Journal of Engineering and Sustainable Development
https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd
https://doi.org/10.31272/jeasd.28.3.4

Vol. 28, No. 03, May 2024
ISSN 2520-0917

Original Research

NEW STRATEGIES FOR IMPROVING NETWORK SECURITY AGAINST CYBER ATTACK BASED ON INTELLIGENT ALGORITHMS

Mahmood Zaki Abdullah[1*], Ali Khalid Jassim[2], Fadia Noori Hummadi[3], and Mohammed Majid M. Al Khalidy[4]

[1]*Computer Engineering Department, Mustansiriyah University, Baghdad, Iraq*
[2]*Electrical Engineering Department, Mustansiriyah University, Baghdad, Iraq*
[3]*Biomedical Engineering Department, Al-Khwarizmi College of Engineering, University of Baghdad, Baghdad, Iraq*
[4]*Department of Electrical and Electronics, College of Engineering, University of Bahrain, Manama, Kingdom of Bahrain*
[1]https://orcid.org/0000-0002-3191-3780
[2]https://orcid.org/0000-0002-4146-4536
[3]https://orcid.org/0000-0002-8179-5305
[4]https://orcid.org/0000-0002-9723-0405

Received 27/10/2023 ***Revised 22/03/2024*** ***Accepted 04/04/2024***

Abstract: Gradually, since the number of linked computer systems that use networks linked to the Internet is raised the information that is delivered through those systems becomes more vulnerable to cyber threats. This article presents proposed algorithms based on Machine Learning (ML) that ensure early detection of cyber threats that cause network breaking through the use of the Correlation Ranking Filter feature selection method. These proposed algorithms were applied to the Multi-Step Cyber-Attack Dataset (MSCAD) which consists of 66 features. The proposed strategy will apply machine learning algorithms like Adaptive Boosting-Deep Learning (AdaBoost-Deep Learning) or (ABDL), Multi-Layer Perceptron (MLP), Bayesian Networks Model (BNM), and Random Forest (RF), the feature would be decreased to high valuable of 46 features were included with a threshold of 0.1 or higher. The accuracy would be increased when the no. of features decreased to 46 with a threshold of ≥ 0.1 with the ABDL algorithm producing an accuracy of 99.7076%. The obtained results showed that the proposed algorithm delivered a suitable accuracy of 99.6791% with the ABDL algorithm even with a higher number of features.

Keywords: *Adaptive Boosting; Multi-Step Cyber-Attack Dataset; Cyber-Attacks; Internet Access; Mitigation*

1. Introduction

In the realm of network security, Intrusion Detection Systems (IDSs) stand as pivotal components, divided into three key categories: misuse-based, anomaly-based, and specification-based. Misuse-based IDSs excel at recognizing familiar attack patterns by comparing them with stored references, yet they grapple with emerging threats[1]. On the other hand, the anomaly-based IDS model aims at exposing unexpected actions and this gives a chance to discover new types of attacks through evaluation of real activities occurring [2].

The other type is the specification-based IDS where rules or patterns of normal network behavior are set as an example. Any deviation from these defined notions is flagged, which means security issues are found as soon as possible. It implies that this mechanism can detect the new unknown patterns of attacks performing the proactive measure [3].

**Corresponding Author*:
drmzaali@uomustansiriyah.edu.iq

The utilization of machine learning algorithms is a cardinal element of IDSs because the algorithms bring more capabilities to machine recognition. The main aim here is to improve the system's ability in the process of detecting and identifying different types of attacks. However, consider the fact that each algorithm has strengths and weaknesses in terms of specific attack types.

The ID system depends also on both its learning model and the dataset it uses. In the past, datasets were the source of the problem of old and bogus data that led to concerns about privacy.

In response to these challenges, the researchers were introduced to the MSCAD, which encompasses two multi-step attack categories.

Additionally, the dataset also trained the IDS with high performance, they were considered robust with a G-mean of 0.83 and fulfilled the twelve important factors for evaluation required of the dataset [4].

This article would enhance intrusion detection by applying machine learning techniques and a thorough feature selection process based on the up-to-date public dataset.

Such an approach performs the identification much faster and speeds up the reaction time of the ID system.

Features are selected from MSCAD using the Correlation Ranking Filter method, and the impact of feature selection on a variety of machine learning algorithms, including AB, MLP, BNM, and RF, is comprehensively evaluated. Additionally, this article conducts an extensive assessment of various machine learning classification algorithms within MSCAD to pinpoint the most suitable classifier and the optimal set of features.

2. Literature Survey

Generally, many researchers have made the best efforts to produce and train various algorithms depending upon machine learning methods within the field of intrusion detection and cyber security.

All previous research in this field focused on developing smart methods for detecting cyber threats and trying to prevent them from affecting network systems connected to the Internet, in addition to preventing them and blocking their impact. A brief survey of previous research in this field is presented below:

Muna A., et al, in 2018 [5], presented a deep learning method involved in analyzing network packet delivery, which is created from TCP/IP protocols data units. The training was done by an encoder and a full-feed forwarded neural network scheme, with testing carried out on realized networked datasets like NS-KDD. The results of the analysis created an improvement and accuracy in detecting threats.

Khan, F. A., et al, in 2019 [6], utilized the ability of an integrated autoencoder of features extraction, which made good classifiers. They used the maximum capabilities of softmax to improve the procedure of classification, they also produced a suitable accuracy in the dataset NS - KDD.

Zhong Y., et al, in 2020 [7], improved an attacker detection system using a hybrid deep learning method. They used a statistical method to search for network characteristics, tested a special encoder to realize a network threat, and further improved the accuracy. Their high-efficiency detection algorithm gives a good performance. detection algorithm gives a good performance.

Mohammad R., et al,2021 [8], introduced a good feature selection method which is called highest wins, for establishing accurate IDS. The highest wins are tested on large scaling datasets and established methods like chi-square and information gain in terms of performance and feature reduction. It is also applied to the NS-KDD dataset with more convenient results and reduced the time of building the model.

