

Efficient Machine Learning Model for DDoS Detection System Based on Dimensionality Reduction

Saad Ahmed Dheyab ¹ , Shaymaa Mohammed Abdulameer ² , Salama Mostafa ³ 

¹ College of Engineering, University of Information Technology and Communications, Baghdad, Iraq

² College of Information Engineering, Al-Nahrain University, Baghdad, Iraq

³ Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia

Corresponding author: Salama Mostafa (salama@uthm.edu.my)

Abstract

Distributed denial of service (DDoS) attacks are one of the most common global challenges faced by service providers on the web. It leads to network disturbances, interruption of communication and significant damage to services. Researchers seek to develop intelligent algorithms to detect and prevent DDoS attacks. The present study proposes an efficient DDoS attack detection model. This model relies mainly on dimensionality reduction and machine learning algorithms. The principal component analysis (PCA) and the linear discriminant analysis (LDA) techniques perform the dimensionality reduction in individual and hybrid modes to process and improve the data. Subsequently, DDoS attack detection is performed based on random forest (RF) and decision tree (DT) algorithms. The model is implemented and tested on the CICDDoS2019 dataset using different data dimensionality reduction test scenarios. The results show that using dimensionality reduction techniques along with the ML algorithms with a dataset containing high-dimensional data significantly improves the classification results. The best accuracy result of 99.97% is obtained when the model operates in a hybrid mode based on a combination of PCA, LDA and RF algorithms, and the data reduction parameter equals 40.

Keywords

Distributed Denial of Service (DDoS); Intrusion Detection Systems (IDS); Machine Learning (ML); Random Forest (RF); Decision Tree (DT); Dimensionality Reduction (DR).

Citation: Dheyab, S. A., Abdulameer, S. M., & Mostafa, S. (2022). Efficient Machine Learning Model for DDoS Detection System Based on Dimensionality Reduction. *Acta Informatica Pragensia*, 11(3), Forthcoming articles. <https://doi.org/10.18267/j.aip.199>

Academic Editor: Stanislav Vojir, Prague University of Economics and Business, Czech Republic

Copyright: © 2022 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

Distributed denial of service (DDoS) attacks are the most critical and common attack types. They target conventional communication networks as well as new-generation ones, such as fifth-generation (5G), Internet of Things (IoT), cloud computing and communication networks (Yungaicela-Naula et al., 2021; Hezam et al., 2021). A DDoS attack can cut off network service by overwhelming the target network server by flooding it with redundant requests in an attempt to overload the server and prevent fulfilling legitimate demands. It utilizes several compromised systems as internet traffic sources to increase their effectiveness (Khalaf et al., 2021). What increases the lethality of the DDoS attacks is that detecting these attacks requires differentiation of legitimate requests from illegitimate ones. A service or site that is unexpectedly sluggish or inaccessible can be considered the most evident DDoS attack sign (Rahal et al., 2020).

DDoS is hard to detect because hackers make attack traffic similar to regular traffic in most cases (Rahman et al., 2019). Nonetheless, there are some common indicators of the presence of a DDoS attack. Firstly, malicious packets utilized for DDoS have identical port and destination addresses. In addition, those packets have some typical size that differs from the legitimate traffic size (Azizan et al., 2021). Organizations have been deploying various preventive measures such as access control lists, firewalls, antivirus programs and intrusion detection systems (IDS) or intrusion prevention systems (IPS) to prevent unauthorized access (Sultana et al., 2019).

Hackers are people exploiting and accessing others' data illegally. Hackers may utilize various strategies to perform passive or active attacks. Passive attacks are not involved in modifying resources; hackers merely sit back and analyse all the data (Priya et al., 2020). On the other hand, in active attack types, hackers are involved in modifying data besides blocking users from performing any actions. A DDoS is a widespread and critical active attack. It uses several distributed resources against the targets, depriving authorized clients of required services (Zhang et al., 2017). The attack targets include system resources, network bandwidth and other resource types.

Machine learning (ML) is one of the sub-sets of artificial intelligence (AI) that contributes to building intrusion detection systems (IDS) for DDoS attacks (Azizan et al., 2021). It can make a system learn and enhance its automatic capability from experience without being programmed explicitly. ML algorithms show high accuracy in detecting DDoS attacks by learning the pattern of legitimate and threat requests (Maseer et al., 2021). Dimensionality reduction (DR) is a technique used to reduce data size, which leads to improved data quality and reduces the execution time of ML algorithms. It increases data diversity and reduces similarity, which leads to enhanced ML classification results (Zong and Huang, 2021).

The main objective of this study is to improve DDoS detection by relying on the outputs of two DR algorithms of principal component analysis (PCA) and linear discriminant analysis (LDA). Subsequently, this paper proposes an IDS model for DDoS detection based on random forest (RF) and decision tree (DT) integrated with the PCA and LDA. The overall model is implemented and tested using four test scenarios to evaluate its performance.

This paper is organized into five sections. The next section is a literature review. Section 3 presents the research methods and materials, which include a description of DDoS, the dataset, ML algorithms and dimensionality reduction algorithms. Section 4 presents the proposed DDoS detection model and reviews the attained results. The conclusion is presented in Section 5 at the end of the paper.

2 Literature Review

Several studies have produced different IDS models to expose DDoS attacks using a data mining approach. The most relevant studies include the work of Bhaya and Ebadymanaa (2017), who apply a combination of unsupervised data mining (DM) techniques as IDS on a collection of datasets (such as

CAIDA-2008, CAIDA-2007 and DARPA-2000). The IDS implemented entropy by windowing on incoming packets, utilizing the clustering using representatives (CURE) algorithm to expose the DDoS attack. The results of the suggested system showed a 96.29% detection accuracy and 0% false acceptance ratio (FAR).

