E-ISSN: 2997-9382
American Journal of Technology Advancement
https://semantjournals.org/index.php/AJTA

| Research Article

Check for updates

Assessing Machine Learning Effectiveness in Intrusion Detection Systems

Hazem Salim Abdullah [1], Mohamed Abdelalmalik Shaker [2], Sader Isam Mahmood [3], Ahmad Saad Abdulmajeed [4]

[1] Directorate of Municipalities, Nineveh Governorate, Mosul, IRAQ
[2] University of Mosul, Daar Ibn Al-Atheer Printing and Publishing, Mosul, IRAQ
[3] Directorate of Electricity Distribution, Nineveh Center, Mosul, IRAQ
[4] Nineveh Education Directorate, Mosul, IRAQ

Abstract

The increased complexity of cyber attacks and advanced persistent threat (APT) scenarios has elevated the importance of Intrusion Detection Systems (IDS) in securing networks from maliciousness. Because of the fast-changing nature of cyber threats and attacks, most traditional techniques, including rules-based systems and signatures, are no longer providing adequate protection against sophisticated cyber-attacks. This study examines the use of various machine learning algorithms (decision tree [DT], k-nearest neighbor [KNN], support vector machine [SVM], and logistic regression [LR]) using the NSL-KDD dataset for intrusion detection classification. Overall, the proposed approach produces results indicating that Machine Learning Algorithms outperform traditional methods, exhibit high classification accuracy, and demonstrate capability of classifying known and unknown attacks. As a result, this research enhances the performance of intrusion detection systems (IDS) by ensuring an effective and versatile solution to securing networks. Performance evaluations of several machine learning algorithms showed a high level of success (99.58%, 98.33%, 97.70%, and 91.96%). Among the four evaluated models, the LR model had the lowest level of performance accuracy while the DT model had the highest level of accuracy.

Keywords: Intrusion Detection Systems, Cyber Security, Machine Learning, Network Intrusion Detection System.

1. Introduction

The quick advancement of wireless communication technology at the physical layer also raises security issues. Wireless eavesdropping, identity spoofing, information manipulation, and other shortcomings in security are common among them and can cause problems for users of wireless networks [1]. There is an increasing number of physical devices connected to networks the

Internet as the technology develops [2]. When rules attached lots of data is generated and saved. The era of 'big data' is taking shape [3]. Increasing internet traffic makes modern networks more susceptible to a variety of attacks, potentially causing network service interruptions and financial damages. Strong Intrusion Detection System (IDS) approaches are required to address these issues. Given the evolution of threats in networks, there is a need to incorporate more advanced machine learning techniques into security protocols, as existing traditional methods are no longer sufficient [4]. Machine learning techniques tailored for intrusion detection are necessary for increasing the accuracy and effectiveness of threat detection. Tailored ML approaches for intrusion detection enable the IDS to evolve with new attack techniques, thus reduce the false positive rates while ensuring fast detection of novel threats [5]. New advances in ML based IDS for IoTs highlight the use of unsupervised learning for zero-day detection, deep learning for higher accuracy, lightweight models for reserve forced plans, advances in adversarial security and blockchain for advanced security and data integrity [6]. This article presents a comprehensive review of deep combinatorial learning methods tailored for intrusion detection systems (IDS) in IoT, with a special focus on lambda architecture for real, time computation and accuracy enhancement. In addition, it suggests future research directions such as automated hyper, parameter tuning and real, world validation [7]. The recent trend in system mastering, primarily based IDS in IoT can be considered as cybersecurity initiatives through predictive models, fashions and optimization strategies, controlling the security risks of connected gadgets, and facilitation of actual, time decision, making even as preserving privacy and moral issues [8].

In fact, the proliferation of laptops, mobile phones, and tablets in the real world has increased the demand for online privacy and data protection. Significantly, Intrusion Detection Systems (IDS) monitor the device activities for violations and thus, they provide management with indispensable reports. It is crucial to develop accurate violation detection models; machine learning is one such method that can be employed to detect unusual network traffic and malicious activities effectively. The paper presents a comparative study of four machine learning models on the NSL, KDD dataset. It assesses the models' performance when dealing with familiar and unfamiliar attacks and proposes practical solutions for implementing lightweight IDS in resource, constrained environments. The paper comprises five sections: Introduction, Literature Review, Methodology, Results and Discussion, and Conclusions with Future Recommendations.

2. Literature Review

IoT expedients are at great risk due to the increasing proportion of cyber attacks and newly it needs more attention. In the literature, many solutions have been proposed utilizing deep learning and machine learning to identify and stop these attacks. For classification, a few popular techniques are employed, including SVM, KNN, Decision Tree, Ensemble approaches, and CNN [9]. For illustration the authors have used autoencoder methods for detecting online intrusions [10]. Online access to NSL KDD data is available for usage as input data. [11]. All symbols are first transformed into numeric functions and then back into symbolic functions in order to preprocess NSL-KDD data. Features are extracted using principal component analysis. The accuracy, precision, and recall of machine learning algorithms used to categorize preprocessed data are compared in this study. Machine learning algorithms include random forests, linear regression, and support vector machines [12]. ANN was employed by the authors to detect network intrusions [13]. The number of false alarms was decreased by the authors' hybrid feature selection approach prior to classification [14]. The scientists deployed an ensemble of ANNs for multiclass intrusion detection and executed 94 .96% accuracy with the KDD99 dataset [15]. For Internet of Medical Things (IoMT) networks, the authors in [16] advised a generative IDS using deep studying. Employing the KD-CUP99 dataset, the authors of [17] received 94.12% accuracy via using an advanced Seagull Optimization Algorithm (SOA) for function selection and a Recurrent Neural Network (RNN) classifier for cyberattack detection. Liu and colleagues. [18]