It is noted that most researchers have concluded and reached a clear fact that most cyber threats and information intrusion occur in routers because they connect many networks and because they are always in the state of sending many information packets and their routing tables include all the IP addresses of neighboring and distant networks, which is more vulnerable to attack, in addition to threats that could infect servers.

Venkatesan S., 2023 [9], introduced many methods based on using machine learning to carry out a reliable intrusion detection system. most of the research works including SVM, Random Forests, and Decision Trees, to enhance the attack detection accuracy using selected features from the NSL-KDD dataset.

It can be shown from Table 1, that most of the researchers applied outdated datasets. In contrast, this article employs the latest cyber-attack dataset, essential in the face of evolving cyber threats. By using current data, the goal of this research is to enhance the intrusion detection methods based on intelligent algorithms to protect the linked networks with their hosts against dynamic cyber security challenges.

Table 1. Summary of Intrusion Detection Studies

Authors	Methodology	Datasets Used	Key Findings
Muna AL Hawawreh (2018) [5]	Deep learning with autoencoder and feed-forward neural network	An NS-KDD, and UNSWNB15.	Improved accuracy, minimal false alarms
Farrukh A. Khan (2019) [6]	Deep Stacked auto encoder for feature extraction.	An UNSW-NB15, and NS-KDD,	Significant improvement in classifications, notable accuracy rates
Zhong Y., et al. (2020) [7]	Damping Incremental Statistic algorithms, autoencoder, weighted LSTM	MAWILab dataset	Good performance compared to state-of-the-art methods
Mohammad R., et al. (2021) [8]	'Highest Wins' (HW), chi-square, and information gain feature selection method	NS-KDD	Outperformed chi-square and information gain, promising results in accuracy and reduced model-building time
Venkatesan S., (2023) [9]	Various machine learning algorithms (SVM, Random Forests, Decision Trees)	NS-KDD	Evaluating machine learning algorithms for improved attack detection accuracy

3. Methodology

The proposed method would include many procedural stages like data preparation, feature selection, classifications, and performance evaluation, setting up an overall strategy to heading the research problem, Fig. 1 illustrates the procedural steps of the proposed research.

3.1 Preparation of Dataset

Within the period of this work, the Multi-Step Cyber-Attack Dataset was between the best and latest datasets in the field of (IDS) [4]. The MSCAD contains 128,799 samples and comprises 66 features that contain six groups of cyber-attacks, as shown in Fig.2.

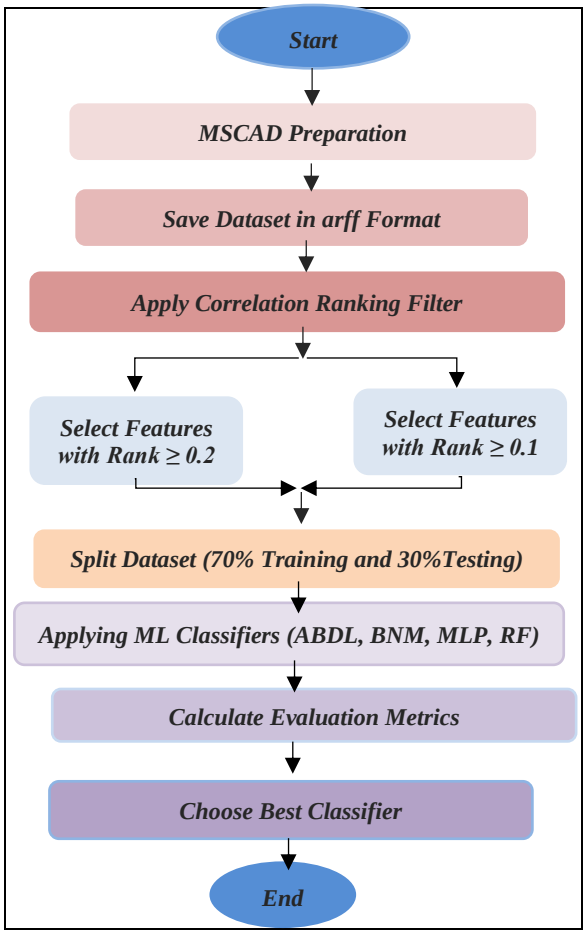

Figure 1. Proposed Method

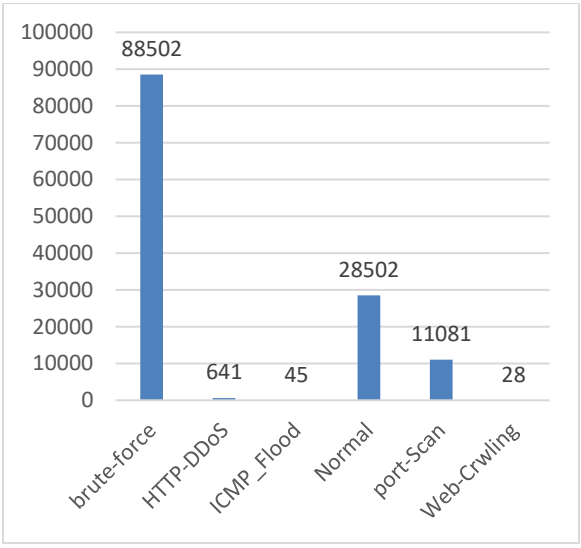

Figure 2. MSCAD Count

The stage of preparation of the dataset involved a meticulous process to adapt the MSCAD for analysis within the Weka program. It began with overall data cleaning to classify the missing and incompatible values. Data transformations were applied to make the dataset compatible with the Weka program. Feature selection simplified the dataset, and it was split into a 70% training set and a 30% testing set. The dataset obtained after the preparation process was saved in (arff) format and loaded to the Weka program for analysis. This procedure made the MSCAD dataset more suitable for machine learning analysis in the Weka program.