Abdulrahman and Ibraheem (2018) studied detection of DDoS attacks based on the CICIDS2017 dataset, which includes benign and DDoS attack network flows. For the classification of seizures, decision tree (DT) type C5.0, random forest (RF), naïve Bayes (NB) and support vector machine (SVM) were utilized. According to the confusion matrix, the obtained maximum accuracy is 75% for the SVM, 86.8% for RF, 86.45% for C5.0, and the accuracy for the last two is 99%. Nevertheless, the FAR for the RF is 0.050%, and for C5.0 it is 0.046%. In 2019, Sharma et al. (2019) applied ML algorithms such as SVM, NB and RF to a Snort dataset. Classification was performed using ML approaches by implementation of WEKA software. A confusion matrix was utilized for evaluations. The overall accuracy scores were 99.70% for the SVM, 97.60% for the RF and 98% for the NB.

The work of Tuan et al. (2019) utilized several of the classifiers for the detection of botnet DDoS attacks, using ML algorithms, ANN, SVM, DT, NB and USML (X-means and K-means) on the KDD99 and UNBS-NB 15 datasets. The unsupervised learning algorithm optimized the expected traffic and the DDoS attacks. The USML obtained the maximum accuracy of 94.78% in the UNBS-NB15 dataset and 98.08% in the KDD-99 dataset. Zhijun et al. (2020) used a multi-feature DDoS attack detection method based on the factorization machine (FM). Features obtained from flow rules were utilized to detect low-rate DDoS attacks, and seeing a low-rate DDoS attack based on the FM approaches of ML was carried out. Experimental results showed that this approach can effectively detect low-rate DDoS attacks that target the SDN data layer. The accuracy of the detection reached 95.8%. It suggests a defence approach based on the dynamic deletion of flow rules and performs experiment simulations and analyses to prove the defence method efficiency. The success rate of forwarding regular packets reached 97.85%.

Table 1. Comparison of previous works.

Authors	Datasets	Detection methods	Detection accuracy (%)
Bhaya and Ebadymanaa (2017)	CAIDA-2008, CAIDA-2007, and DARPA-2000	CURE	96.2
Abdulrahman and Ibraheem (2018)	CICIDS2017	DT	86.4
		RF	86.8
		SVM	75.0
Sharma et al. (2019)	Private Snort haven	NB	98.0
		RF	97.6
		SVM	99.7
Tuan et al. (2019)	KDD99 and UNBS-NB15	NB	96.74, 71.6
		DT	93.3, 94.43
		USML	98.08, 94.78
		ANN	97.44, 63.97
		SVM	91.55, 84.32
Zhijun et al. (2020)	NSL-KDD, DARPA98, CAIDA	FM	92.5, 93.2, 95.8
Abbas and Almhanna (2021)	MIX (PORTMAP+LDAP)	RF+NB	99.9

In 2021, Abbas and Almhanna worked on detecting DDoS attacks using the RF for extracting data patterns with classification of types of given features in the training step. NB was used to classify data for comparing its classification results with the results obtained from the RF classifier. The MIX dataset was

used for the training as well as testing of the suggested model that resulted from the merging of 2 datasets (PORTMAP+LDAP), and both were utilized in the CICDDoS2019 dataset. The model accuracy score is 99.9764%, the detection rate score is 100%, the FAR score is nearly 0, and the F-measure score is 99.90% when PCA equals 25. Table 1 presents a summary of the related work, including the datasets, detection algorithms and accuracy results.

3 Research Methods

The transmission control protocol (TCP) flood attack is the most common attack of DDoS. A TCP flood attack sends a massive amount of TCP connection requests to the victim with no acknowledgment of the victim's server SYN-ACK. The victim's server is left with several half-opened connections. Those half connections take up most of its sources, making it unavailable for authentic user (Novaes et al., 2020). A different DDoS type is the internet control message protocol (ICMP) flood attack, which is also referred to as the smurf attack. ICMP flood attacks send ICMP packets with a spoof source internet protocol (IP) address. The owner of a spoof IP address will be a possible target because it will become the destination of several ICMP responses and be flooded. The user datagram protocol (UDP) flooding attack randomly sends too many UDP packets to various target ports (Khalaf et al., 2019). The DNS amplification attack is where attackers spoof the victim's source address. The attacker sends a small request to the DNS server, which replies with a large response (Ferrag et al., 2021). As shown in Figure 1, in this class, TCP-based attacks include the MSSQL SSDP. In contrast, UDP-based attack types include network time protocol (NTP), trivial file transfer protocol (TFTP) and character generator protocol (CharGen), as shown in Figure 1.

Network-based attacks represent threats that originate from and are regulated by devices other than those under attack. A DDoS attack is a network-based attack example in which systems of intrusion prevention and firewalls are countermeasures to this attack type (Maranhão et al., 2021). A host-based IDS monitors and analyses the computing system internals. A general IDS utilizes a database of the system objects that it should be monitoring (Borkar et al., 2017).

Data mining approaches are utilized for implementing ML in IDS to detect anomalies and misuse. In detecting abuse, training data are labelled as "intrusion" or "normal". After that, the ML classifier is modelled to detect anomalies and known intrusions (Azizan et al., 2021). ML represents a suitable extraction of hidden predictive information entirely captured or stored in primary data centres. Lately, many commercial and free data mining and analysis tools have been designed to solve classification and prediction problems across fields (Khalaf et al., 2019).

