Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz. *Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection*
To appear in the Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2022.

Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection

Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz
Multimedia Communications Lab
Technical University of Darmstadt, Germany
{pegah.golchin, ralf.kundel, tim.steuer, rhaban.hark, ralf.steinmetz}@kom.tu-darmstadt.de

***Abstract*—DDoS attack detection is crucial in computer networks to meet the reliability and accessibility requirements of online services. The ability of machine learning to discriminate between DDoS attacks and benign flows makes it a promising candidate for DDoS detection. Correctly classifying the flows with high performance in near real-time is a critical issue for an ML-based DDoS detector to reduce the damages of DDoS attacks. In order to improve the performance of classification and reduce the prediction time, we propose a multi-aspect Ensemble Feature Selection (EFS) for DDoS attack detection in this work. The presented EFS selects the most relevant features of each attack separately, leveraging a combination of statistical filtering approaches and machine learning methods. We evaluate our method on two different datasets to demonstrate the EFS robustness toward model-specific biases. Last, we demonstrate that the prediction time is reduced leveraging the proposed EFS.**

I. Introduction

The ever-increasing number of Distributed Denial-of-Service (DDoS) attacks is leading to a considerable problem in today's computer networks. DDoS attack is a kind of cyber-attack that attempts to disrupt legitimate users' access by overwhelming the networking and computing resources of the target. According to the Worldwide Infrastructure Security Report [1], the frequency of DDoS attacks increases with peak bandwidth usage of up to 1.12 *TBPS* in 2020, which is dominating in the field of cyber-attacks. There are various reasons to launch a DDoS attack, including financial gains, revenge, and intellectual challenges [2].

Intrusion Detection Systems (IDS) can predict, detect and mitigate DDoS attack attempts. However, DDoS attacks have become highly sophisticated. Therefore, IDS technological approaches such as entropy-based or predetermined rules are no longer adequate for DDoS detection [3]. Even though not being detectable with the aforementioned methods, most DDoS attacks have common prominent features that allow discriminating between attack and benign flows with sufficient precision.

The ability of Machine Learning (ML) approaches to learn linear and non-linear relations makes it an appropriate method to detect anomalous network behavior [4]. Furthermore, its advantages over statistical and payload inspection-based techniques allow a more precise DDoS attack detection. Consequently, there are some ML-based approaches for IDS to improve the accuracy and reliability in terms of DDoS attack mitigation.

The first step for the ML-based DDoS detector is gathering the network information to classify DDoS attack and benign flows [5]. Therefore, the DDoS detector faces a large amount of network data to be processed as the network extends greatly [6]. It increases the computational complexity and classification time.

Applying Feature Selection (FS) to speed up this classification process is a promising way to decrease the classification time and computational complexity. FS can extract a subset of relevant features to reduce the feature space dimension. In addition, eliminating non-relevant features such as noisy features can have a positive impact on the accuracy of the IDS [7]. Further, it improves the generalization of classifiers while decreasing the overfitting problem.

In this work, we use the CICDDoS2019 [8] and the InSDN [9] datasets, including reflection- and exploitation-based attacks. Moreover, we investigate Exploratory Data Analysis (EDA) to understand the distribution of features for each attack available in these datasets. The outcome of EDA presents different prominent features for each attack. Build upon this, we present a multi-aspect Ensemble Feature Selection (EFS), which extracts the essential features of each attack separately. The proposed EFS combines statistical methods such as Pearson correlation and variance filtering with different types of ML models, including Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM). The features with higher contributions are selected to reduce the variance of the classifiers, which can improve the generalization. Last, we apply the proposed EFS method on CICDDoS2019 and the InSDN datasets and evaluate DDoS attack detection utilizing ML models. Our evaluation results show that the EFS method reduces the feature space and computation time without losing classification accuracy. These results can be applied in real networks to improve the DDoS attack mitigation, *e.g.* in 5G networks running on high performance networking hardware [10].

The rest of this paper is organized as follows: Section II discusses related work and the differences to our approach in detail. In Section III the proposed method will be introduced. The evaluation scheme and the results are provided in Section IV. Finally, Section V draws the conclusion and provides an outlook on possible future works.

Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz. *Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection*
To appear in the Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2022.

II. Related work

FS is an ML preprocessing phase with the aim of improving detection accuracy while reducing computational complexity [11]. In [12], an overview of Filter, Wrapper, Embedded FS methods is performed and applied on various standard datasets. In [13], a brief survey on FS methods is provided, and the taxonomies of ensemble filter-based selectors is explained. The authors in [14], investigate different EFS methods, and they claim that utilizing the EFS methods can improve the ML performance compared to using a single FS method.