3.2. Features Selection Stage

In the stage of feature selection, MSCAD used the Correlation Ranking Filter technique for the feature selection process, this method assigned a rank to each feature according to its weight [10]. As an initial phase, the no. of selected features would be 46, increasing the threshold rank to 0.1 or higher. Later on, the selected features were reduced to cover 31 features, requiring a threshold rank of 0.2 or higher.

3.3 Classifiers Used

The up-to-date dataset of MSCAD was used in this article to evaluate the performance of the four algorithms of machine learning classifiers, these classifiers were chosen because of their beneficial characteristics and ability to identify particular kinds of cyber attacks [11].

Adaptive Boosting Deep Learning (ABDL), is an iterative layered intelligence technique that divides the information into subsets based on features to arrive at conclusions, this iterative intelligence makes it a more accurate method as compared with the other methods [12].

Multi-Layer Perceptron (MLP), MLP is a sort of artificial neural network that makes a good decision at handling intricate patterns. It is made up of numerous layers of linked nodes or neurons [13].

Bayesian Networks Model (BNM), the Bayesian method depends on the foundation of the probabilistic algorithm known as Bayesian Networks [14].

Random Forest, is a learning technique that combines different output decisions to achieve a good convenient accuracy and fitting. It has the ability of used with large-scaling datasets [15] [17].

All the mentioned classifiers would be tested to choose the suitable one for the intrusion detection system [18].

3.4 Evaluation Matrix

This article applied a lot of assessments for the metrics according to the confusion matrix shown in Table 2. It applied equations 1 to 4 to make all calculated metrics [6]. It is important to give a short definition for each metric as stated below.

Accuracy: This represents the proportion of the corrected sample occurrence among all samples as shown in equation (1).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: This represents the correct computation of proportional estimated positive samples to all positive expected instances as shown in equation (2).

$$Pr = \frac{TP}{TP+FP} \quad (2)$$

Sensitivity: It is also known as a (Recall), it calculates the capability of the method to differentiate all relevant instances from the real positive samples as shown in equation (3).

$$Se = \frac{TP}{TP+FN} \quad (3)$$

F-measure: Represents the integration of Sensitivity (Recall) and Precision into one item as shown in equation (4).

$$F - measure = \frac{2\times se\times pr}{se+pr} \quad (4)$$

Table 2 Confusion Matrix

	Predicted class	
Actual class	TP	FN
	FP	TN

- TP (True Positive): Represents the no. of samples that are correctly estimated as positive.
- FN (False Negative): Represents the no. of samples that are incorrectly estimated as negative.
- FP (False Positive): Represents the no. of samples that are incorrectly estimated as positive.
- TN (True Negative): Represents the no. of samples that are correctly estimated as negative. [19-23]

4. Results and Analysis

After choosing the most important features that were found in the Multi-Step Cyber-Attack Dataset using the Correlation Ranking Filter, the aim was to make the system more accurate at detecting cyber threats. The results were divided mainly into two parts based on feature selection thresholds of ranking ≥ 0.1, and ranking ≥ 0.2 to show the effect on the performance of the classifiers.

4.1 Features with Ranking ≥ 0.1

Started by the consideration of features with a ranking of ≥ 0.1, as shown in Fig. 3, which contained a large scale of characteristics. Then these features were tested using the four machine-learning classifiers. The performance evaluation of the systems expanded all standard classification metrics like True Positive, False Positive, Precisions, Recalls, F1-Score, and ROC area as shown in Figs 4,5,6, and 7.