Using the KDD99 dataset, CNN is applied for feature extraction earlier than MLP is hired to identify usual and uncommon user conduct. A DNN-primarily based IDS system was recommended with the aid of the authors [19]. They asserted that DNN with ant layers outperforms other classifiers for gadget studying. Several datasets, including the UNSW_NB-15, NSL-KDD, and CIC-IDS-2017 datasets, have been used to assess the model. The UNSW-NB15 dataset was utilized by the authors of [20] to broaden a community anomaly detection technique. The model was also tested with different classifiers, and a classification accuracy of 87.37% and 99.94% was achieved in the case of the Worms class by using a Reduced Error Pruning Tree classifier. To ensure accurate results, the author in the paper [21] proposed different models using meta-classification techniques in ensemble models. The model delivered results with an accuracy of 94.27% and 82.22% when tested with UNSW-NB15 and UGR'16 datasets. In a similar context, different speech classifiers were used by the authors in [22] to develop different models with an accuracy rate of 99.7%. The literature abundantly proves that better models need to be designed to handle issues related to sophisticated cyberattacks in the Internet of Things environment. Moreover, to improve the effectiveness of ML-based IDS in the near future, enhanced accuracy can be ensured by integrating ensemble learning techniques.

3. Methodology

3.1 Intrusion Detection Systems (IDSs)

Intrusion detection systems (IDS) are key software or hardware instruments for monitoring and analyzing security related incidents of computer system or network. IDSS are now an integral and critical part of the network security infrastructure in organizations, due to the rise in number of network intrusions. They are an effective means to reduce risks from external attacks and insider abuse of privileges. Security administrators should concentrate more on selecting appropriate IDS technologies as it has become crucial for safeguarding IT infrastructures [24].

There are many reasons to implement intrusion detection systems (IDS) because they deliver multiple benefits which include detecting security breaches and defending against potential attacks and assessing present threats and enhancing security management and design processes which create security risks and which provide vital information about real security breaches for effective diagnosis and recovery [25]. The intrusion detection system enables companies to control user behavior through its security threat detection system which increases user awareness of attacker identification and punishment results.

3.2 Machine Learning Based Intrusion Detection System

The Machine learning is frequently used to create an intrusion detection system (IDS) that automatically identifies and categorizes network intrusions and host-level threats. Malicious assaults vary and occur in large numbers, necessitating a scalable solution. Cyber security researchers may conduct research and associated tasks using public malware databases. Security and analysis at the data and physical levels Data protection has become increasingly vital as data volumes increase. IDSs gather and analyses data to detect system or network intrusions and avoid data loss. The volume, diversity, and speed of network data make data analysis for identifying assaults difficult. IDS develops data security mechanisms using ML algorithms that are precise and efficient. Computers may learn from data using a wide choice of techniques acknowledged as machine learning is characterized by a set of essential characteristics, including:

- Its algorithm, which performs better with structured and ordered data, may be effectively built with a small to medium quantity of data.
- It breaks down the issue into minor matters and resolves each separately.
- Its algorithms, like decision trees and linear regression, have a fundamental structure.

- A central processing unit (CPU) is capable of running it.
- More human involvement is needed to choose and analyze characteristics in its algorithms to select the correct input.
- It takes less time during training but slows down during testing.

Some machine learning algorithms remain the quickest despite this. Therefore, Machine learning is becoming a crucial tool for identifying and reducing cyber threats in smart cities. Machine learning approaches fall into three primary categories: anomaly-based, signature-based, and hybrid systems [26]. A hybrid system combines the assets of both approaches for greater accuracy and efficiency. Anomaly-based detection employs system intelligence that has been educated using several methods, whereas Network traffic is compared using signature-based detection.

to pre-existing attack patterns or signatures see the Figure 1. Researchers have created multiple IDSs using different methodologies, algorithms, and target systems, comparing their accuracy and precision with other algorithms. Figure 2 illustrates ML Techniques-based-intrusion detection system and Figure 3 show Machine-learning -based-intrusion-detection cycle. We will Give a brief synopsis of the relevant IDS algorithms in this part, as follows:

