

A Hybrid Metaheuristic Algorithm for Features Dimensionality Reduction in Network Intrusion Detection System

Bukola Fatimah Balogun[1], Kazeem Alagbe Gbolagade[1], Micheal Olaolu Arowolo[2], and Yakub Kayode Saheed[3(✉)]

[1] Department of Computer Science, Kwara State University, Malete, Nigeria
[2] Department of Computer Science, Landmark University, Omu-Aran, Nigeria
[3] School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria
yakubu.saheed@aun.edu.ng

Abstract. The advent of the Internet computer, and thus the amounts of connected computers in the last few decades, has opened vast quantities of intelligence to attackers and intruders. Firewalls are designed to identify, and block potentially harmful incoming traffic based on a predefined rule set. But, as attack tactics evolve, it becomes more difficult to differentiate anomalous traffic from regular traffic. Numerous detection strategies using machine-learning approaches have been suggested. However, there are issues with the high dimensional data of network traffic, the performance accuracy, and the high rate of false-positive and false-negative. In this paper, we propose a hybrid metaheuristic features dimensionality reduction method for Intrusion Detection Systems (IDSs). We used metaheuristic Bat algorithm for feature selection. The Bat algorithm selects sixteen (16) attributes. Subsequently, RNS was used to obtain the residues of the sixteen features selected. Then, the PCA was used to get the residues by extracting it. The experimental analysis was performed on NSLKDD dataset. The propose Bat-RNS + PCA + RF achieved 98.95% accuracy, sensitivity of 99.40% and F-score of 97.70%. The findings were also benchmarked with existing studies and our results were superior.

Keywords: Bat algorithm · Residue number system · Principal component analysis · Random forest · Intrusion detection system · Dimensionality reduction

1 Introduction

Network technology has been a critical component in advancing the industrial revolution 4.0, serving as a platform for the development of industrial technology automation and as a means of sharing knowledge [1]. However, the advancement in communication and information technology is also harmed by several threats to network security. Numerous threats posed by black attackers or ransomware can result in widespread attacks on information technology and computer system [2–5]. Due to the widespread use of the

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 101–114, 2021.
https://doi.org/10.1007/978-3-030-87013-3_8

internet computer and increased usage, and access to online content, cyber-crime is also growing in frequency [6, 7]. Intrusion detection is the first line of defense against a security breach [8–10]. As a result, studies focus heavily on security technologies like IDSs, intrusion prevention systems (IPS), firewalls, and unified threat modeling. Intrusion detection detects computer threats by the review of different network records [11, 12].

Intrusion detection models are divided into two types: those that detect misuse and those that detect anomalies. Misuse detection may identify intrusions based on a recognized pattern, referred to as a signature [13, 14]. By detecting deviations from the usual network traffic pattern, anomaly detection may recognize malicious activities [15–17]. As a result, anomaly detection will detect novel anomalies. The issue with current developmental strategies is their high-rate of false-positives and low-rate of false- negatives. Therefore, how to reduce the high rate of false-positives and low-rate of false-negatives have been an important issue in cybersecurity field.

Numerous machine learning based approaches such as genetic algorithm [18–20], SVM [21, 22], RF [22, 23], DT [24], PSO [25], KNN [26, 27], ANN [28, 29], and NB [30, 31] have been proposed to address these issues in the past. None of the techniques, however, can detect all intrusion attempts and resulting in a higher detection accuracy and a lower false alarm rate. As a result, it is essential to incorporate feature selection and classifier techniques in order to improve the efficiency of IDSs.

Feature selection phase is a critical step in solving classification tasks because it leads to the elimination of redundant and irrelevant features, and does not just decreases learning-time nonetheless, also advances classification performance [32, 33]. The methods for selecting features can be categorized as filter, wrapper, or hybrid. Filter methods pick a function subgroup as the pre- processing phase based on a predefined standard and without regard for the classifier's efficiency. Thus, filter techniques are typically less computationally intensive than costly wrapper methods that test feature selection methods using the classification results. Though wrapper methods typically outperform filter techniques in terms of classifier accuracy, the obtained results are often unusable when the classifier is modified. Thus, the methods that combined the filter method and wrapper method is known as hybrid methods. The method considered in this paper can be regarded as wrapper methods.

Efficient and cost-effective IDS are typically built using data mining algorithms as a result of the fact that they can spot intrusions and performing simplifications with ease [34]. Though, such systems are inherently complicated to incorporate and deploy. The inherent complexities of the systems can be divided into dissimilar collections of problem on the basis of accuracy, capability, and adaptability criteria [35, 36]. Yet, IDS developed because of data mining methods, particularly those that are focused on anomaly detection, have a superior rate of false-positive instances than earlier techniques for identification focused on hand-crafted signatures.