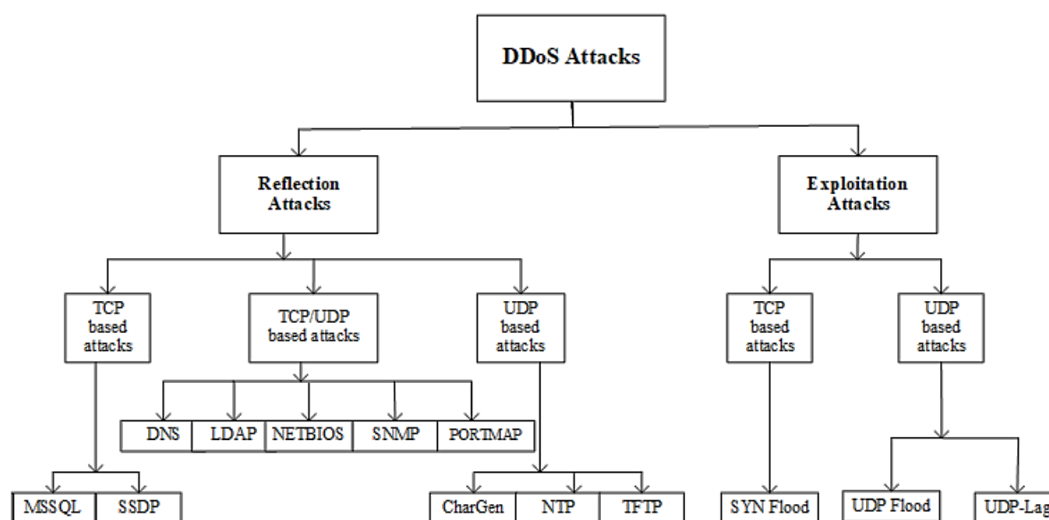


Figure 1. DDoS attack taxonomy. Source: (Sharafaldin et al., 2019).

A classifier is an algorithm or a technique used to group or classify available data into various categories depending on some characteristics. In ML, a classifier is used to learn the pattern in available data, and then the classifier is used to classify the data into various groups according to the patterns (Saini et al., 2020).

3.1 CICDDoS2019 dataset description

The CICDDoS2019 dataset contains 50,063,112 records, including 56,863 rows for benign traffic and 50,006,249 for DDoS attacks, in which every row has 83 features. The attack statistics in the dataset training and testing are listed in Table 1. The training dataset includes 12 types of DDoS attacks: DNS, network time protocol (NTP), Microsoft SQL Server (MSSQL), lightweight directory access protocol (LDAP), simple network management protocol (SNMP), network basic input output system (NetBIOS), user datagram protocol (UDP), simple service discovery protocol (SSDP), Waldos, UDP-Lag, TFTP and SYN, whereas the testing dataset includes seven attacks: Net-BIOS, MSSQL, LDAP, PortScan, UDP, SYN and UDP-Lag in the testing day (Sharafaldin et al., 2019; João Paulo et al., 2021), as shown in Table 2.

Table 2. Classes of CICDDoS2019. Based on (Sharafaldin et al., 2019; Elsayed et al., 2020).

Classes	Flow count
Benign	56,863
DDoS_NetBIOS	4,093,279
DDoS_SNMP	5,159,870
DDoS_NTP	1,202,642
DDoS_TFTP	20,082,580
DDoS_SSDP	2,610,611
DDoS_SYN	1,582,289
DDoS_UDP-Lag	366,461
DDoS_DNS	5,071,011
DDoS_MSSQL	4,522,492
DDoS_LDAP	2,179,930
DDoS_UDP	3,134,645
DDoS_WebDDoS	439

3.2 Machine learning algorithms

The selected ML algorithms in this work are decision tree (DT) J48 type and random forest (RF). They are selected to perform classification of DDoS attacks as they are popular candidates for supervised and ensemble learning algorithms.

- **Decision tree (DT):** The DT J48 is an excellent ML algorithm for categorically and continuously examining data. The classification process is modelled using a binary tree and implemented in all database tuples. J48 is used to classify various applications and presents accurate classification results: J48 represents one of the best ML algorithms utilized for continuous and categorical data examination (Kousar et al., 2021). All aspects of information are split into smaller subsets to base a decision. J48 considers standardized gain of data, representing the result of information splitting by selecting an attribute. Smaller subsets are received by the algorithm once again. Split strategies are terminated when a subset has an index with a similar class in every one of the instances (Khalaf et al., 2019).
- **Random forest (RF):** The RF can be defined as a supervised learning approach representing part of a decision tree-based approach. The RF is a collection of decision trees where a random vector sample produces every classifier from the input vector (Alduailij et al., 2022). It results in several trees, one for each of the features in training data. The decision tree produces aggregated results, and the optimal estimator is considered. The benefit of the RF classifier is that its running time

values are unbalanced and concise, and the missing data are treated (Disha and Waheed, 2022). In RF, the new dataset or testing data are distributed to all the sub-trees that have been created. Each decision sub-tree in this forest can decide the dataset classes.

3.3 Dimensionality reduction techniques

Dimensionality reduction (DR) can be defined as the operation of transforming a high-dimensional data representation into low dimension representations. With a massive growth in high-dimensional data, various DR techniques have become popular in many areas (Alharbi et al., 2021). Reducing dimensionality is one common method of removing redundant noise features. It fetches irrelevant data, which disturbs the operation and performance by decreasing the sample feature ratios. The gene expression datasets are of a high dimensionality, which leads to heavy computation weight as well as decreasing performance of the classification model algorithms (Arowolo et al., 2021). The DR may be carried out using feature extraction (transformation) and feature selection. Feature selection has been utilized to reduce the dimensionality impact on the dataset by finding a subset of features that efficiently define the data (Zong and Huang, 2021). Initially, the evaluation approaches in the area of feature selection have been of 4 types: filter, embedded, hybrid and wrapper. In the present study, the feature extraction algorithms utilize LDA and PCA (Fazili et al., 2020).