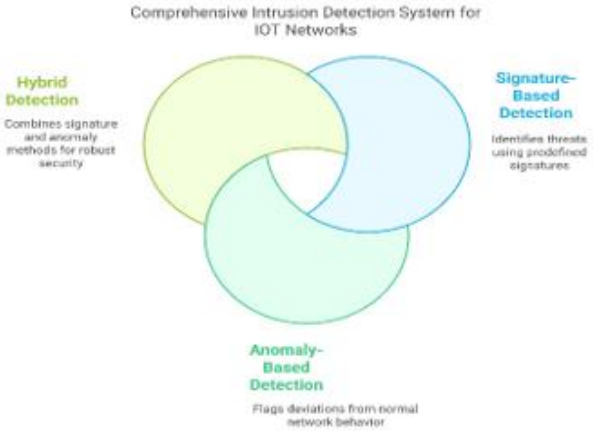

Figure 1. A comparison between functions of IDS types

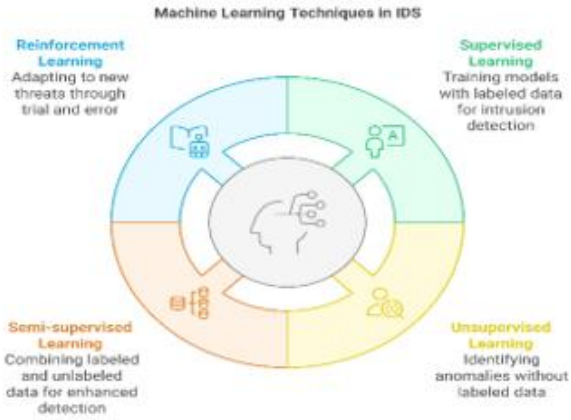

Figure 2. Machine-learning Techniques in IDS

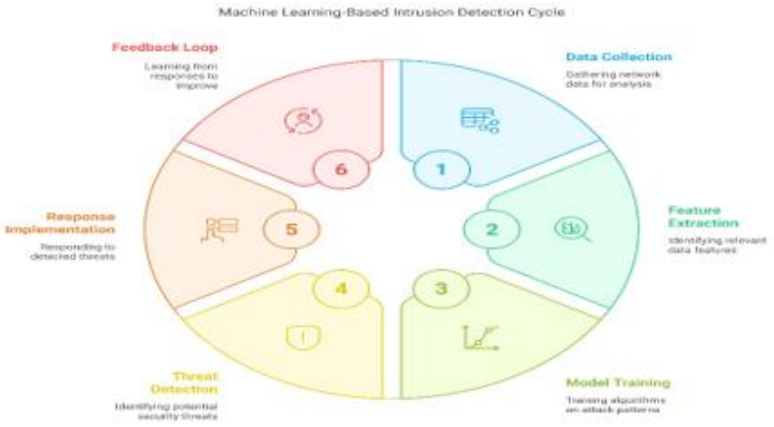

Figure 3. Machine-learning -based-intrusion-detection cycle

3.2.1 K -nearest neighbor's algorithm (KNN)

K- nearest neighbors assess new data by comparing its similarity to existing data for regression and classification tasks. Known as a lazy learner, it postpones action until new data arrives,

relying on previously collected information. The algorithm operates in phases: first, it determines the number of neighbors; second, it identifies the area around each neighbor; third, it selects the K closest points; and finally, it totals the data points in each category to predict the class with the highest number of neighbors [32].

3.2.2 Logistic Regression algorithm (LR)

Logistic Regression (LR) is a machine learning approach used in binary classification tasks and multiclass classification when using one-vs-rest approaches. It applies a linear model to the sigmoid function or its variants, squashing the output between 0 and 1 and determining a class's probability by an output closer to 1, which can then be mapped to two or more discrete classes [27].

3.2.3 Decision Tree (DT)

A supervised, non-parametric machine learning method known as DT is employed in solve regression and categorical problems. Prediction of the cost of a dataset's output arises from attractive selection rules from dataset features. It is easy to understand and interpret and can be visualized. It can handle problems with multiple outputs [28]. It is usually recycled in IDS. The decision node, which has some outlets and is shown to form choices and leaf nodes, no longer contains any branches and the output of these choices are two nodes in the DT. The first decision node is called the root node. To construct a selection tree, to choose the fact, the Gini index is subjected to the attribute selection measure (ASM) benefit and feature [29].

3.2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a leading supervised learning process for classifying data into positive or negative classes. It works by first plotting input data into a higher-dimensional nose planetary to uncover hidden patterns. The next step involves identifying the optimal hyperplane that best separates the mapped features. This maximizes the separation margin, defined as the separation between the hyperplane and the closest data point. thereby improving both classification and regression results [38].

3.3.1 NSL-KDD Dataset

The KDD99 dataset, developed in 1999, is a prominent resource for studies in cyber security, but has notable issues such as data duplication and excessive entry counts in both training and testing datasets. These challenges hinder the effective use of the complete dataset in research contexts. To resolve these deficiencies .The NSL-KDD dataset was introduced as an better version , gaining recognition since 2009 as the leading dataset for cyber security research. NSL-KDD consists of two subsets: KDDTrain+, containing 125,973 records and KDDTest+, which comprises 22,544 records. Every record is characterized by 41 properties categorized into fundamental features, connection-based traffic features time-based traffic features, and content features with 21 assigned label classes indicating attack or normal recordings. In the domain of cyber security see the Figure 4 showed Protocol Type, each dataset entry represents a session, defining a connection between two hosts in a network. Notably, there are differences in the probability distribution of KDDTrain+ and KDDTest+, prompting the need to evaluate models' performance on datasets that encompass attack types present in the test data but absent from the training data. This assessment approach is critical in assessing a model's potential to apply in the real world situations, in which intrusion detection structures (IDS) need to identify both current and capacity risks. The check dataset offers 14 awesome assault kinds now not observed within the training dataset, while the latter includes 24 special attack types. The approach taken with the aid of NSL-KDD, especially in checking out classifiers on unknown attacks, ensures models maintain accuracy for recognized threats whilst possessing the adaptability necessary to counter new cyber threats. By addressing problems of redundancy, bias, and erroneous reviews seen in KDD99, the NSL-KDD dataset

notably enhances intrusion detection structures and contributes to the improvement of extra powerful cyber hazard models [30]. A comprehensive overview of the NSL-KDD record information can be found in Table 1.