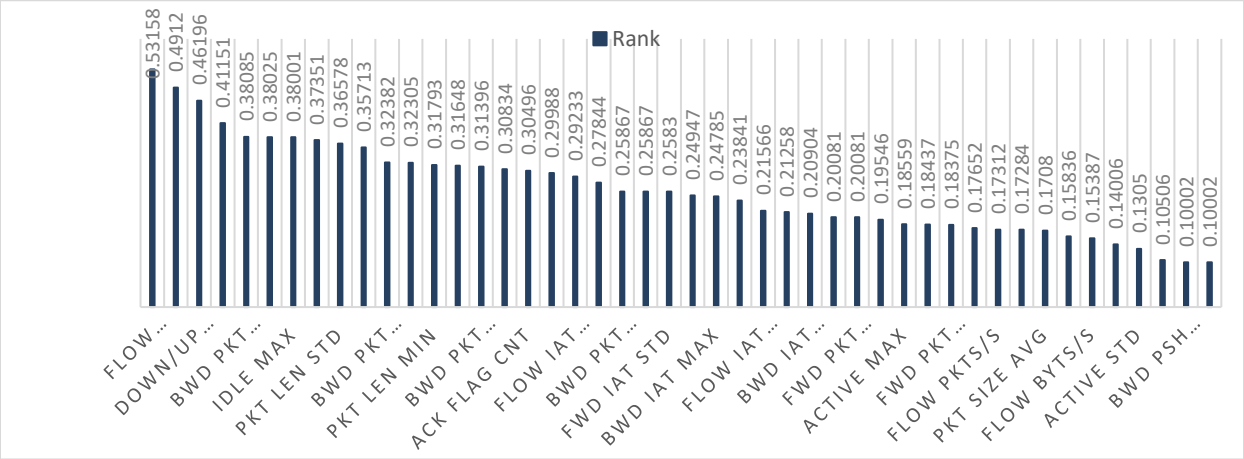

Figure 3. Features with Rank ≥ 0.1

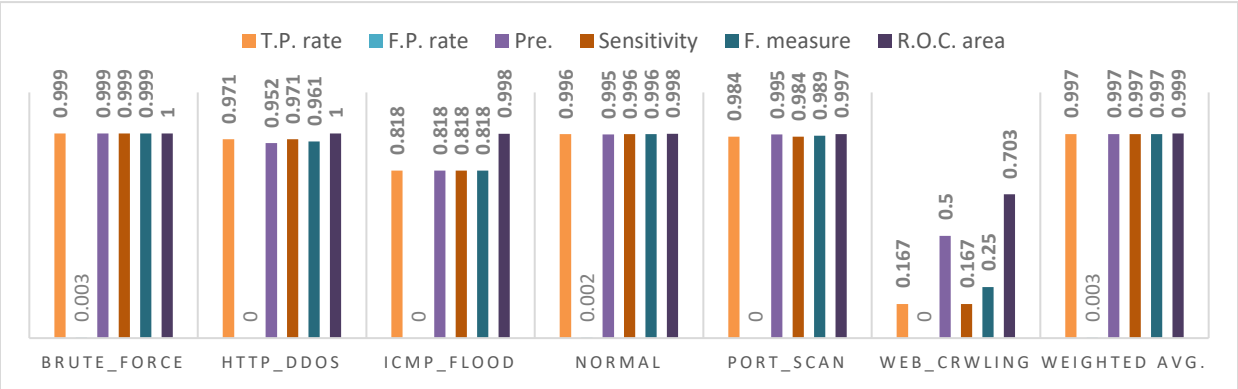

Figure 4. ABDL Performance Metrics (Rank ≥ 0.1)

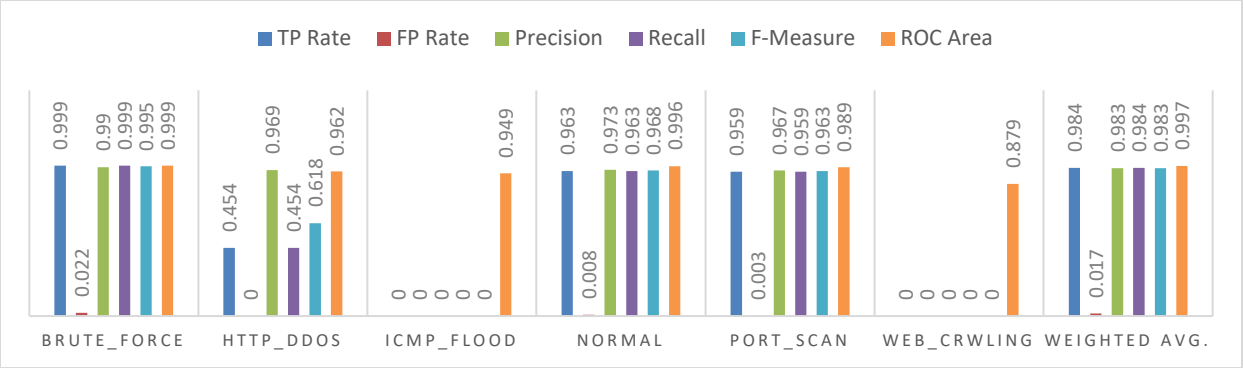

Figure 5. MLP Performance Metrics (Rank ≥ 0.1)

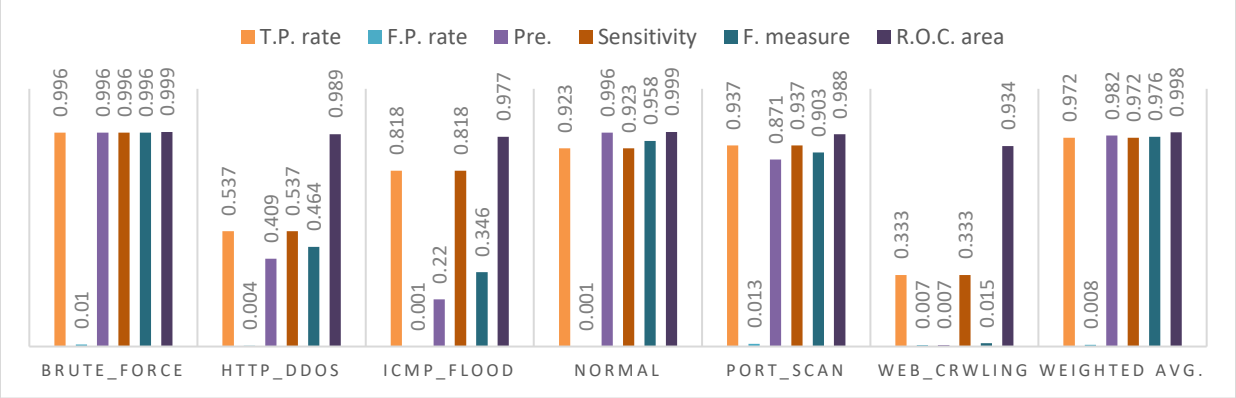

Figure 6. BNM Performance Metrics (Rank ≥ 0.1)

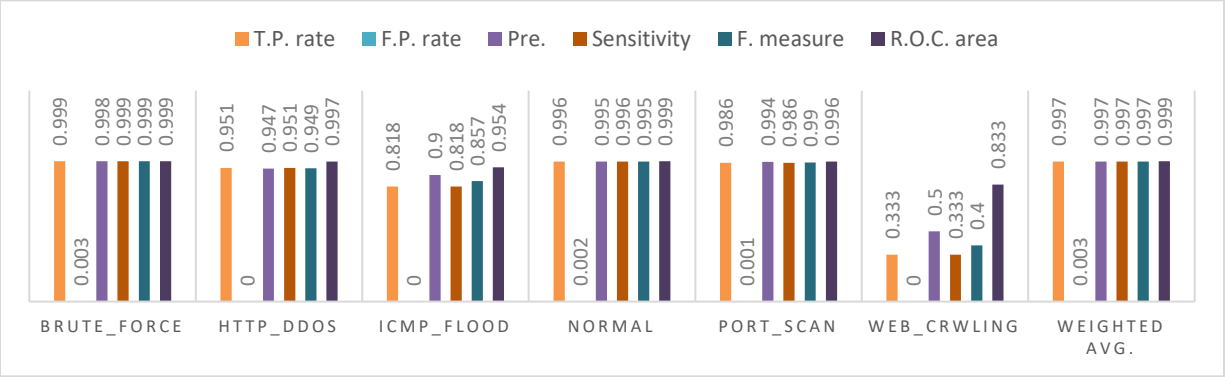

Figure 7. RF Performance Metrics (Rank ≥ 0.1)

4.1 Features with Rank ≥ 0.2

This section focuses on the selected features that had a valuable impact with a threshold rank of greater equal to 0.2 as shown in Fig.8. These can be considered as important attributes that the system required to mark a cyber-attack.