- **Principal component analysis (PCA):** PCA can be defined as an unsupervised linear approach commonly utilized to reduce data dimensions. It can preserve statistical information of data as much as possible by embedding data in low-dimensional linear space (Han and Zhipeng, 2020). PCA projects feature spaces from high to lower dimensions by reconstructing k-dimensional unrelated features from the n-dimensional feature of the area (Reddy et al., 2020). PCA presents a variety of strategies for DR, reducing the feature space and preserving the maximum variance amount of original data. PCA can be calculated using a variety of algorithms, including eigenvalues, factor analysis, latent variable analysis or LR (Ayesha et al., 2020). With data variance as standard, PCA performs a measure of the amount of information in a dataset. Higher variance corresponds to a larger amount of data. The calculation of PCA includes singular value decomposition and transformation. The original high-dimensional data are mapped to a linear subspace that is formed by a small number of the principal components with relatively higher eigenvalues during data projection (Li et al., 2020).
- **Linear discriminant analysis (LDA):** LDA is another popular method of DR for preprocessing in ML and data mining applications. The primary goal of LDA is to project a dataset with many features onto a space of lower dimensionality, which will reduce the computational costs (Peng et al., 2020). LDA benefits from the eigenvalue decomposition of scattering matrices based on the whole dataset. It operates by minimizing within-class scatter and maximizing between-class scatter of training data. A lot of data must be provided to get the scatter matrices to produce good data representations (Navya and Savakis, 2021).

4 Solutions and Results

4.1 Model design and setting

The present study aims to build a precise model of DDoS detection using the CICDDoS2019 dataset for the model training. This dataset has 83 features. It includes 12 different DDoS attack types. The dataset is quite large and of high dimensionality. It has a mix of data sources that are integrated into one file, which becomes 50,063,112 records. The resulting dataset is utilized for training and testing the DDoS IDS attack model. The proposed DDoS IDS model has four main modules: preprocessing, feature selection, dimensionality reduction and classification, as shown in Figure 2 and described below.

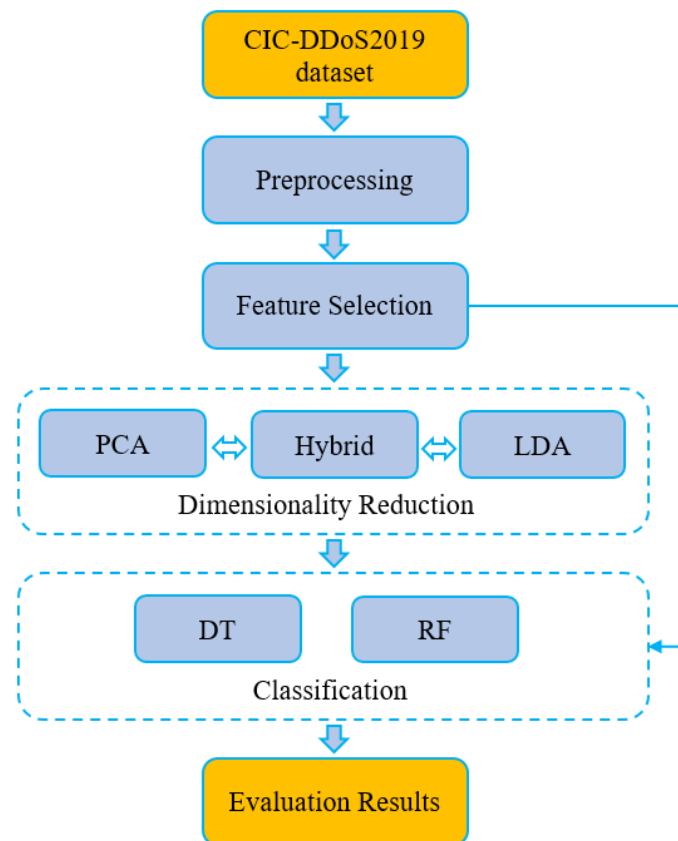


Figure 2. DDoS detection model.

- **Data preprocessing:** Choosing the appropriate algorithms for data preprocessing has the greatest impact on improving any ML results. The dataset must be cleaned to remove undesirable information, and means and commotion are eliminated from the dataset. Then it should be standardized, parsed and reduced before it is utilized for training and testing the model. Preprocessing entails eliminating noise, handling missing values and transforming non-numeric values into numeric ones. The unmitigated information is changed into mathematical values to achieve better accuracy. All the data components, such as symbolic features, are changed from their straight-out factors into a form that can be utilized to estimate all the features precisely and can be used with ML. The symbolic features in network-related data are the source and destination IPs, the flow ID and the protocol that is utilized by the network. Encoding categorical data has been found to enhance system results and accelerate the normalization step.
- **Feature selection:** Deciding on the best subset of features, which results in elevated detection of the required pattern, is very important in the classification. Repetitive and unimportant features can discompose classifying incoming network traffic. The extra trees classifier (ETC) is a feature selection algorithm that has a randomized and effective best feature determination strategy. The ETC is used in the DDoS detection model to pick the top attributes/features out of 86 features available in the dataset. This step is only valid for the first test to generate more reliable benchmark test results.
- **Dimensionality reduction (DR):** DR entails converting the data from high dimensions to low dimensions to improve the data properties. The PCA and LDA techniques perform the DR in individual and hybrid modes to process and improve the data. These algorithms have been applied after carrying a log 2 step to the CICDDoS2019 dataset over eight separate feature reduction options (5, 10, 15, 20, 25, 30, 35, 40).
- **Classification:** The DDoS detection model performs the data classification with the aid of two separate classifiers: DT and RF. These classifiers run parallel and build a different model based on

the training dataset. They are used to detect the DDoS attack before and after data dimensionality reduction in order to verify the best dimensional reduction algorithm to be adapted in the DDoS detection model.