One of the most significant obstacles to assessing the efficiency of network intrusion detection systems is the absence of a robust network-protocol based data collection [37]. Majority of the anomaly IDS studies in the literatures were experimented on KDDCup'99 dataset [38]. We utilized the NSLKDD [39] dataset which is a standard dataset for IDS in this research. PCA is used in this work to produce a new collection of uncorrelated

characteristics, removing noise and avoiding the use of small variance attributes (that is, variables that are single-valued). Additionally, these newest attributes are chosen for their discriminative potential. Following that, the Random forest algorithm is being used to represent and categorize the feature vectors. Numerous models for intrusion detection have been produced using technique for selection and classification of features. The proposed metaheuristic Bat-RNS model differs from conventional IDS methods in that the Bat algorithm is optimized using residue number system.

This paper is structed as follows. Section 2 is the related work. We highlight the methodology employed in Sect. 3 and in Sect. 4, we present the results and discussion. The conclusion is presented in Sect. 5.

2 Related Work

Numerous investigations have been carried out on IDSs with ML technique. Their primary objective is to address the feature selection issue which is known to be costly in time and efforts, increase the output accuracy, sensitivity, F-score and recommend the standard dataset.

Recently, Ref. [40] proposed a feature selection approach for improving the IDS performance. The authors proposed Information Gain and Ranker technique as the feature selection method and they performed classification with KNN, NB and SVM. The evaluation was done on NSLKDD dataset. The results obtained gave more than 95% accuracy for the three classifiers with low false positive and false alarm rate.

Ref. [41] proposed a model utilizing the LDA and GA as the feature selection algorithms. The LDA is implemented to find the best features set while the GA is presented to obtain the novel features from the fitness score. The experimental analysis was done on KDD Cup'99 IDS dataset.

Ref. [42] presented a wrapper feature-selection approach for network IDS. They used ANN and SVM for the classification method using the NSLKDD dataset. The findings exposed that the ANN with wrapper feature selection outperformed the SVM with wrapper feature selection. The results obtained was efficient as compared to reported works.

Ref. [43] proposed a method on IDS by using a ranking based method with chi-square for feature selection. The classification was done with SVM. The work was evaluated on both KDD Cup99 and the NSL-KDD datasets. The method gave better results on NSL-KDD in comparison to KDDCup'99. Authors in [21] presented feature selection weight and Genetic algorithm for optimizing parameter of SVM. The results obtained showed a reduced rate of false positive and false negative.

Ref. [44] presented a method that used PCA to select subsets of the features based on the PCA eigenvalues. They presented a genetic algorithm based principal components in order to select attributes with higher eigen values to pick the SVM subset attributes for the classification.

Ref. [45] presented three methods for feature selection issues in IDS. They used information reinforcement, gain ratios and correlation based. The authors also suggest a new feature selection method using attribute vitality reduction approach to find less

significant features. The findings of their results showed that with features selection they obtained superior performance.

Ref. [46] proposed two feature selection approaches which are correlation coefficient measure and information gain method. The proposed method gave high performance accuracy in detecting attacks on the IDS.

Ref. [47] utilized correlation feature selection and particle swarm optimization as feature selection method to eliminate redundant features. This was done to increase the performance of the classifier accuracy. The experiment was done on KDDCup '99 dataset and findings obtained reached an accuracy up to 99%.

The authors [48] utilize K-means method of clustering for feature engineering. In this method, the evaluation value of the correlation is first determined in which a cluster is molded founded on the clustering algorithm K-means.

3 Materials and Method

We presented the proposed methodology in this section. In the first line of our methodology, we used metaheuristic Bat algorithm for feature selection that belong to the family of wrapper techniques. The Bat algorithm selects sixteen (16) attributes. Subsequently, the fusion of RNS was performed on the sixteen features selected from the Bat algorithm to obtain the residues of the sixteen features selected. Then, the PCA was used to get the residues by extracting it. PCA is used in this work to produce a new collection of uncorrelated characteristics, removing noise and avoiding the use of small variance (that is, attributes that are single-valued). Additionally, these newest attributes are chosen for their discriminative potential. Following that, the Random forest algorithm was implemented to model and categorize the feature vector.