The same ML classifiers were used to test their performance with these Features The evaluation of the performance involved the use of common classifications metrics, containing True Positive (TP), False Positive (FP), Precision, Sensitivity, F-Measure, and ROC Area, are shown in Fig. 9, 10, 11, and 12.

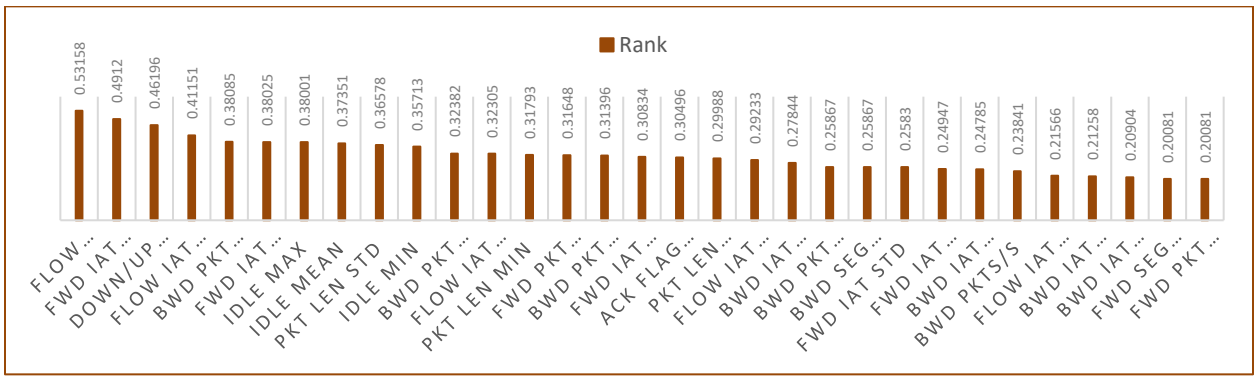

Figure 8. Features with Rank ≥ 0.2

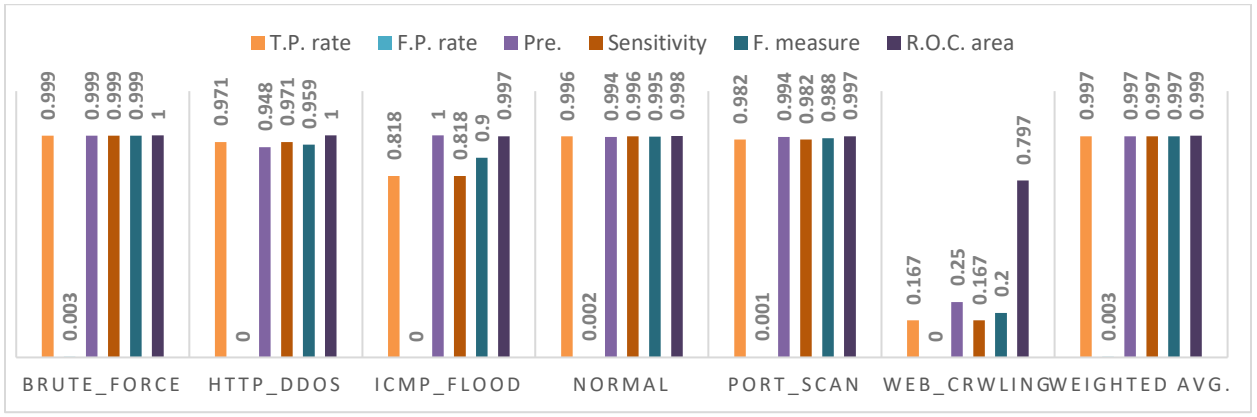

Figure 9. ABDL Performance Metrics (Rank ≥ 0.2)

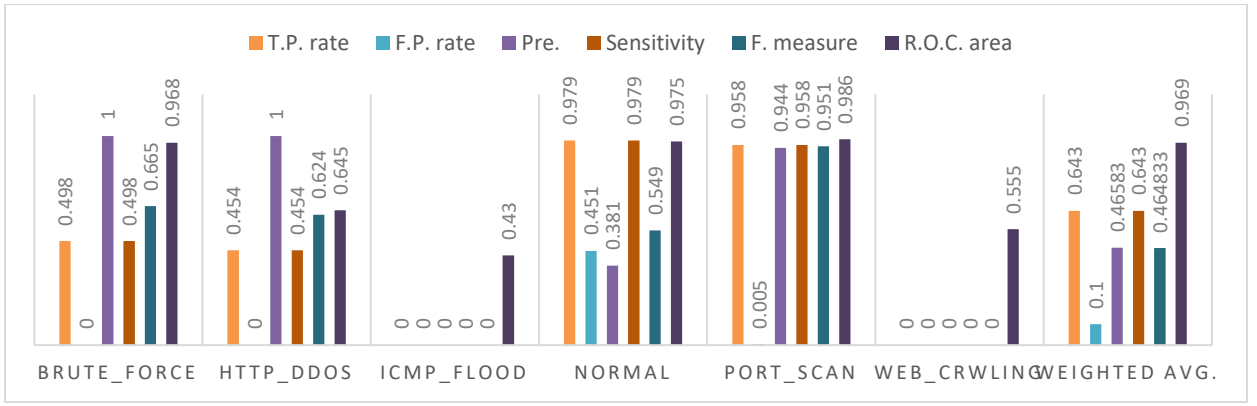

Figure 10. MLP Performance Metrics (Rank ≥ 0.2)