Table 1. NSL-KDD Record Details

	All Records	Normal	DOS	Probe	R2L	U2R
KDDTrain+	125,973	67343	45927	11656	995	52
KDDTest+	22,544	9711	7458	2421	2754	200

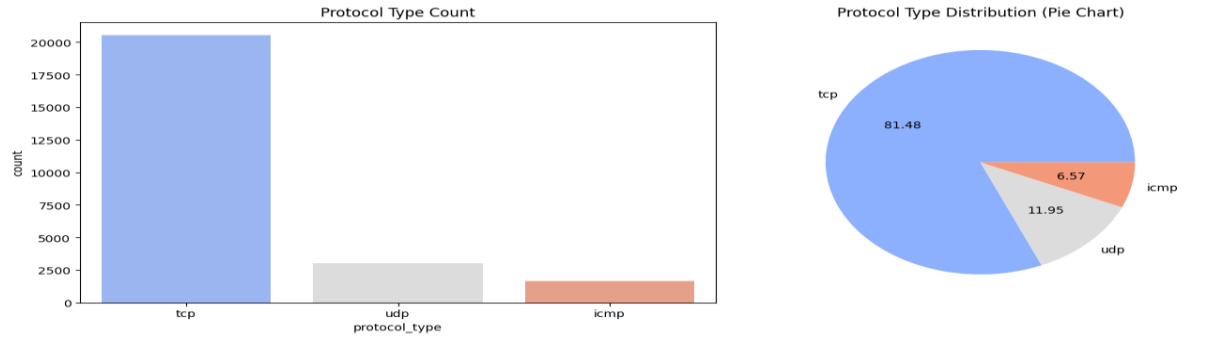

Figure 4. Protocol Type

3.3.2 Evaluation Metrics

Accuracy in percentage terms is the number of data points that were accurately predicted out of all the data points. Equation 1 provides the accuracy calculation.

Accuracy = (TP + TN) / (TP + FP + TN + FN) (1).

It represents the numbers of examples that are true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

- Sensitivity or recall is the percentage of positive class instances that are correctly forecasted to be positive. Equation 2 illustrates this formula.

Recall = TP/(TP + FN) (2).

- Precision, whose formula is in Equation 3, calculates the probability that a positive prediction will come true. The optimal degree of accuracy is 1.

Precision = TP / (TP + FP) (3).

- F1- Score is an indicator of the test's accuracy. It computes the score by considering the test's precision and recall. Equation 4 illustrates this formula.

F1-Score = 2TP / (2TP + FP + FN) (4).

- To evaluate the effectiveness of the device gaining knowledge of algorithms in detecting cyber threats, Receiver Operating Characteristic (ROC) curves were employed. The ROC curve is a plot showing the exchange between the True Positive Rate (TPR) and the False Positive Rate (FPR).as that ratio is varied, at a range of class thresholds.

By reading the Area Under the Curve (AUC) , we can quantify the model's capacity to distinguish between regular and malicious sports.

A model with an AUC closer to 1.0 indicates a stronger ability to detect cyber threats with fewer false alarms[31].

3.3.3 Implementation of proposed ML models

The Python 3.7 language is used to create and implement the K-Nearest Neighbors KNN, Logistic Regression LR, Decision Tree DT, and Support Vector Machine SVM are machine learning techniques. The system used for analysis is an i5 CPU, Windows 11 operating system, and 16 GB of RAM. The NSL-KDD dataset is separated into an 80:20 ratios for training and testing, ensuring a balance between bias and variance. The 20% test set provides a sufficient sample for generalization performance, while the 80% training set efficiently trains each model. Equations (1)– (4) are used to evaluate each model see the figure 5 explain flowchart of the suggested IDS.