4.2 Results and discussion

The experimental results are discussed in this section. It includes four test case scenarios with and without the DR algorithms. The DT (J84) and RF classifiers are used for DDoS attack detection. The CICDDoS2019 dataset is divided into 70% training and 30% testing data for all the performed test cases. The algorithm training phase includes detecting 12 types of DDoS attacks, and the testing phase includes detecting seven types of DDoS attacks (Sharafaldin et al., 2019). The four test cases are implemented as follows:

Test 1: First, the CICDDoS2019 dataset is examined without DR. This test is used as a benchmark for the DR tests. The ETC selects only the best of the 83 features for the first test. The following test uses DR algorithms. Then the J84 and RF classify seven types of DDoS attacks. The classification accuracy for these algorithms shows that the RF algorithm performs slightly better than J48. Figure 3 shows the performance of the abovementioned model on the dataset based on accuracy. As shown in Figure 3, the RF algorithm accuracy is 91.1%, and that of the J48 is 89.38%.

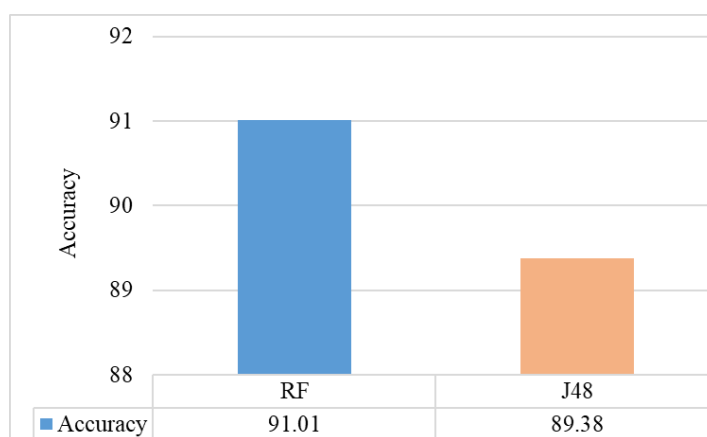


Figure 3. Accuracy without DR.

Test 2: In this test, the CICDDoS2019 dataset that is utilized in the suggested model is converted from high to low dimensions using the PCA algorithm. Then the PCA algorithm assesses the dataset with the attributes reduced from 83 to (5, 10, 15, 20, 25, 30, 35, 40) reduction options by applying the RF and J48 algorithms. The classification accuracy for these algorithms shows that the RF algorithm performs slightly better than the J48. Figure 4 shows the performance of the abovementioned algorithms on the dataset based on accuracy after the DR using the PCA algorithm, in which the accuracy rate increases compared to the first test.

Test 3: The dataset dimensions are reduced from 83 to eight different reduction options (5, 10, 15, 20, 25, 30, 35, 40) with the use of LDA. As mentioned earlier, the LDA technique is quite different from PCA. The LDA-based model shows that the RF algorithm performs slightly better than the J48. Nevertheless, the accuracy of the PCA-based model is better than that of the LDA-based model because the PCA algorithm does not interfere with the class label, which reduces class diversity. Figure 5 shows the performance of the LDA-based DDoS detection model.

Test 4: In this test, the hybrid model chooses the best feature of the PCA and the best features of the LDA. The model takes half of the features from each DR algorithm after excluding the overlapping choices and with subsequent reduction attempts (10, 20, 30, 40). In the first attempt, the PCA chooses the first five features to be reduced, and the second five are chosen by the LDA. Then, accordingly, the other features are sent to the classification algorithms, once to the RF and once to the DT J48. The same procedure is

applied for the other reductions (20, 30, 40). The classification accuracy for these algorithms shows that the RF algorithm performs slightly better than the J48 for all the reduction options. Figure 6 shows the performance of the abovementioned algorithms on the dataset based on the accuracy after combining the PCA and LDA reductions. The RF algorithm achieves the best results when the number of reductions equals 40, for which the highest accuracy is 99.97%.

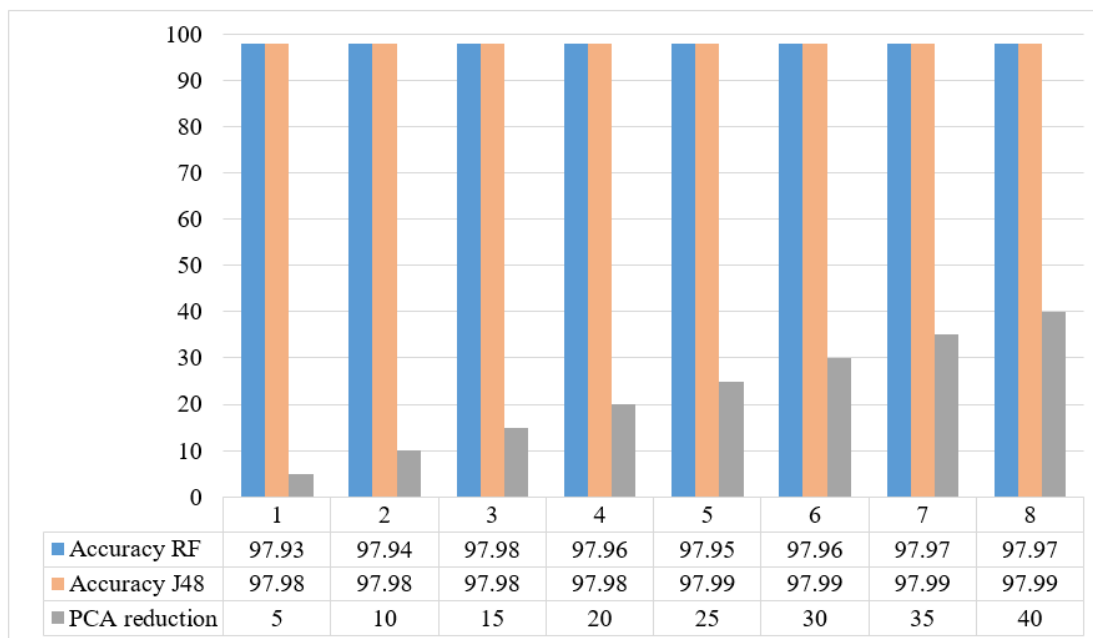


Figure 4. Accuracy of classifiers with PCA DR.