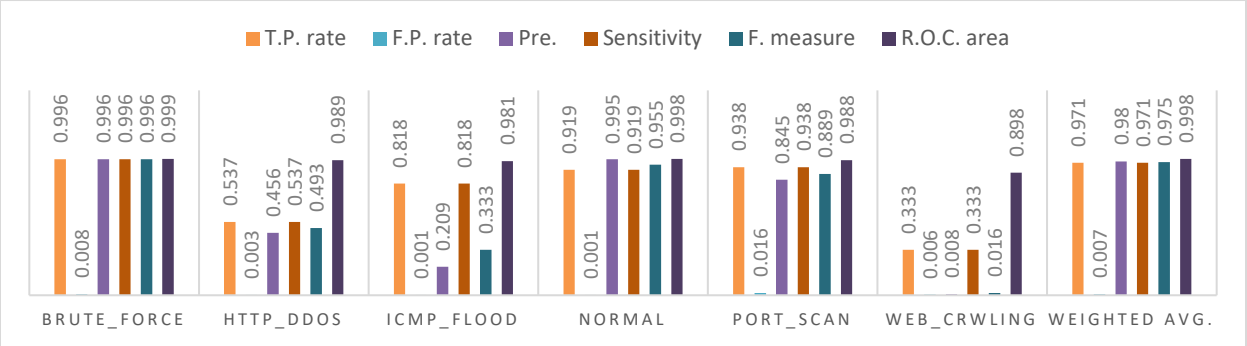

Figure 11. BNM Performance Metrics (Rank ≥ 0.2)

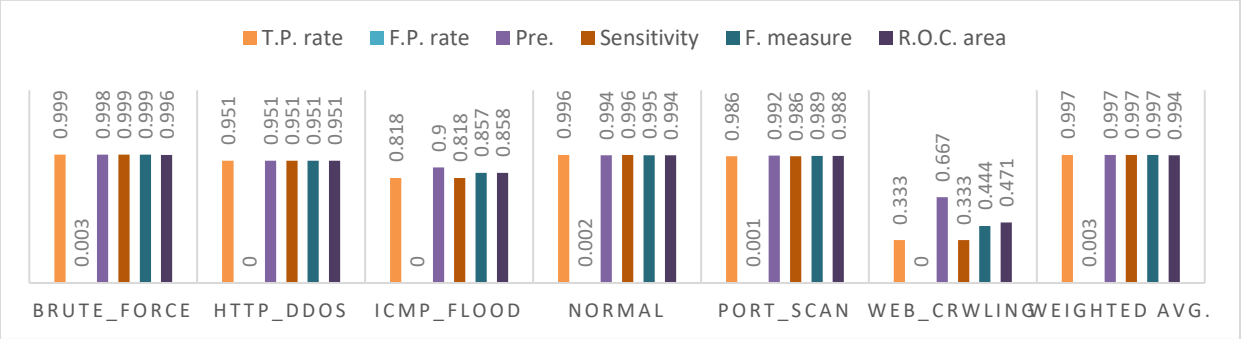

Figure 12. RF Performance Metrics (Rank ≥ 0.2)

The weighted average performance metrics of machine learning algorithms can be shown in Fig. 13 and Fig. 14. Fig 13 illustrates the results of the chosen features with a threshold rank greater or equal to (0.1), while Fig.14 shows the results of features based on a threshold rank of greater or equal to (0.2). These obtained results provide a good visualization of the overall performance of proposed feature selection methods.

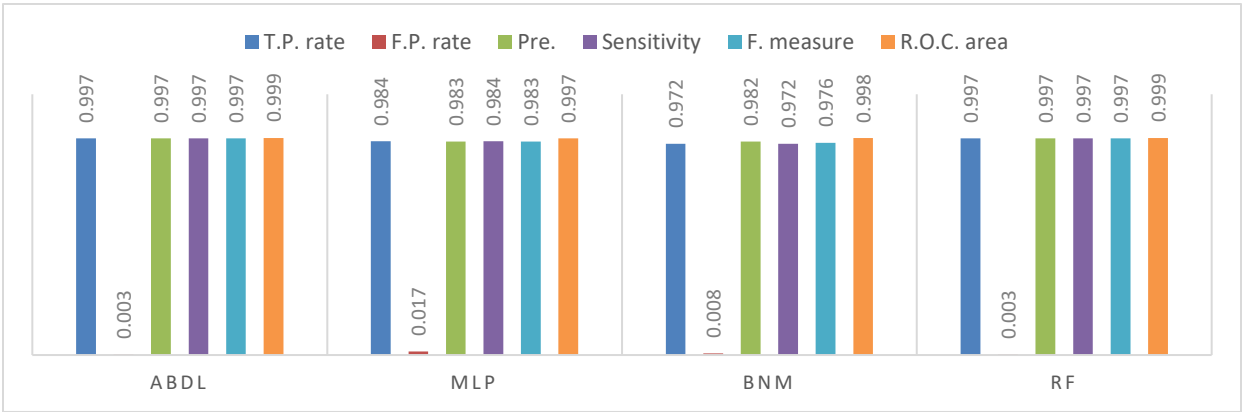

Figure 13. Weighted average performance metrics of machine learning models (Rank ≥ 0.1)