3.1 Bat Algorithm

The Bat Algorithm was created in response to bats' echolocation activity. Bats typically produce a brief tone pulse and detect a fractional echo in a second, that is employed to assess their distance after an entity [49]. Additionally, they possess the ability to discriminate between obstacles and target, which enables them to search in completely dim environments [50]. Each bat represents a solution in the BAT algorithm, which is algorithm for population-based evolution. It was created using the echolocation activity of virtual bats as a guide. This ability allows them to detect their prey's location. For echolocation, frequency modulated pulses are used [51]. Because as bat gets closer to its prey, the volume of their reverberation decreases and the rate of the sound pulse increases.

3.2 Principal Component Analysis

PCA is a commonly used algorithm for mining the most important information from a dataset in a variety of applications [52, 53]. Indeed, it has been successfully applied to applications involving facial recognition. PCA is used in this case to generate an entirely new collection of unrelated attributes drawn from a collection of related attributes [54, 55]. Therefore, PCA produces a collection of orthogonal functions that are used to shown data as a weighted sum of those basis vectors [56].

3.3 Residue Number System

For many decades, the residue number system (RNS) has been a significant area of investigation in computer arithmetic, owing to its carry-free design, which enables the development of architectures for high-performance computing with superior delay requirements [57]. RNS have several benefits over binary number systems in general. This is due to their intrinsic characteristics, which include the absence of carry, parallelism, modularity, and fault tolerance [58].

3.4 Random Forest Algorithm

Random Forests are algorithms for supervised learning [59]. RF is a type of ML technique that combines many decision trees to construct an algorithm for generating an accurate and robust prediction equation for results. These multiple trees are created arbitrarily and trained for a particular action that becomes the model's final outcome [60]. RF is usually utilized in distributed denial of service attack prediction [61] and anomaly prediction.

3.5 Description of NSLKDD Dataset

The experimental work was carried out using the NSLKDD dataset, which is a typical dataset for intrusion detection. NSLKDD was created by Tavallaee et al. to address the inherent issues with the KDDCupp '99 data collection, as mentioned in [62]. However, some of the issues addressed by McHugh in [63] persist in this latest version of the dataset, and it may not be a complete illustration of actual real networks. This dataset is still used in most modern NIDS research, so we assume it is still a useful benchmark for researchers to compare various methods. The NSL-KDD database has a similar structure to the KDD Cup '99 dataset (that is it has normal traffic or twenty-two (22) attack patterns, and 41 attributes). For our evaluations, we will use the entire NSL-KDD dataset, the structure of which has been also shown in Table 1.

4 Results and Discussions

We used metaheuristic Bat algorithm with the fusion of RNS for feature selection in which sixteen (16) attributes were selected out of forty (40) attributes with one class label. PCA is then used for feature extraction. Thus, RF model was used for the classification in terms accuracy, false alarm, sensitivity, F-score, precision, and specificity. The feature selection time and the classifier training time was also noted in our experimental analysis. About 25% of the NSLKDD dataset was used for testing the model and 75% was used for training.

4.1 Performance Evaluation Metrics

i. True Positive rate (TP)/Sensitivity: Data about an attack that has been correctly identified as such.
ii. False Positive rate (FP): Data that is normal that has been mislabeled as an attack data.

Table 1. Attributes of the NSLKDD dataset

Attribute no	Attribute name	Attribute no	Attribute name	Attribute no	Attribute name
A1	Duration	A15	Su attempted	A29	Same server rate
A2	Type of protocol	A16	Number root	A30	Diff server rate
A3	Service	A17	Number file creation	A31	Srv diff host rate
A4	Flag	A18	Number shells	A32	Dst host count
A5	Source bits	A19	Number access files	A33	Dst host server count
A6	Destination bits	A20	Number outbound commands	A34	Dst host same server rate
A7	Land	A21	Is host log in	A35	Dst host diff server rate
A8	Error fragment	A22	Is guest log in	A36	Dst host same source port rate
A9	Critical	A23	Count	A37	Dst host server diff host rate
A10	Hot	A24	Server count	A38	Dst host serror rate
A11	Number Unsuccessful logins	A25	Server error rate	A39	Dst host server error rate
A12	Logged in	A26	Srverror rate	A40	Dst host rerror rate
A13	Num compromise	A27	R error rate	A41	Dst host srvr error rate
A14	Source shell	A28	Srverror rate	A42	Class Label

iii. True Negative rate (TN)/Specificity: Data that is normal that has been correctly categorized.
iv. False Negative rate (FN): Data that is attack that has been labeled incorrectly as normal data.

The following performance measures were adopted in this paper to evaluate the performance of the model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

The proportion of the total number of valid classifications is measured by the accuracy. The most critical measures in an intrusion detection system is accuracy [40, 64]. The accuracy, sensitivity, specificity, precision and F-score of our proposed model were used to evaluate its efficiency.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

The sensitivity is a calculation that divides the number of valid classifications by the number of missing entries.