While the above studies provide surveys on FS and EFS in general, the following works address DDoS attack classification utilizing EFS:

In [15], the potential features were selected by comparing the calculated affinity of the selected features with a predefined threshold. Although this method is efficient, it depends strongly on the predefined threshold. In addition, new application-layer attacks which are more sophisticated and similar to benign flows were not considered. In [16], a combination of Information Gain (IG) and correlation methods is leveraged to select the best features efficiently. Their method was applied on the CICDDoS2019 dataset utilizing the J48 classifier. Even though IG is implemented to determine the relevance of attributes, it faces a bias problem with features having large values in the dataset. In addition, they applied their proposed FS on the dataset consisting of all attacks. However, as the distribution of features for each attack is different, applying a method on the entire dataset might cause ignoring some important features of few attacks and reduce the robustness of the FS method.

A deep learning-based FS method is proposed in [17], utilizing a Multi-Layer Perceptron (MLP) network to select and learn important weights for the input feature set. This work calculated the context attention score based on the feature set and its values and transformed it into a weight. Then the authors multiplied it with the correspondent feature value. Although selecting features leveraging MLP might improve the bias issue available in IG, utilizing MLP for feature selection makes the method very complex in terms of computational resource and time consumption. Also, in this work, the authors applied FS on the entire dataset leading to eliminating some important features of the minority attacks.

III. Proposed Method

In this section, we provide an overview of the preprocessing, EDA, and the proposed multi-aspect EFS algorithm.

A. Preprocessing

Data preprocessing is a crucial step as the quality of data can affect the ability of the model to optimize an objective function. The raw files in the CICDDoS2019 and the InSDN datasets consist of 88 features with various ranges and types of values. The flow-level based features in the datasets are collected utilizing the passive monitoring method which leads to extract more features. In this work, the features types are converted to floating numbers and categorical features like *protocol* are transformed to numeric ones using one-hot encoding. Since gradient descent-based algorithms such as logistic regression and neural networks, and distance-based algorithms such as SVM are sensitive to the different features scale, z-score standardization is applied [18].

B. Exploratory Data Analysis

We execute an EDA on the selected dataset in order to investigate the feature distribution in different attacks. Firstly, we check the linear correlation between features leveraging Pearson correlation coefficient [19]:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{1}$$

Here, r represents the Pearson correlation, x and y are features, and $\overline{x}$ and $\overline{y}$ are the mean of x and y. The result demonstrates a high positive and negative linear correlation among some features. Therefore, we filter high correlated features using a 95% threshold.

Furthermore, we investigate available protocols in the dataset. There are three labels for transport layer protocols, including TCP (*protocol-6*), UDP (*protocol-17*), and HOPOPT (*protocol-0*) for non-understandable IPv6 extension headers, which are employed while creating this dataset. Considering the protocol distribution in attacks, the majority of malicious flows exploited UDP except for SYN attack which is in TCP flood category.

In addition to the protocols, after investigating the distribution of TCP flags, it becomes clear that the distribution of the ACK flag in the SYN attack is different from the other attacks. The mean and standard deviation of the ACK flag in the SYN attack are 0.9994 and 0.0228, respectively, while for all other attacks, the values are 0.0097 and 0.0984, respectively. As we expect, It is one of the prominent features in SYN attacks because it uses the TCP three-way handshake vulnerability and needs to modify the ACK flag. As exploitation attacks such as UDP and TCP floods are also available in this dataset, we investigate the forward packets count distribution. The mean and standard deviation of the forward packets count for the flooding attacks are 258.0164 and 5.1572, respectively, while these values for the other kind of attacks are 7.6624 and 3.4625, respectively. This is what we expect from the flooding attacks to have more forward packets compared to other kinds of attacks. Moreover, as the automated tools mostly produce DDoS attack flows, including fix-size packets, compared to benign flows that have different packet lengths, the ”packet length std” feature might be a prominent feature for benign flows. The mean and standard deviation values of the ”packet length std” feature for benign flows are 88.2735 and 230.5472, respectively, while these values for DDoS attacks are 9.1670 and 14.4972, respectively. In addition, the dataset is highly imbalanced because most of the flows have the attack label while the minority has the benign label. About 40% of the entire dataset is TFTP attack, while benign flows are about 1% of the entire dataset, and the rest 59% are the other attacks.

Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz. *Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection*
To appear in the Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2022.

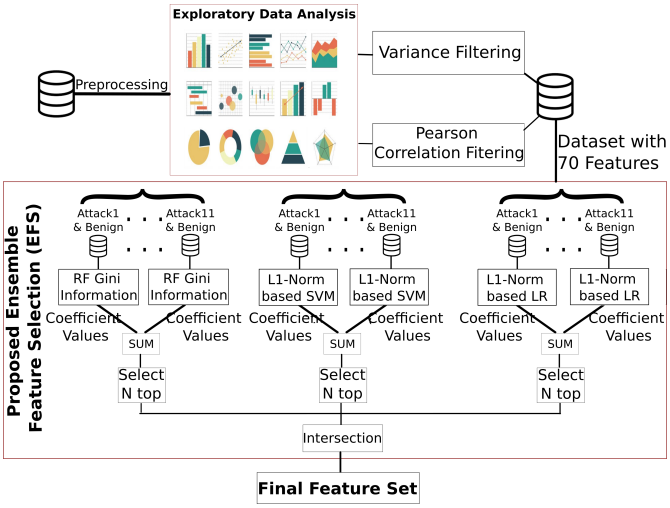

Fig. 1. Proposed multi-scale Ensemble Feature Selection. The figure illustrates filtering, ML-based feature reduction, and final feature selection.