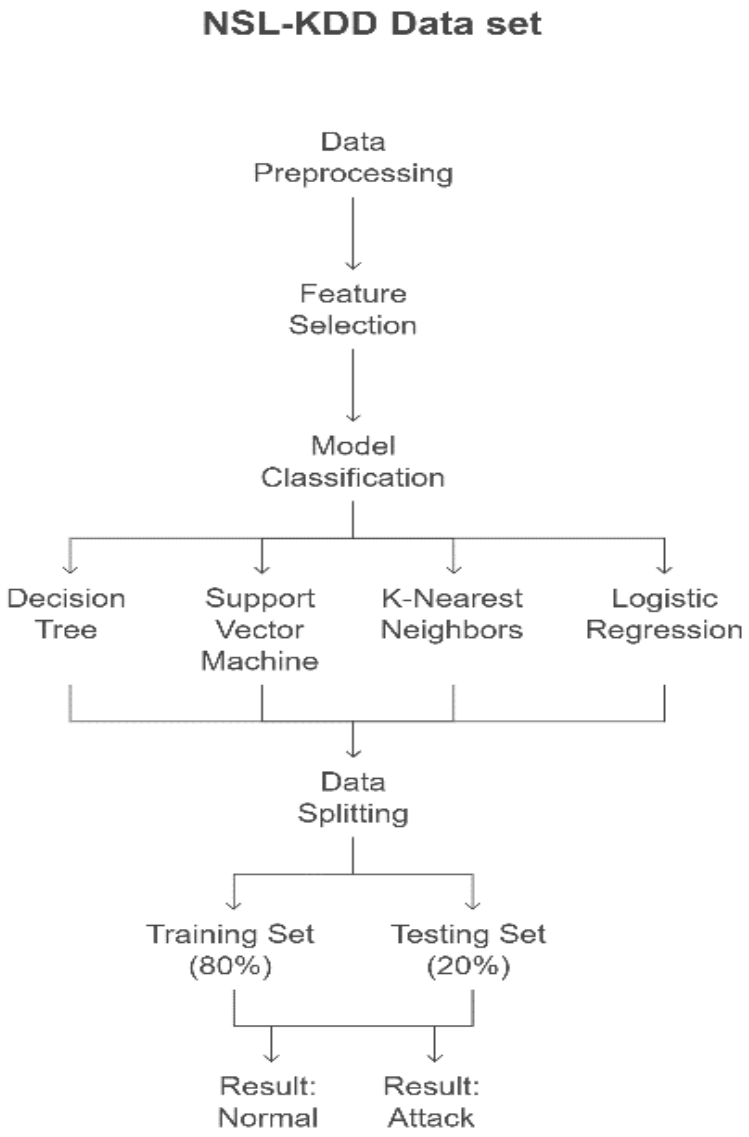

Figure 5. Flowchart of the proposed IDS ML Model

4. Results and Discussion

Table 2 display the classification reports and different performance accuracies obtained from the confusion matrices shown in Figure 6. Different ML models' performances, shown in Figure 7, are measured after they have been trained using the training dataset, validated using the validation datasets, and then tested using the test dataset. displays the various performance accuracies (99.58%, 98.33%, 97.70%, and 91.96%). It has been observed that the LR model has the lowest accuracy among the different models, while the DT model has the highest accuracy.

		Predicted Condition	
Total Pop. = P+N		Pos. (PP)	Neg. (PN)
Actual Cond.	Pos. (P)	TP	FN
	Neg. (N)	FP	TN
Total Pop. = P+N		(P)	(N)