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{3}$$

Specificity is proportional to the number of records admitted correctly, and it refers to the number of normal records known as normal.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

Precision is computed by separating the number of correctly classified instances by the amounts of incorrect categories.

$$\text{False Alarm} = \frac{FP}{FP + TN} \tag{5}$$

The false alarm rate calculates the percentage of benign incidents reported incorrectly as malicious.

$$\text{F - Score } = 2 \cdot \frac{Precision.Sensitivity}{Precision + Recall} \tag{6}$$

The F-score is a derived effectiveness metric that computes the harmonic mean of recall and precision.

4.2 Experimental Results of Bat-RNS + PCA + RF

The time taken by the metaheuristic Bat algorithm with the fusion of RNS for feature selection is depicted in Table 2 with the RF training time. The time taken to selects the sixteen features was 56.43 s with the RF training time of 22.05 s.

The performance measures of the proposed method Bat-RNS+PCA+RF is given in Table 3. The proposed method gave an accuracy of 98.95%, 99.4% sensitivity, 95.26% specificity, 96.0% precision and F-score of 97.7%

The feature selection time versus the training time is revealed in Fig. 1. The time taken by the proposed model to select the significant attributes is about 56.43 s. Whereas, the

Table 2. Feature selection time and training time

Feature selection time (s)	Training time (s)
56.43	22.05

Table 3. Performance measures of metaheuristic Bat-RNS + PCA + RF

Model/Metrics	Accuracy	Sensitivity	Specificity	Precision	F-score
Bat-RNS + PCA + RF	98.95	99.40	95.26	96.00	97.70



Fig. 1. Time taken for feature selection and training the proposed model

training time is 22.05 s. This is as a result of the facts that the metaheuristic Bat algorithm has eliminated the redundant features which made it possible for the RF classification algorithm to have few times for training the model.

The performance measures of the proposed model are depicted in Fig. 2. The proposed model gave higher sensitivity as compared to the accuracy, specificity, precision and F-score. Whereas, the accuracy gave higher value than the specificity, precision and f-score. The specificity gave the lowest performance as compared to other measures.

4.3 Comparison with the Recent Studies in IDS

The performance measures of the proposed model were bench marked with the existing model as presented in Table 4. As evident in Table 4, the proposed Bat-RNS+PCA+RF gave outstanding accuracy, sensitivity and F-score. However, the existing studies also gave better results in terms of accuracy. Though, we cannot

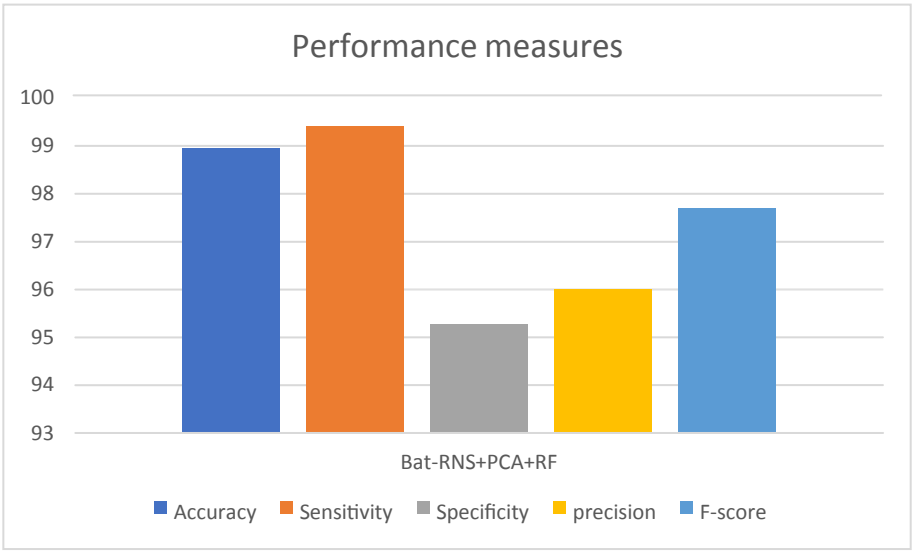

Fig. 2. Performance measures of the proposed Bat-RNS+PCA+RF

claim that the proposed Bat-RNS+PCA+RF outperforms the existing studies in all IDS performance measures, since all other metrics were unknown for the previous studies. In terms of validation dataset, the NSLKDD dataset was established to be the standard dataset. Our proposed method was evaluated on NSLKDD. However, some previous studies as shown in Table 4 still relied on KDDCup'99 which has some inherent issues as noted in [62].