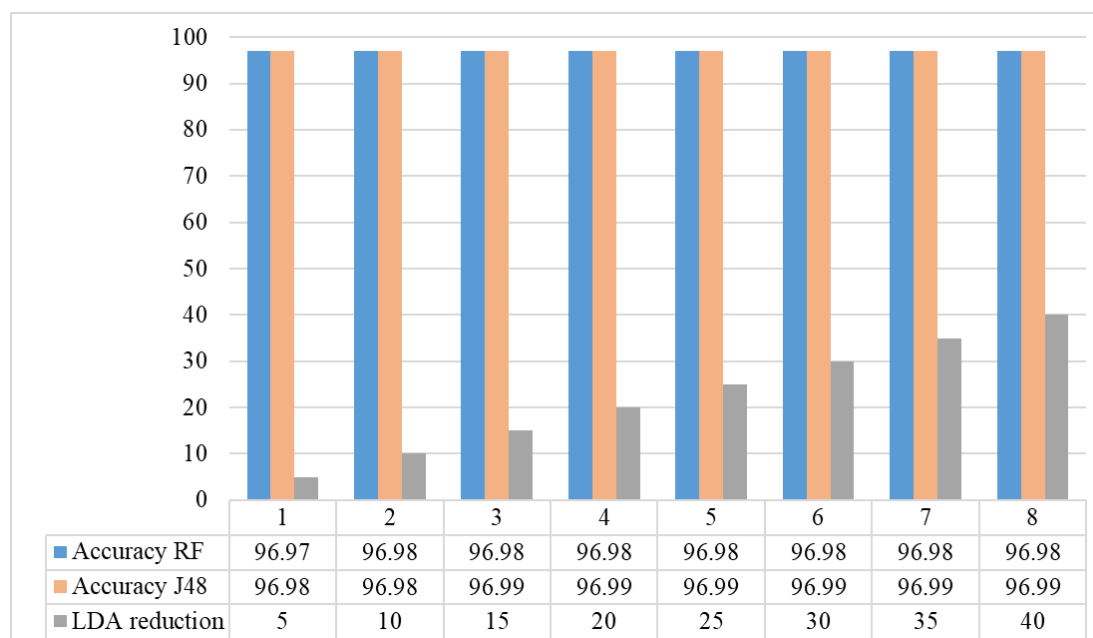


Figure 5. Accuracy of classifiers with LDA DR.

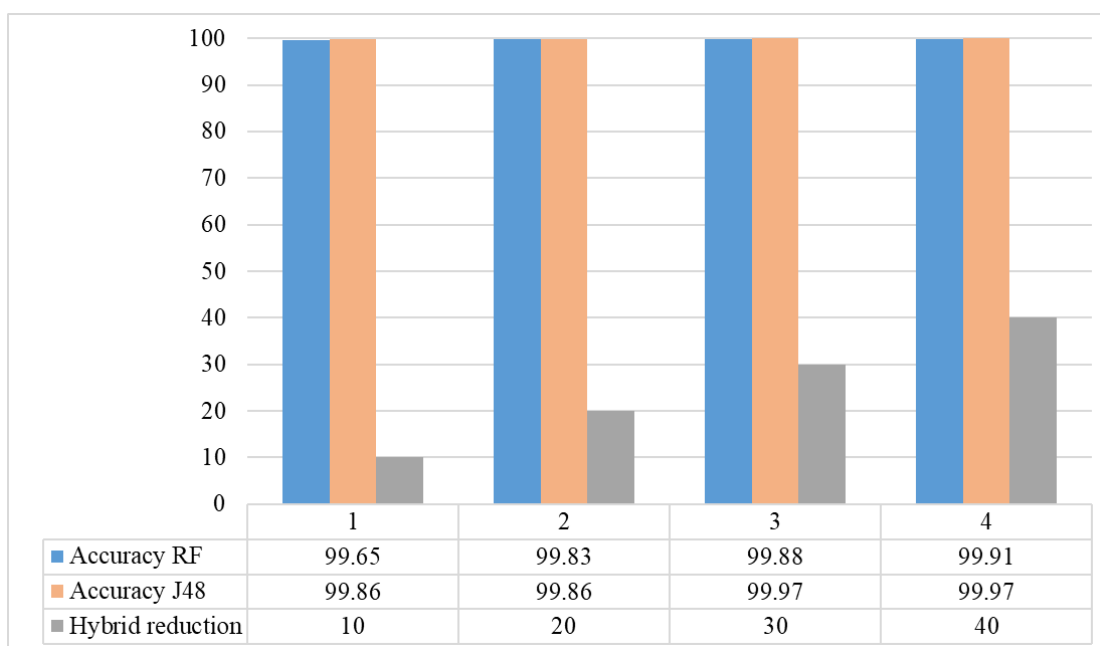


Figure 6. Accuracy of classifiers with hybrid PCA and LDA DR.

5 Conclusion

DDoS attacks are the most fatal and widespread attacks targeting the Internet. Different ML algorithms contribute to building IDS for DDoS attacks. These algorithms accurately detect DDoS attacks by learning the pattern of legitimate and threat requests. Usually, however, these algorithms deal with massive amounts of requests and transmission information, which affects the performance of these algorithms. Therefore, a filtering or data reduction mechanism is required to enhance the performance of the algorithms and reduce the execution and detection times. In this paper, we propose an IDS model that consists of two ML algorithms of DT and RF and two DR algorithms of PCA and LDA for DDoS attack detection. The results show that the suggested model took the best features after reducing the data dimensions, which eventually increased the accuracy rate. The RF algorithm with a combined PCA and LDA achieves the best accuracy result of 99.97% when the number of reductions equals 40.