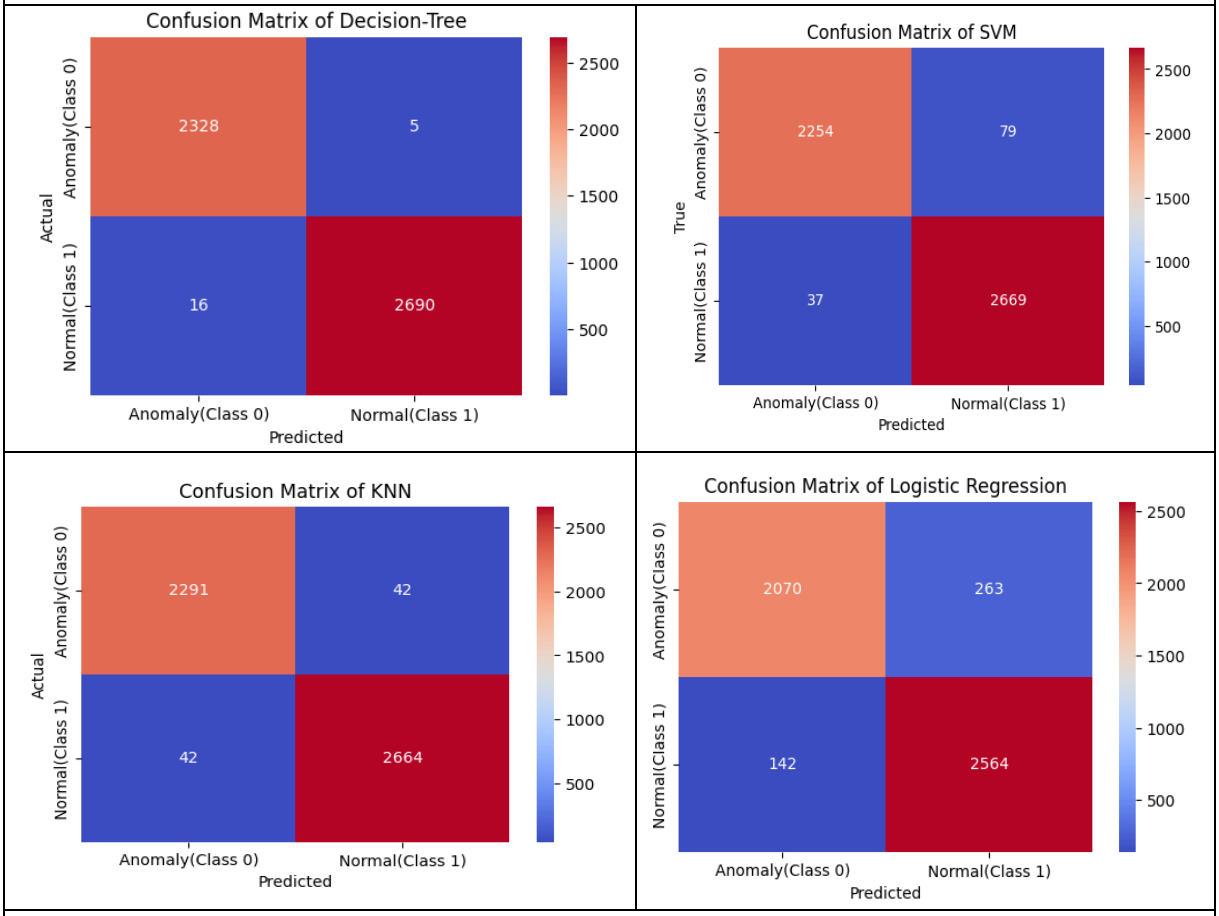

Figure 6. Confusion Matrices of ML models

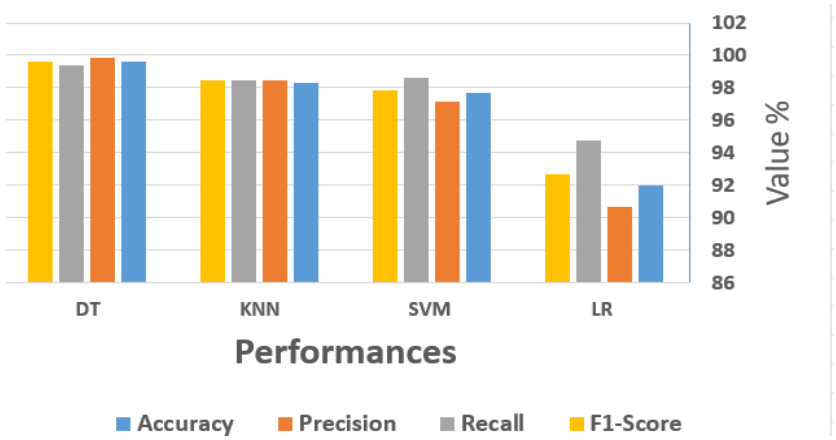

Figure 7. The Different Performance Value

The DT model outperforms SVM, KNN, and LR models for intrusion detection or classification tasks on large, high-dimensional datasets due to its versatility, ability to manage complex linkages and nonlinear feature interactions, and iterative prediction improvement capabilities. It also has high accuracy, scalability to parallel processors with optimized hyper parameters, and the ability to handle outliers or missing data.

Table 2. The Performance Values (%) of ML Models

ML Models	Acc.	Prec.	Recall	F1-sc.
Decision Tree(DT)	99.58	99.81	99.41	99.61
K-Nearest Neighbors (KNN)	98.33	98.45	98.45	98.45
Support Vector Machine(SVM)	97.70	97.13	98.63	97.87
Logistic Regression (LR)	91.96	90.70	94.75	92.68

ROC techniques were used to plot the algorithms as shown in Figure 8. The first algorithm, DT, achieved optimal performance with an AUC of 1.00, while the results of the SVM and KNN algorithms showed 0.99, while the LR algorithm achieved 0.97, indicating strong performance for the results of the studied algorithms.

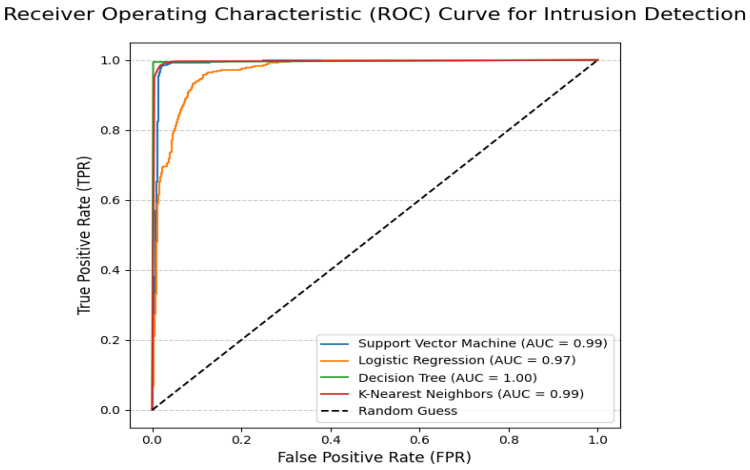

Figure 8. ROC CURVE for intrusion detection system

By comparing the results of the tested models with the results of some ML models included in Table 3, it was found that each of the proposed models was the most efficient among its peers.

Table 3. A Comparison between different ML models based on NSL-KDD Dataset

Ref. No.	Model	Acc. (%)	Model Type
Proposed	**DT**	**99.58**	**Machine Learning (ML) Models**
[32]	DT	99.15	
[33]	DT	98.6	
Proposed	**KNN**	**98.33**	
[34]	KNN	95.5	
[37]	KNN	76.57	
Proposed	**SVM**	**97.70**	
[35]	SVM	96.6	
[37]	SVM	78.14	
Proposed	**LR**	**91.96**	
[36]	LR	68.97	
[38]	LR	81.51	

Logistic regression (LR) often performs less accurately on complex datasets such as NSL-KDD due to a drawback of the following:

- ✓ Logistic regression, a linear classifier, struggles with complex, nonlinear data patterns. More flexible, nonlinear decision boundaries are often required to accurately classify attack types and network behaviors in the NSL-KDD dataset.
- ✓ The NSL-KDD dataset's numerous characteristics interact in intricate ways, making logistic regression less adept at handling these interactions than nonlinear models like DT or neural networks, which can learn intricate feature correlations.
- ✓ Logistic regression might not work as well if the classes are unevenly distributed unless you do things like weight the classes or resample the data.
- ✓ You need really good features to make logistic regression work properly. But with more advanced models like deep learning, you don't need to do as much pre-processing. These models can even choose more complex patterns.