Table 4. Performance of Bat-RNS + PCA + RF with the recent works in IDS

Authors	Accuracy	Sensitivity	F-score	Datasets
Saheed et al. [40]	94.00	x	X	NSLKDD
Kasliwal et al. [41]	83.25	x	85.2	KDDCup'99
Taher et al. [42]	94.02	x	X	NSLKDD
Thaseen et al. [43]	98.00	X	X	NSLKDD
Mukherjee et al. [45]	98.80	X	X	NSLKDD
Tsang et al. [65]	98.70	X	X	KDDCup'99
Raman et al. [66]	97.14	96.27	X	KDDCup'99
Proposed method	**98.95**	**99.40**	**97.70**	**NSLKDD**

5 Conclusion and Future Work

The traffic in network dataset is very huge with large number of distributions of connections of protocol. IDS with constraint of time find it difficult to process the whole network dataset. Hence, feature dimensionality is important for IDSs. Also, the feature selection phase is known to be critical both in efforts and time. Therefore, finding an effective and efficient feature dimensionality reduction is an issue in IDS. We proposed a metaheuristic Bat-RNS for feature dimensionality in this paper to selects the significant attributes in the network dataset. We adopt PCA for the feature extraction and performed our classification with RF algorithm. The experimental analysis was performed on the NSLKDD dataset. The results findings revealed an accuracy of 98.95%, sensitivity of 99.40% and F-score of 97.70%. Our results were also bench mark with existing studies and our findings showed better performance. The future work would be to introduce a deep learning method for IDS with the fusion of RNS.

References

1. Aziz, M.N., Ahmad, T.: Cluster analysis-based approach features selection on machine learning for detecting intrusion. Int. J. Intell. Eng. Syst. **12**(4), 233–243 (2019). https://doi.org/10.22266/ijies2019.0831.22
2. Gunduz, M.Z., Das, R.: Cyber-security on smart grid: threats and potential solutions. Comput. Netw. **169**, 107094 (2020). https://doi.org/10.1016/j.comnet.2019.107094
3. Alenezi, M.N., Alabdulrazzaq, H., Alshaher, A.A., Alkharang, M.M.: Evolution of malware threats and techniques: a review. Int. J. Commun. Networks Inf. Secur. **12**(3), 326–337 (2020)
4. Yaacoub, J.P.A., Salman, O., Noura, H.N., Kaaniche, N., Chehab, A., Malli, M.: Cyber-physical systems security: limitations, issues and future trends. Microprocess. Microsyst. **77**, 103201 (2020). https://doi.org/10.1016/j.micpro.2020.103201
5. Ogonji, M.M., Okeyo, G., Wafula, J.M.: A survey on privacy and security of Internet of Things. Comput. Sci. Rev. **38**, 100312 (2020). https://doi.org/10.1016/j.cosrev.2020.100312
6. Alaei, P., Noorbehbahani, F.: Incremental anomaly-based intrusion detection system using limited labeled data. In: 2017 3rd International Conference on Web Research ICWR 2017, pp. 178–184 (2017). https://doi.org/10.1109/ICWR.2017.7959324
7. Song, H., Lynch, M.J., Cochran, J.K.: a macro-social exploratory analysis of the rate of interstate cyber-victimization. Am. J. Crim. Justice **41**(3), 583–601 (2015). https://doi.org/10.1007/s12103-015-9308-4
8. Khan, K., Mehmood, A., Khan, S., Khan, M.A., Iqbal, Z., Mashwani, W.K.: A survey on intrusion detection and prevention in wireless ad-hoc networks. J. Syst. Arch. **105**, 101701 (2020). https://doi.org/10.1016/j.sysarc.2019.101701
9. Rubio, J.E., Alcaraz, C., Roman, R., Lopez, J.: Current cyber-defense trends in industrial control systems. Comput. Secur., 101561 (2019). https://doi.org/10.1016/j.cose.2019.06.015
10. Patil, S.S., Sonavane, S.P.: Data Science and Big Data: An Environment of Computational Intelligence, vol. 24, pp. 49–81 (2017). https://doi.org/10.1007/978-3-319-53474-9
11. Naganhalli, N.S., Terdal, S.: Network intrusion detection using supervised machine learning technique. Int. J. Sci. Technol. Res. **8**(9), 345–350 (2019)
12. Bhattacharya, S., et al.: A novel PCA-firefly based XGBoost classification model for intrusion detection in networks using GPU. Electron. **9**(2), 219 (2020). https://doi.org/10.3390/electronics9020219