Future work may include using the hybrid DR algorithm to optimize the features of deep learning DDoS detection models. Also, applying the hybrid DR algorithm to other DDoS attack datasets can further evaluate the algorithm robustness. Finally, we are looking for an opportunity of integrating and testing the model on real software such as firewall or network intrusion prevention software.

Additional Information and Declarations

Acknowledgments: We would like to express our heartfelt gratitude to the people in the College of Engineering at the University of Information Technology and Communications and the College of Information Engineering at the Al-Nahrain University for their support of this research. Additionally, the authors would like to thank the Center of Intelligent and Autonomous Systems (CIAS), Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia (UTHM), for supporting this work.

Funding: This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Grant TIER 1, Vot No. H786.

Conflict of Interests: The authors declare no conflict of interest.


Author Contributions: S.A.D.: Conceptualization, Methodology, Software, Data curation, Writing – Original draft preparation, Visualization, Investigation, Software, Validation, Writing – Reviewing and Editing. S.M.A.: Conceptualization, Methodology, Software, Data curation, Writing – Original draft preparation, Visualization, Investigation, Software, Validation, Writing – Reviewing and Editing. S.M.: Conceptualization, Methodology, Software, Data curation, Writing – Original draft preparation, Visualization, Investigation, Supervision, Software, Validation, Writing – Reviewing and Editing.

Data Availability: The data that support the findings of this study is available online and has a proper citation in the article.

References

- Abbas, S. A., & Almhanna, M. S. (2021). Distributed Denial of Service Attacks Detection System by Machine Learning Based on Dimensionality Reduction. *Journal of Physics: Conference Series*, 1804(1), 012136. <https://doi.org/10.1088/1742-6596/1804/1/012136>
- Abdulrahman, A. A., & Ibrahim, M. K. (2018). Evaluation of DDoS Attacks Detection in a CICIDS2017 Dataset Based on Classification Algorithms. *Iraqi Journal of Information and Communications Technology*, 1(3), 49–55.
- Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M., & Malik, F. (2022). Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method. *Symmetry*, 14(6), 1095. <https://doi.org/10.3390/sym14061095>
- Alharbi, Y., Alferaidi, A., Yadav, K., Dhiman, G., & Kautish, S. (2021). Denial-of-Service Attack Detection over IPv6 Network Based on KNN Algorithm. *Wireless Communications and Mobile Computing*, 2021, Article ID 8000869. <https://doi.org/10.1155/2021/8000869>
- Arowolo, M. O., Adebisi, M. O., Adebisi, A. A., & Olugbara, O. (2021). Optimized hybrid investigative based dimensionality reduction methods for malaria vector using KNN classifier. *Journal of Big Data*, 8(1), 1-14. <https://doi.org/10.1186/s40537-021-00415-z>
- Ayesha, S., Hanif, M. K., & Talib, R. (2020). Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion*, 59, 44–58. <https://doi.org/10.1016/j.inffus.2020.01.005>
- Azizan, A. H., Mostafa, S. A., Mustapha, A., Foozy, C. F. M., Wahab, M. H. A., Mohammed, M. A., & Khalaf, B. A. (2021). A Machine Learning Approach for Improving the Performance of Network Intrusion Detection Systems. *Annals of Emerging Technologies in Computing*, 5(5), 201–208. <https://doi.org/10.33166/aetic.2021.05.025>
- Bhaya, W., & EbadyManaa, M. (2017, March). DDoS attack detection approach using an efficient cluster analysis in large data scale. In *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)* (pp. 168–173). IEEE. <https://doi.org/10.1109/NTICT.2017.7976110>
- Borkar, A., Donode, A., & Kumari, A. (2017, November). A survey on Intrusion Detection System (IDS) and Internal Intrusion Detection and protection system (IIDPS). In *2017 International conference on inventive computing and informatics (ICICI)* (pp. 949–953). IEEE. <https://doi.org/10.1109/ICICI.2017.8365277>
- Disha, R. A., & Waheed, S. (2022). Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique. *Cybersecurity*, 5(1). <https://doi.org/10.1186/s42400-021-00103-8>
- Elsayed, M. S., Le-Khac, N. A., Dev, S., & Jurcut, A. D. (2020, August). DDoSNet: A deep-learning model for detecting network attacks. In *2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)* (pp. 391–396). IEEE. <https://doi.org/10.1109/WoWMoM49955.2020.00072>
- Fazili, S., Grover, J., Wazir, S., & Mehta, I. (2021). Recent Trends in Dimension Reduction Methods. In *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development*. EUDL. <https://doi.org/10.4108/eai.27-2-2020.2303136>
- Ferrag, M. A., Shu, L., Djallel, H., & Choo, K.-K. R. (2021). Deep Learning-Based Intrusion Detection for Distributed Denial of Service Attack in Agriculture 4.0. *Electronics*, 10(11), 1257. <https://doi.org/10.3390/electronics10111257>
- Han, J., & Ge, Z. (2020). Effect of dimensionality reduction on stock selection with cluster analysis in different market situations. *Expert Systems with Applications*, 147, 113226. <https://doi.org/10.1016/j.eswa.2020.113226>
- Hezam, A. A., Mostafa, S. A., Baharum, Z., Alanda, A., & Salikon, M. Z. (2021). Combining Deep Learning Models for Enhancing the Detection of Botnet Attacks in Multiple Sensors Internet of Things Networks. *JOIV: International Journal on Informatics Visualization*, 5(4), 380–387. <https://doi.org/10.30630/joiv.5.4.733>
- Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abdulllah, W. M. (2019). Comprehensive Review of Artificial Intelligence and Statistical Approaches in Distributed Denial of Service Attack and Defense Methods. *IEEE Access*, 7, 51691–51713. <https://doi.org/10.1109/ACCESS.2019.2908998>
- Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., Mahmoud, M. A., Al-Rimy, B. A. S., Abd Razak, S., Elhoseny, M., & Marks, A. (2021). An Adaptive Protection of Flooding Attacks Model for Complex Network Environments. *Security and Communication Networks*, 2021, Article ID 5542919. <https://doi.org/10.1155/2021/5542919>