5. Conclusion

This research addresses the Evaluating and comparing the effectiveness of an Intrusion Detection System (IDS) which applies machine learning and proves to be effective in providing accurate results for identifying network intrusion. The system combines both the techniques based on inspect anomalous behavior alongside recognizing known behavior to enhance detection capability for emerging and recurring cyber threats. A combination of machine learning methods (SVM, DT, LR, and KNN) was implemented to classify the network traffic of the standard NSL-KDD dataset. The system is trained and tested, and achieve credible results that correlate with the findings. This research proposes to create an NIDS that considers historical data for initiating preventive measures to stop attacks and or warn a network administrator. A new direction working with datasets is also proposed using non-linearity properties, neural networks, and ensemble methods for intrusion detection. With the explosion of IoT devices, a promising area for investigators would be to research real-time IDS and developing routing technology from a security perspective to secure both data and network communication.

REFRENCES

1. Z. Xue, J. Wang, G. Ding, Q. Wu, Y. Lin, and T. A. Tsiftsis, “Device-to-device communications underlying UAV-supported social networking,” IEEE Access, vol. 6, pp. 34488–34502, 2018.
2. M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, “Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks,” IEEE Trans. Veh. Technol., vol. 68, pp. 641–653, Jan. 2019.
3. G. Gui, H. Huang, Y. Song, and H. Sari, “Deep learning for an effective nonorthogonal multiple access scheme,” IEEE Trans. Veh. Technol., vol. 67, no. 9, pp. 8440–8450, Sep. 2018.
4. Abdulhammed, R., Faezipour, M., Abuzneid, A., & AbuMallouh, A. (2019). Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. *IEEE Sensors Letters*, *3*(1), 1–14.
5. Verma, A., & Ranga, V. (2018). On the evaluation of network intrusion detection systems: Statistical analysis of CIDDS-001 dataset using machine learning techniques. *Pertanika Journal of Science & Technology*, *26*(1), 1307–1332.

6. Z. Iqbal, A. Sajid, M. N. Zakki, A. Zafar, and A. Mehmood, “Role of machine and deep learning algorithms in secure intrusion detection systems (IDS) for IoT & smart cities,” Int. J. Inf. Technol. Res. Appl., vol. 3, no. 4, pp. 1–16, Nov. 2024, doi: 10.59461/ijitra.v3i4.111.
7. R. Alghamdi and M. Bellaiche, “An ensemble deep learning based IDS for IoT using Lambda architecture,” Cybersecurity, vol. 6, no. 1, pp. 1–17, Mar. 2023, doi: 10.1186/s42400-022-00133-w.
8. N. Rane, S. K. Mallick, Ö. Kaya, and J. Rane, “Machine learning and deep learning architectures and trends: A review,” Oct. 2024, doi: 10.70593/978-81-981271-4-3_1.
9. T. Saba, T. Sadad, A. Rehman, Z. Mehmood, and Q. Javaid, “Intrusion detection system through advanced machine learning for the internet of things networks,” IT Prof., vol. 23, no. 2, pp. 58–64, 2021, doi: 10.1109/mitp.2020.2992710.
10. Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, “Kitsune: An ensemble of autoencoders for online network intrusion detection,” 2018, [Online]. Available: https://arxiv.org/abs/1802.09089.
11. C. W. Tsai, T. P. Hong, and G. N. Shiu, “Metaheuristics for the lifetime of WSN: A review,” IEEE Sensors J., vol. 16, no. 9, pp. 2812–2831, 2016, doi: 10.1109/jsen.2016.2523061.
12. S. Latif, Z. Idrees, Z. Zou, and J. A. Drann, “A deep random neural network model for intrusion detection in industrial IoT,” in Proc. 2020 Int. Conf. UK-China Emerging Technol. (UCET), Glasgow, UK, IEEE, 2020, pp. 1–4, doi: 10.1109/UCET51115.2020.9205361.
13. Y. R. Zhao, Y. Liu, D. Wang, W. R. Lv, and J. L. Zhou, “An ANN based sequential detection method for balancing performance indicators of IDS,” Proc. 2019 7th Int. Symp. Comput. Networking (CANDAR), Nagasaki, Japan, 2019, pp. 239–244, doi: 10.19723/j.issn.1671-167X.2019.02.007.
14. N. Moustafa and J. Slay, “A hybrid feature selection for network intrusion detection systems: central points,” 2017, [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1707/1707.05505.pdf.
15. M. M. Baig, M. M. Awais, and E. S. M. El-Alfy, “A multiclass cascade of artificial neural network for network intrusion detection,” J. Intell. Fuzzy Syst., vol. 32, no. 4, pp. 2875–2883, 2017, doi: 10.3233/jifs-169230.
16. S. P. Rm, P. K. R. Maddikunta, M. Parimala, S. Koppu, T. R. Gadekallu, C. L. Chowdhary, and M. Alazab, “An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture,” Comput. Commun., vol. 160, pp. 139–149, 2020, doi: 10.1016/j.comcom.2020.05.048.
17. A. A. Ewees, R. R. Mostafa, R. M. Ghoniem, and M. A. Gaheen, “Improved seagull optimization algorithm using Lévy flight and mutation operator for feature selection,” Neural Comput. Appl., vol. 34, no. 10, pp. 7437–7472, 2022, doi: 10.1007/s00521-021-06751-8.
18. Y. Liu, S. Yang, and G. Li, “A survey of intelligent methods in cyberattack detection for IoT networks,” J. Supercomput., vol. 77, pp. 1–14, 2021.
19. R. Lohiya, A. Thakkar, and A. Thakkar E-Mail, "Intrusion detection using deep neural network with antirectifier layer," Lecture Notes in Networks and Systems, vol. 187, pp. 89–105, 2021.
20. A. Roy and K. J. Singh, "Multi-classification of UNSW-NB15 dataset for network anomaly detection system BT," Proc. Int. Conf. on Communication and Computational Technologies, Singapore, Springer, pp. 429–451, 2021.