13. Kaur, S., Singh, M.: Hybrid intrusion detection and signature generation using Deep Recurrent Neural Networks. Neural Comput. Appl. **32**(12), 7859–7877 (2019). https://doi.org/10.1007/s00521-019-04187-9
14. Gupta, A.R., Agrawal, J.: The multi-demeanor fusion based robust intrusion detection system for anomaly and misuse detection in computer networks. J. Ambient. Intell. Humaniz. Comput. **12**(1), 303–319 (2020). https://doi.org/10.1007/s12652-020-01974-4
15. Shijoe Jose, D., Malathi, B.R., Jayaseeli, D.: A survey on anomaly based host intrusion detection system. J. Phys.: Conf. Ser. **1000**, 012049 (2018). https://doi.org/10.1088/1742-6596/1000/1/012049
16. Anand, K., Kumar, J., Anand, K.: Anomaly detection in online social network: a survey. In: Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017, pp. 456–459 (2017). https://doi.org/10.1109/ICICCT.2017.7975239
17. Zhou, L., Guo, H.: Anomaly detection methods for IIoT networks. In: Proceedings of the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2018, pp. 214–219 (2018). https://doi.org/10.1109/SOLI.2018.8476769
18. Gauthama Raman, M.R., Somu, N., Kirthivasan, K., Ramiro Liscano, V.S., Sriram, S.: An efficient intrusion detection system based on hypergraph - Genetic algorithm for parameter optimization and feature selection in support vector machine. Knowl.-Based Syst. **134**, 1–12 (2017). https://doi.org/10.1016/j.knosys.2017.07.005
19. Saheed, Y., Babatunde, A.: Genetic Algorithm Technique in Program Path Coverage For Improving Software Testing, vol. 7, no. 5, pp. 151–158 (2014)
20. Resende, P.A.A., Drummond, A.C.: Adaptive anomaly-based intrusion detection system using genetic algorithm and profiling. Secur. Priv. **1**(4), e36 (2018). https://doi.org/10.1002/spy2.36
21. Tao, P., Sun, Z., Sun, Z.: An improved intrusion detection algorithm based on GA and SVM. IEEE Access **6**, 13624–13631 (2018). https://doi.org/10.1109/ACCESS.2018.2810198
22. Ahmad, I., Basheri, M., Iqbal, M.J., Rahim, A.: Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE Access **6**, 33789–33795 (2018). https://doi.org/10.1109/ACCESS.2018.2841987
23. Aung, Y.Y., Min, M.M.: An analysis of random forest algorithm based network intrusion detection system. In: Proceedings - 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. SNPD 2017, pp. 127–132 (2017). https://doi.org/10.1109/SNPD.2017.8022711
24. Ahmim, A., Maglaras, L., Ferrag, M.A., Derdour, M., Janicke, H.: A novel hierarchical intrusion detection system based on decision tree and rules-based models. In: Proceedings - 15th Annual International Conference on Distributed Computing in Sensor Systems. DCOSS 2019, pp. 228–233 (2019). https://doi.org/10.1109/DCOSS.2019.00059
25. Syarif, A.R., Gata, W.: Intrusion detection system using hybrid binary PSO and K-nearest neighborhood algorithm," Proceedings of 11th International Conference on Information and Communication Technology and Systems. ICTS 2017, vol. 2018-January, pp. 181–18 (2018). https://doi.org/10.1109/ICTS.2017.8265667
26. Frp, V.J., et al.: *Hqhwlf 3Urjudpplqj Dqg . 1Hduhvw 1Hljkerxu &Odvvlilhu %Dvhg,Qwuxvlrq 'Hwhfwlrq 0Rgho, pp. 42–46 (2017)
27. Reazul, M., Rahman, A., Samad, T.: A network intrusion detection framework based on bayesian network using wrapper approach. Int. J. Comput. Appl. **166**(4), 13–17 (2017). https://doi.org/10.5120/ijca2017913992
28. Dias, L.P., Cerqueira, J.J.F., Assis, K.D.R., Almeida, R.C.: Using artificial neural network in intrusion detection systems to computer networks. In: 2017 9th Computer Science and Electronic Engineering Conference CEEC 2017 - Proceeding, pp. 145–150 (2017). https://doi.org/10.1109/CEEC.2017.8101615