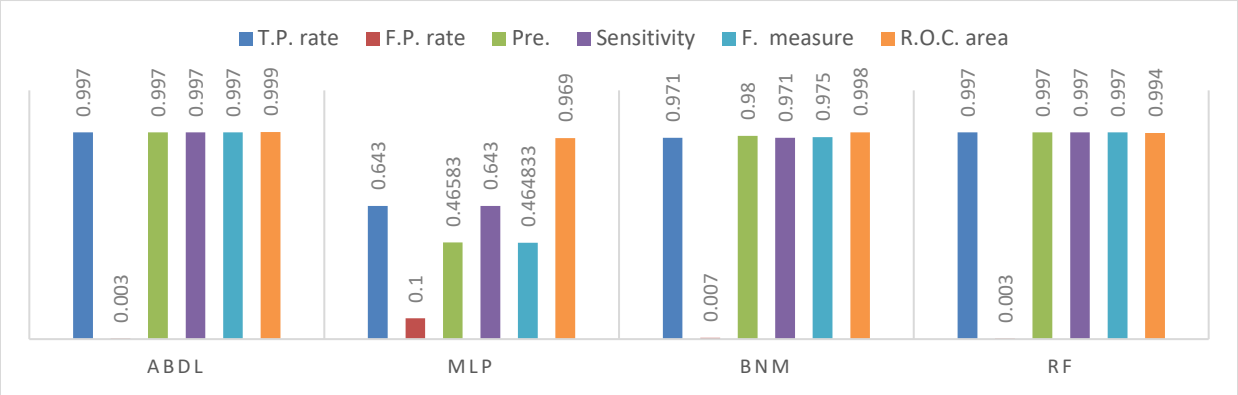

Figure 14. Weighted average performance metrics of machine learning models (Rank ≥ 0.2)

This article implemented different methods to identify cyber threats. Four algorithms were applied, ABDL, RF, MLP, and BNM. The comparison focused on their performance using two types of rules, one with features threshold above 0.1 and another with features threshold above 0.2. The selected rules remarkably affected how well the methods performed and how they quickly processed data. ABDL and RF still obtained a high accuracy even when using less strict rules for choosing features. However, MLP did not perform as well when the rules were less strict. BNM permanently performed well, regardless of the rules applied. Choosing the most accurate method depends on specific needs due to the trade-off between speed and accuracy. In this article, ABDL with strict rules (rank above 0.2) came out as the best choice according to its balanced performance in terms of time and accuracy.

In summary, this article highlights the significance of choosing algorithms and features depending on specific rules. ABDL method especially for strict rules (rank above 0.2) appears to be a good choice for identifying attacks with a good balance between time and accuracy as shown in Table 3.

Table 3. Accuracy and Time Comparison for Four Classifiers at Different Feature Rank Thresholds

Classifier	Feature Selection	RMSE	Accuracy %	Build Time (s)	Test Time (s)
ABDL	Rank ≥ 0.2	0.0308	99.7076	35.69	0.09
	Rank ≥ 0.1	0.0321	99.6791	17.29	0.05
MLP	Rank ≥ 0.2	0.0688	98.4369	1087.15	0.37
	Rank ≥ 0.1	0.3394	64.3427	611.63	0.53
BNM	Rank ≥ 0.2	0.091	97.2464	7.92	0.45
	Rank ≥ 0.1	0.0935	97.1273	5.83	0.53
RF	Rank ≥ 0.2	0.03	99.6843	184.62	1.08
	Rank ≥ 0.1	0.0309	99.4648	184.62	1.08

Table 4 gives a summary of large-scale research works, the use of Feature Selection algorithms applied classification algorithms, the use of datasets, and the suitable accuracy obtained in each work. The proposed algorithm conforms to its impact as it produced high accuracy when applied to the dataset MSCAD, this result shows the activity of the correlation ranking filter and the classification methods used in the research.

Table 4. Summary of Feature Selection and Classification Performance in Research Studies

Ref.	Algorithm of Selection Feature	Algorithms of Classification	Dataset	Analyses Measurements
Muna AL Hawawreh, in 2018[5]	Autoencoder	Deep Neural Network	An NS-KDD, and UNSWNB15.	A good accuracy was obtained in NSL KDD at 98.6%, and UNSW-NB15, accuracy was 92.4%
Farrukh A. Khan, in year 2019 [6]	Deep Stack Autoencoder	Softmax	An UNSW-NB15, and NSLKDD,	A good accuracy obtained in NSLKDD is 98.6%, and UNSWNB15, accuracy was 92.4%
Ying Z., and their team, in 2020[7]	Autoencoder	HELAD	MAWILab dataset	Better accuracy
Mohammad R. et al, 2021 [8]	Highest Wins (HW), chi-square, and information gain.	naive Bayes and decision tree	NS-KDD	99.56% with decision tree
Venkatesan S., 2023 [9]	Choose critical features that improve the accuracy	SVM, Random Forests, and Decision Trees	NS-KDD	99% using decision tree
The proposed method	Correlation Ranking Filter	AB, MLP, BNM, and FR algorithms	MSCAD	99.7076% using ABDT with Rank ≥ 0.2

5. Conclusions

This research illustrates the importance of feature selection and its significant impact on the effectiveness of systems used to identify cyber threats when they attack conventional networked systems. It is clear that by implementing the correlation ranking filter method an improvement in the accuracy of the classifiers would occur especially with the ADBL algorithm which produced more accurate results when compared with the other algorithms. The findings confirm a relation between the precision of identification and the computational resources needed. The choice of feature selection criteria plays an important role in achieving accurate and fast detection of attacks. ABDL with Rank ≥ 0.2 appeared as an excellent choice providing a convenient compromise between accuracy and system performance. In the future, there will be chances to improve feature selection techniques, examine advanced algorithms, and carry out in-depth research with a variety of the latest datasets. In the future research can take into consideration real-time applications that adjust the system against new cyber threats and maximize computational efficiency without losing accuracy. Additionally, exploring the integration of artificial intelligence and machine learning methodologies for adaptive feature selection may enhance the flexibility of cyber threats detection and prevention systems even further.

Conflict of interest

The authors confirm that there is no conflict in publishing this article.

Author Contribution Statement

Mahmood Zaki Abdullah proposed the idea and algorithms of the research. Ali Khalid Jasim, Fadia Noori Hummadi, and Mohammed Majid M. Al Khalidy participated in implementing calculations and algorithms.

References

1. Chaloop, S. G., Abdullah, M. Z., Enhancing Hybrid Security Approach Using AES And RSA Algorithms. J. Eng. Sustain. Dev., Vol. 25, Issue 4, pp. 58–66, 2021. https://doi.org/10.31272/jeasd.25.4.6

2. Sharipuddin S., Winanto E., Purnama B., Kurniabudi K., Stiawan D., Hanapi D., Bin Idris M., Kerim B., and Budiarto R., Enhanced Deep-Learning Intrusion Detection in IoT Heterogeneous Network With Feature-Extraction, Indones. J. Electr. Eng. Informatics, Vol. 9, Issue 3, pp. 747–755, 2021. https://doi.org/10.52549/.V9I3.3134.