21. S. Rajagopal, P. P. Kundapur, and K. S. Hareesha, "A stacking ensemble for network intrusion detection using heterogeneous datasets," Security and Communication Networks, vol. 2020, p. 4586875, 2020, doi: 10.1155/2020/4586875.

22. T. Saba, A. R. Khan, T. Sadad, and S. P. Hong, "Securing the IoT system of smart city against cyber threats using deep learning," Discrete Dynamics in Nature and Society, vol. 2022, p. 1241122, 2022, doi: 10.1155/2022/1241122.

23. Y. Li, Y. Xu, Z. Liu, H. Hou, Y. Zheng, Y. Xin, Y. Zhao, and L. Cui, "Robust detection for network intrusion of industrial IoT based on multi-CNN fusion," Measurement, vol. 154, p. 107450, 2020, doi: 10.1016/j.measurement.2019.107450.

24. H. S. Abdullah, "Artificial intelligence method for cyber security intrusion detection," M.S. thesis, Dept. of Computer Science, AUL, Nov. 2023.

25. E. S. A. Alars and S. Kurnaz, "Enhancing network intrusion detection systems with combined network and host traffic features using deep learning: deep learning and IoT perspective," Discov. Comput., vol. 27, p. 39, 2024, doi: 10.1007/s10791-024-09480-3.

26. M. N. Chohan, U. Haider, M. Y. Ayub, H. Shoukat, T. K. Bhatia, and M. F. Ul Hassan, "Detection of cyber attacks using machine learning based intrusion detection system for IoT based smart cities," EAI Endorsed Trans. Smart Cities, vol. 7, no. 2, p. e4, 2023, doi: 10.4108/eetsc.3222.

27. V. Pai et al., "Comparative analysis of machine learning algorithms for intrusion detection," IOP Conf. Ser.: Mater. Sci. Eng., vol. 1013, p. 012038, 2021.

28. A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A Novel Hierarchical Intrusion Detection System Based on Decision Tree and Rules-Based Models," in Proc. 15th Int. Conf. on Distributed Computing in Sensor Systems (DCOSS), Santorini, Greece, 2019, pp. 228-233, doi: 10.1109/DCOSS.2019.00059.

29. M. A. Uddin, M. M. Islam, M. A. Talukder et al., "Machine learning based diabetes detection model for false negative reduction," Biomed. Mater. Dev., pp. 1–17, 2023.

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

31. A. Hazem Salim, S. Salim Abdullah and M. Mario, “Exploring Machine Learning Techniques for Questionnaire Analysis”, International Journal of Informatics and Data Science Research, vol. 2, no. 7. Scientific Bulletin, pp. 33–45, Jul. 21, 2025. doi: 10.5281/zenodo.16752613.

32. A. Pathak and S. Pathak, "Study on Decision Tree and KNN Algorithm for Intrusion Detection System," Int. J. Eng. Res. and, vol. 9, no. 05, May 2020, doi: 10.17577/IJERTV9IS050303.

33. A. S. Ahanger, S. M. Khan, and F. Masoodi, "An Effective Intrusion Detection System using Supervised Machine Learning Techniques," in 2021 5th Int. Conf. on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1639–1644, doi: 10.1109/ICCMC51019.2021.9418291.

34. S. Priya and K. Kumar, “Feature selection with deep reinforcement learning for intrusion detection system,” Computer Systems Science and Engineering, vol. 46, pp. 1–15, 2023, doi: 10.32604/csse.2023.030630

35. S. Rastogi, A. Shrotriya, M. K. Singh, and R. V. Potukuchi, “An analysis of intrusion detection classification using supervised machine learning algorithms on NSL-KDD dataset,” Journal of Computing Research and Innovation, 2022.

36. I. Priyadarshini, “Anomaly detection of IoT cyberattacks in smart cities using federated learning and split learning,” Big Data Cogn. Comput., vol. 8, no. 3, p. 21, 2024, doi: 10.3390/bdcc8030021.

37. K. Z. Haider, Q. G. K. Safi, M. Awais, U. Fatima, and M. M. Iqbal, “Evaluating machine learning-based intrusion detection in software defined networks using NSL-KDD dataset,” Journal of Computing & Biomedical Informatics, vol. 9, no. 2, 2025.

38. A. Mahfouz, A. Abuhussein, D. Venugopal, and S. Shiva, “Ensemble classifiers for network intrusion detection using a novel network attack dataset,” Future Internet, vol. 12, no. 11, p. 180, 2020, doi: 10.3390/fi12110180.