29. Sumaiya Thaseen, I., Saira Banu, J., Lavanya, K., Rukunuddin Ghalib, M., Abhishek, K.: An integrated intrusion detection system using correlation-based attribute selection and artificial neural network. Trans. Emerg. Telecommun. Technol. **32**(2), 1–15 (2021). https://doi.org/10.1002/ett.4014
30. Gu, J., Lu, S.: An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Comput. Secur. **103**, 10215 (2021). https://doi.org/10.1016/j.cose.2020.102158
31. Talita, A.S., Nataza, O.S., Rustam, Z.: Naïve bayes classifier and particle swarm optimization feature selection method for classifying intrusion detection system dataset. J. Phys.: Conf. Ser. **1752**, 012021 (2021). https://doi.org/10.1088/1742-6596/1752/1/012021
32. Zargari, S., Voorhis, D.: Feature selection in the corrected KDD-dataset. In: Proceedings - 3rd International Conference on Emerging Intelligent Data and Web. EIDWT 2012, pp. 174–180 (2012). https://doi.org/10.1109/EIDWT.2012.10
33. Saheed, Y.O.Y.K., Hambali, M.A., Arowolo, M.O.: Application of GA feature selection on Naive Bayes, Random Forest and SVM for Credit Card Fraud Detection. In: 2020 International Conference on Decision Aid Sciences and Application (DASA), pp. 1091–1097 (2020)
34. Aljawarneh, S., Aldwairi, M., Yassein, M.B.: Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. J. Comput. Sci. **25**, 152–160 (2018). https://doi.org/10.1016/j.jocs.2017.03.006
35. De la Hoz, E., De La Hoz, E., Ortiz, A., Ortega, J., Prieto, B.: PCA filtering and probabilistic SOM for network intrusion detection. Neurocomputing **164**, 71–81 (2015). https://doi.org/10.1016/j.neucom.2014.09.083
36. Ravale, U., Marathe, N., Padiya, P.: Feature selection based hybrid anomaly intrusion detection system using K Means and RBF kernel function. Procedia Comput. Sci. **45**, 428–435 (2015). https://doi.org/10.1016/j.procs.2015.03.174
37. Moustafa, N., Slay, J.: UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW- NB15 network data set). In: 2015 Military Communications and Information Systems Conference MilCIS 2015 - Proceedings 2015. https://doi.org/10.1109/MilCIS.2015.7348942
38. Zargari, S.: Feature Selection in UNSW-NB15 and KDDCUP'99 datasets
39. Dhanabal, L., Shantharajah, S.P.: A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. Int. J. Adv. Res. Comput. Commun. Eng. **4**(6), 446–452 (2015). 17148/IJARCCE.2015.4696
40. Saheed, Y.K., Hamza-usman, F.E.: Feature Selection with IG-R for Improving Performance of Intrusion Detection System, vol. 12, no. 3, pp. 338–344 (2020)
41. Kasliwal, B., Bhatia, S., Saini, S., Thaseen, I.S., Kumar, C.A.: A hybrid anomaly detection model using G-LDA. In: Souvenir 2014 IEEE International Advance Computing Conference. IACC 2014, pp. 288–293 (2014). https://doi.org/10.1109/IAdCC.2014.6779336
42. Taher, K.A., Mohammed Yasin Jisan, B., Rahman, M.M.: Network intrusion detection using supervised machine learning technique with feature selection. In: 1st International Conference on Robotics, Electrical and Signal Processing Techniques. ICREST, pp. 643– 646 (2019). https://doi.org/10.1109/ICREST.2019.8644161
43. Sumaiya Thaseen, I., Aswani Kumar, C.: Intrusion detection model using fusion of chi-square feature selection and multi class SVM. J. King Saud Univ. - Comput. Inf. Sci. **29**(4), 462–472 (2017). https://doi.org/10.1016/j.jksuci.2015.12.004
44. Guo, C., Zhou, Y., Ping, Y., Zhang, Z., Liu, G., Yang, Y.: A distance sum-based hybrid method for intrusion detection. Appl. Intell. **40**(1), 178–188 (2013). https://doi.org/10.1007/s10489-013-0452-6
45. Mukherjee, S., Sharma, N.: Intrusion detection using naive bayes classifier with feature reduction. Procedia Technol. **4**, 119–128 (2012). https://doi.org/10.1016/j.protcy.2012.05.017