   Shah, A. H., Miry, A. H., Salman, T. M. Automatic Modulation Classification Using Deep Learning Polar Feature. J. Eng. Sustain. Dev., Vol. 27, Issue 4, pp. 477–486,2023. https://doi.org/10.31272/jeasd.27.4.5

3. Almseidin, M., Al-Sawwa, J., Alkasassbeh, M. Generating a benchmark cyber multi-step attacks dataset for intrusion detection. J. Intell. Fuzzy Syst., Vol. 43, Issue 3, pp. 3679–3694, 2022. https://doi.org/10.3233/JIFS-213247

4. Muna, A.-H., Moustafa, N., Sitnikova, E. Identification of malicious activities in industrial Internet of Things based on deep learning models. J. Inf. Secur. Appl., Vol. 41, pp. 1–11, 2018.

5. https://doi.org/10.1016/j.jisa.2018.05.002

6. Khan, F. A., Gumaei, A., Derhab, A., Hussain, A. A novel two-stage deep learning model for efficient network intrusion detection. IEEE Access, Vol. 7, pp. 30373–30385, 2019. https://doi.org/0.1109/ACCESS.2019.2899721

7. Zhng Y., Chen W., Wang Z., Chen Y., Wang K., Li Y., Yin X., Shi X. Yang J., and Li K., A Novel Network Anomaly Detection Model Based on Heterogeneous Ensemble Learning, Comp. Net., vol. 169, 2020. https://doi.org/10.1016/j.comnet.2019.107049

8. Mohammad R., and Alsmadi, M., Intrusion Detection Using Highest-Wins Feature Selection Algorithm, Neural Computer. Appl., Vol. 33, p. 9805 – 9816, 2021. https://doi.org/10.1007/s00521-021-05745-w

9. Venkatesan, S. Design an intrusion detection system based on feature selection using ML algorithms. Math. Stat. Eng. Appl., Vol. 72, Issue 1, pp. 702–710. 2023.https://doi.org/10.17762/msea.v72i1.2000

10. Mwadulo, M., A Review on Feature Selection Methods for Classification Tasks, International Journal of Computer Applications Technology and Research, vol. 5, Issue 6, p. 395–402, 2016. http://dx.doi.org/10.7753/IJCATR0506.1013

11. Bagaa, M., Taleb, T., Bernabe, J. B., Skarmeta, A. A machine learning security framework for IoT systems. IEEE Access, vol. 8, 114066–114077. 2020. https://doi.org/10.1109/ACCESS.2020.2996214

12 Sornsuwt P., and Jaiyn S. A New Hybrid Machine Learning for Cyber Security Threat Detection Based on Adaptive Boosting, Applied Artifecial Intelligence,

Vol. 33, no.5, p. 462 – 482, 2019. https://doi.org/10.1080/08839514.2019.1582861

13 Wang M., Lu Y., and Qin J., A Dynamic MLP Based DDoS Attack Detection Method Using Feature Selection and Feedback, Comput. Secur., Vol. 88, 2020. https://doi.org/10.1016/j.cose.2019.101645

14 Alexander R. Reducing Threats by Using Bayesian Networks to Prioritize and Combine Defense in Depth Security Measures, Journal of Information Security, vol. 11, no. 3, 2020. https://doi.org/10.4236/jis.2020.113008

15 Ahuja N., G. Singal, D. Mukhopadhyay, and N. Kumar, "Automated DDOS attack detection in software defined networking," J. Netw. Comput. Appl., vol. 187, no. November 2020, p. 103108, 2021. https://doi.org/10.1016/j.jnca.2021.103108.

16. Joshi, M., Hadi, T. H. A Review of Network Traffic Analysis and Prediction Techniques, 2015. https://doi.org/10.48550/arXiv.1507.05722

17. Khammas B., Ransomware Detection using Random Forest Technique, ICT Express 6(4):325-331, December 2020. http://dx.doi.org/10.1016/j.icte.2020.11.001

18. Dhanya K., Sulakshan V., Kartik S., Kumard, T., and Gireesh K., Detection of Network Attacks using Machine Learning and Deep Learning Models, Procedia Computer Science, vol. 218, pp. 57–66, 2023. https://doi.org/10.1016/j.procs.2022.12.401

19. Zeljko Đ., Classification Model Evaluation Metrics, (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 12, Issue. 6, 2021. https://doi.org/10.14569/IJACSA.2021.0120670

20. Zaheer A., and Seunghwan M., Enhancing Industrial Cybersecurity, Focusing on Formulating a Practical Strategy for Making Prediction through Machine Learning Tools in Cloud Computing Environment, electronics, 12 (12), 2650, June 2023.https://doi.org/10.3390/electronics12122650

21. Cavusoglu U., Akgun D., and Hizal S., A Novel Cyber Security Model Using Deep Transfer Learning, Arabian Journal for Science and Engineering vol.49, p. 3623-3632, 2023. http://dx.doi.org/10.1007/s13369-023-08092-1

22. Elmaghraby R., Aziem N., Sobh M., and Bahaa-Eldin A., Encrypted Network Traffic Classification Based on Machine Learning, Ain Shams Engineering Journal, vol. 15, Issue 2, 2024. https://doi.org/10.1016/j.asej.2023.102361

23. Talukder A., Islam M., Udin A., Hassan K., Sharmn S., Alyami S., and Moni M., Machine Learning Based Network Intrusion Detection for Big and Imbalanced Data using Oversampling Stacking Feature Embedding and Feature Extraction, Journal of big data, Vol. 11, Article no.33, 2024. https://doi.org/10.1186/s40537-024-00886-w