- Kousar, H., Mulla, M. M., Shettar, P., & Narayan, D. G. (2021, June). Detection of DDoS Attacks in Software Defined Network using Decision Tree. In *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 783-788). IEEE. <https://doi.org/10.1109/CSNT51715.2021.9509634>
- Li, M., Wang, H., Yang, L., Liang, Y., Shang, Z., & Wan, H. (2020). Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction. *Expert Systems with Applications*, 150, 113277. <https://doi.org/10.1016/j.eswa.2020.113277>
- Maranhão, J. P. A., da Costa, J. P. C. L., Javidi, E., de Andrade, C. A. B., & de Sousa, R. T. (2021). Tensor based framework for Distributed Denial of Service attack detection. *Journal of Network and Computer Applications*, 174, 102894. <https://doi.org/10.1016/j.jnca.2020.102894>
- Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., & Foozy, C. F. M. (2021). Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset. *IEEE Access*, 9, 22351–22370. <https://doi.org/10.1109/access.2021.3056614>
- Nagananda, N., & Savakis, A. (2021). GILDA++: Grassmann Incremental Linear Discriminant Analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4453-4461). https://openaccess.thecvf.com/content/CVPR2021W/DiffCVML/html/Nagananda_GILDA_Grassmann_Incremental_Linear_Discriminant_Analysis_CVPRW_2021_paper.html
- Novaes, M. P., Carvalho, L. F., Lloret, J., & Proença, M. L. (2020). Long Short-Term Memory and Fuzzy Logic for Anomaly Detection and Mitigation in Software-Defined Network Environment. *IEEE Access*, 8, 83765–83781. <https://doi.org/10.1109/ACCESS.2020.2992044>
- Peng, X., Ma, Z., & Xu, H. (2020). Maximum Discriminant Difference Criterion for Dimensionality Reduction of Tensor Data. *IEEE Access*, 8, 193593–193607. <https://doi.org/10.1109/access.2020.3032346>
- Priya, S. S., Sivaram, M., Yuvaraj, D., & Jayanthiladevi, A. (2020, March). Machine learning based DDoS detection. In *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 234-237). IEEE. <https://doi.org/10.1109/ESCI48226.2020.9167642>
- Rahal, B. M., Santos, A., & Nogueira, M. (2020). A Distributed Architecture for DDoS Prediction and Bot Detection. *IEEE Access*, 8, 159756–159772. <https://doi.org/10.1109/access.2020.3020507>
- Rahman, O., Quraishi, M. A. G., & Lung, C. H. (2019, July). DDoS attacks detection and mitigation in SDN using machine learning. In *2019 IEEE world congress on services (SERVICES)* (Vol. 2642, pp. 184-189). IEEE. <https://doi.org/10.1109/SERVICES.2019.00051>
- Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access*, 8, 54776–54788. <https://doi.org/10.1109/ACCESS.2020.2980942>
- Saini, P. S., Behal, S., & Bhatia, S. (2020, March). Detection of DDoS attacks using machine learning algorithms. In *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 16-21). IEEE. <https://doi.org/10.23919/INDIACom49435.2020.9083716>
- Sharafaldin, I., Lashkari, A. H., Hakak, S., & Ghorbani, A. A. (2019, October). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *2019 International Carnahan Conference on Security Technology (ICCST)* (pp. 1-8). IEEE. <https://doi.org/10.1109/CCST.2019.8888419>
- Sharma, V., Verma, V., & Sharma, A. (2019, June). Detection of DDoS attacks using machine learning in cloud computing. In *International Conference on Advanced Informatics for Computing Research* (pp. 260-273). Springer, Singapore. https://doi.org/10.1007/978-981-15-0111-1_24
- Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-To-Peer Networking and Applications*, 12(2), 493–501. <https://doi.org/10.1007/s12083-017-0630-0>
- Tuan, T. A., Long, H. V., Son, L. H., Kumar, R., Priyadarshini, I., & Son, N. T. K. (2020). Performance evaluation of Botnet DDoS attack detection using machine learning. *Evolutionary Intelligence*, 13(2), 283-294. <https://doi.org/10.1007/s12065-019-00310-w>
- Yungaicela-Naula, N. M., Vargas-Rosales, C., & Perez-Diaz, J. A. (2021). SDN-Based Architecture for Transport and Application Layer DDoS Attack Detection by Using Machine and Deep Learning. *IEEE Access*, 9, 108495–108512. <https://doi.org/10.1109/ACCESS.2021.3101650>
- Zhang, B., Zhang, T., & Yu, Z. (2017, December). DDoS detection and prevention based on artificial intelligence techniques. In *2017 3rd IEEE International Conference on Computer and Communications (ICCC)* (pp. 1276-1280). IEEE. <https://doi.org/10.1109/CompComm.2017.8322748>
- Zhijun, W., Qing, X., Jingjie, W., Meng, Y., & Liang, L. (2020). Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network. *IEEE Access*, 8, 17404–17418. <https://doi.org/10.1109/access.2020.2967478>
- Zong, Y., & Huang, G. (2019). A feature dimension reduction technology for predicting DDoS intrusion behavior in multimedia internet of things. *Multimedia Tools and Applications*, 80(15), 22671–22684. <https://doi.org/10.1007/s11042-019-7591-7>

Editorial record: The article has been peer-reviewed. First submission received on 13 August 2022. Revision received on 15 October 2022. Accepted for publication on 11 November 2022. The editor in charge of coordinating the peer-review of this manuscript and approving it for publication was Stanislav Vojir .

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951