46. Amiri, F., Rezaei Yousefi, M., Lucas, C., Shakery, A., Yazdani, N.: Mutual information-based feature selection for intrusion detection systems. J. Netw. Comput. Appl. **34**(4), 1184–1199 (2011). https://doi.org/10.1016/j.jnca.2011.01.002
47. Ahmad, T., Aziz, M.N.: Data preprocessing and feature selection for machine learning intrusion detection systems. ICIC Express Lett. **13**(2), 93–101 (2019). https://doi.org/10.24507/icicel.13.02.93
48. Fouedjio, F.: A hierarchical clustering method for multivariate geostatistical data. Spat. Stat. **18**, 333–351 (2016). https://doi.org/10.1016/j.spasta.2016.07.003
49. Natesan, P., Rajalaxmi, R.R., Gowrison, G., Balasubramanie, P.: Hadoop based parallel binary bat algorithm for network intrusion detection. Int. J. Parallel Program. **45**(5), 1194–1213 (2017). https://doi.org/10.1007/s10766-016-0456-z
50. Yang, X.S.: A new metaheuristic Bat-inspired Algorithm. Stud. Comput. Intell. **284**, 65–74 (2010). https://doi.org/10.1007/978-3-642-12538-6_6
51. Sreeram, I., Vuppala, V.P.K.: HTTP flood attack detection in application layer using machine learning metrics and bio inspired bat algorithm. Appl. Comput. Inf. **15**(1), 59–66 (2019). https://doi.org/10.1016/j.aci.2017.10.003
52. Uddin, M.P., Al Mamun, M., Hossain, M.A.: Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification. Int. J. Remote Sens. **40**(18), 7190–7220 (2019). https://doi.org/10.1080/01431161.2019.1601284
53. Bouwmans, T., Javed, S., Zhang, H., Lin, Z., Otazo, R.: On the applications of robust PCA in image and video processing. Proc. IEEE **106**(8), 1427–1457 (2018). https://doi.org/10.1109/JPROC.2018.2853589
54. Nobre, J., Neves, R.F.: Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets. Expert Syst. Appl. **125**, 181–194 (2019). https://doi.org/10.1016/j.eswa.2019.01.083
55. Rajab, K.D.: New hybrid features selection method: a case study on websites phishing. Secur. Commun. Netw. **2017**, 1–10 (2017). https://doi.org/10.1155/2017/9838169
56. Bouhlel, J., et al.: Comparison of common components analysis with principal components analysis and independent components analysis: application to SPME-GC-MS volatolomic signatures. Talanta **178**, 854–863 (2018). https://doi.org/10.1016/j.talanta.2017.10.025
57. Navi, K., Molahosseini, A.S., Esmaeildoust, M.: How to teach residue number system to computer scientists and engineers,. IEEE Trans. Educ. **54**(1), 156–163 (2011). https://doi.org/10.1109/TE.2010.2048329
58. Gbolagade, K.A., Chaves, R., Sousa, L., Cotofana, S.D.: An improved RNS reverse converter for the {22n+1–1,2 n,2n-1} moduli set. ISCAS 2010 - 2010 International Symposium on Circuits and Systems, Nano-Bio Circuit Fabrics and Systems, pp. 2103–2106 (2010). https://doi.org/10.1109/ISCAS.2010.5537062
59. Al-Garadi, M.A., Mohamed, A., Al-Ali, A.K., Du, X., Ali, I., Guizani, M.: A survey of machine and deep learning methods for Internet of Things (IoT) security. IEEE Commun. Surv. Tutorials **22**(3), 1646–1685 (2020). https://doi.org/10.1109/COMST.2020.2988293
60. Tahsien, S.M., Karimipour, H., Spachos, P.: Machine learning based solutions for security of Internet of Things (IoT): a survey. J. Netw. Comput. Appl. **161**(February), 102630 (2020). https://doi.org/10.1016/j.jnca.2020.102630
61. Doshi, R., Apthorpe, N., Feamster, N.: Machine learning DDoS detection for consumer internet of things devices. In: IEEE Symposium on Security and Privacy Work. SPW 2018, no. Ml, pp. 29–35 (2018). https://doi.org/10.1109/SPW.2018.00013
62. Tavallaee, M., Bagheri, E., Lu, W., Ghorbani, A.A.: A detailed analysis of the KDD CUP 99 data set in Computational Intelligence for Security and Defense Applications. Comput. Intell. Secur. Def. Appl. no. Cisda, pp. 1– 6 (2009)

63. Mchugh, J.: Testing intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by lincoln laboratory. ACM Trans. Inf. Syst. Secur. **3**(4), 262–294 (2000). https://doi.org/10.1145/382912.382923
64. Yin, C., Zhu, Y., Fei, J., He, X.: A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access **5**, 21954–21961 (2017). https://doi.org/10.1109/ACCESS.2017.2762418
65. Tsang, C.H., Kwong, S., Wang, H.: Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. Pattern Recognit. **40**(9), 2373–2391 (2007). https://doi.org/10.1016/j.patcog.2006.12.009
66. Raman, M.R.G., Somu, N., Kirthivasan, K., Sriram, V.S.S.: A hypergraph and arithmetic residue-based probabilistic neural network for classification in intrusion detection systems. Neural Netw. **92**, 89–97 (2017). https://doi.org/10.1016/j.neunet.2017.01.012