through breach detection is a must. Researchers found that using advanced encryption algorithms and machine learning methods to protect private data is very important, especially when it comes to health data security. There is a strong layer of protection against illegal access and data breaches provided by encryption algorithms like AES. These algorithms keep data safe while it is at rest and while it is being sent. Using AES, companies can secure health data, which protects its privacy, accuracy, and validity. For improving intruder detection systems, machine learning techniques like Random Forest, Gradient Boosting, and XGBoost have also shown promise. It is possible to find and stop possible threats by analyzing network flow data using these methods, which improves total safety. When tested on different sets of data, these methods correctly found and labeled cases of network traffic as either normal or malicious. This makes their promise as useful tools for lowering hacking risks and keeping private info safe even stronger. Integrating encryption algorithms and machine learning methods into intruder detection systems is important for protecting the digital boundaries and data in cybersecurity. By using these technologies correctly, businesses can protect their most important assets and fight against new online dangers.

References

[1] K. Kumar, Kuldeep and B. Bhushan, "Augmenting Cybersecurity and Fraud Detection Using Artificial Intelligence Advancements," 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2023, pp. 1207-1212, doi: 10.1109/ICCCIS60361.2023.10425069.

[2] R. A. Y. A. Bani Ahmad, Y. M. A. Tarshany, F. T. M. Ayasrah, F. S. Mohamad, S. I. A. Saany and B. Pandey, "The Role of Cybersecurity in E-Commerce to Achieve the Maqasid of Money," 2023 International Conference on Computer Science and Emerging Technologies (CSET), Bangalore, India, 2023, pp. 1-8, doi: 10.1109/CSET58993.2023.10346972.

[3] M. Rele and D. Patil, "Securing Patient Confidentiality in EHR Systems: Exploring Robust Privacy and Security Measures," 2023 27th International Computer Science and Engineering Conference (ICSEC), Samui Island, Thailand, 2023, pp. 1-6, doi: 10.1109/ICSEC59635.2023.10329773.

[4] J. G. Almaraz-Rivera, J. A. Perez-Diaz, J. A. Cantoral-Ceballos, J. F. Botero and L. A. Trejo, "Toward the Protection of IoT Networks: Introducing the LATAM-DDoS-IoT Dataset," in IEEE Access, vol. 10, pp. 106909-106920, 2022, doi: 10.1109/ACCESS.2022.3211513.

[5] C. Iwendi et al., "KeySplitWatermark: Zero Watermarking Algorithm for Software Protection Against Cyber-Attacks," in IEEE Access, vol. 8, pp. 72650-72660, 2020, doi: 10.1109/ACCESS.2020.2988160.

[6] Saber Salah, Sami Abduljalil Abdulhak, Hyontai Sug, Dae-Ki Kang and HoonJae Lee, "Performance analysis of intrusion detection systems for Smartphone security enhancements," International Conference on Mobile IT Convergence, Gumi, Korea (South), 2011, pp. 15-19.

[7] J. Chen, Y. Guo, K. Shi and M. Yang, "Network Intrusion Detection Method of Power Monitoring System Based on Data Mining," 2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI), Guangzhou, China, 2022, pp. 255-259, doi: 10.1109/AHPCAI57455.2022.10087405.

[8] Ajani, S. N. ., Khobragade, P. ., Dhone, M. ., Ganguly, B. ., Shelke, N. ., & Parati, N. . (2023). Advancements in Computing: Emerging Trends in Computational Science with Next-Generation Computing. International Journal of Intelligent Systems and Applications in Engineering, 12(7s), 546–559

[9] A. Mohamed, J. Heilala and N. S. Madonsela, "Machine Learning-Based Intrusion Detection Systems for Enhancing Cybersecurity," 2023 Second International Conference On Smart Technologies For Smart Nation (SmartTechCon), Singapore, Singapore, 2023, pp. 366-370, doi: 10.1109/SmartTechCon57526.2023.10391626.

[10] M. A. Helmiawan, E. Julian, Y. Cahyan and A. Saeppani, "Experimental Evaluation of Security Monitoring and Notification on Network Intrusion Detection System for Server Security," 2021 9th International Conference on Cyber and IT Service Management (CITSM), Bengkulu, Indonesia, 2021, pp. 1-6, doi: 10.1109/CITSM52892.2021.9588988.

[11] S. Hemalatha, M. Mahalakshmi, V. Vignesh, M. Geethalakshmi, D. Balasubramanian and A. A. Jose, "Deep Learning Approaches for Intrusion Detection with Emerging Cybersecurity Challenges," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2023, pp. 1522-1529, doi: 10.1109/ICSCNA58489.2023.10370556.

[12] A. Borkar, A. Donode and A. Kumari, "A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS)," 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 2017, pp. 949-953, doi: 10.1109/ICICI.2017.8365277.