Therefore, it is essential to ensure that the trained model is not biased toward the high frequent class, namely DDoS attacks.

C. Proposed Feature Selection

The ideal DDoS detector can precisely discriminate between DDoS attacks and benign flows with high speed and low complexity. In order to develop a DDoS detector close to the ideal one, we propose a new ensemble FS (EFS) algorithm.

According to the EDA outcomes, some features such as TCP flags, protocols, and number of forward packets have different distributions in each attack. Therefore, we argue that identifying the relevant features for each attack can improve the classification performance. In fact, FS on all attacks together may result in statistically biased FS due to the different sample sizes of each attack.

As we show in the proposed EFS in Fig. 1, the first step is using Pearson correlation filtering based on Equation 1. One of the features out of a pair that has more than 95% correlation was removed. In addition, we apply variance filtering to remove the features with zero variance, such as PSH or URG flags. According to Section III-B (EDA), only SYN attack flows exploit TCP protocol. Since it uses the three-way handshake vulnerability, the ACK flag is essential for this attack not the other flags. As a result, leveraging these two filtering methods, the feature space is reduced from 88 features to 70 features. In the second stage, we apply L1-norm-based on linear classifiers to select the non-zero coefficient values and obtain the most relevant features for each attack. Among ML methods, linear classifiers like LR and SVM construct sparse solutions using an L1-norm-based cost function [20].

According to Equation 2, a penalty term is added to the cost function to reduce the weight of irrelevant features in general [21].

$$E(w) = \frac{1}{m}\sum_{i=1}^{m} Cost(\widehat{y}_i - y_i) + \alpha \sum_{j=1}^{d} |w_j| \qquad (2)$$

Here E represents the cost function, w is a weight coefficient, y is the true label, $\widehat{y}$ is the predicted value, d is the dimension of features, m is the number of samples, and α is the controller parameter. In order to collect features using this method, we choose LR and linear SVM as a linear classifier and train them for each attack on the CICDDoS2019 training dataset.

Furthermore, we collect another set of features using RF as it is one of the accurate and powerful learning methods, and it can extract the most relevant features using Gini importance I_G of the features, which is calculated as follows [22]:

$$I_G(\theta) = \sum_{T}\sum_{\tau} \Delta i_\theta(\tau - T), \qquad (3)$$

where θ demonstrates a specific feature for which I_G is calculated and $\Delta i_\theta(\tau - T)$ gives information about the number of times that θ is selected for a split at each node τ within the binary tree of T. Therefore, it is an important ranking value for features.

In the next step, we take the sum over all coefficient values extracted from each of these three FS methods on each DDoS attack and benign flows separately and select the top 20, 10, and 5 features, respectively. Last, we make an intersection over the selected top features of each FS method, which results in 34, 23, and 14 features as the final feature set.

IV. Evaluation

In order to evaluate the proposed EFS method, we utilize five well-known ML models, including RF, LR, SVM, Naive Bayes (NB), and Multi-Layer Perceptron (MLP). The models are trained on the CICDDoS2019 training set, including eleven attacks, and evaluated on the unseen test set, including seven attacks. In addition, we evaluate our EFS method on the InSDN dataset to demonstrate that the performance results are not biased to a specific dataset. For that, the model is trained on 70% of the InSDN dataset, and the performance results are evaluated on the remaining unseen 30% as test set. In order to illustrate the performance of the classifier, we calculate the macro average of Precision, Recall and F1-score to consider equal weights for the results of both the majority (attacks) and the minority (benign) classes. In this case, correct prediction of both attack and benign flows is essential, and it makes the evaluation metrics more valid [8]. Table I presents the macro average values of Precision, Recall and F1-score for each classifier on the CICDDoS2019 dataset. It compares the classification performance results of the models with and without feature space reduction leveraging the proposed EFS method. The bold numbers in Table I highlight the best value of the evaluation metric for each classifier. According to Table I, there are improvements on all methods when the feature space is reduced to the relevant features leveraging the proposed EFS. The recall of LR and SVM has improved by reducing the feature space utilizing the proposed EFS. RF and NB classifiers have about 40% improvement in the F1-score when the proposed method is applied, and the classifiers train on 23 features. In fact, there are some features in the raw

Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz. *Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection*
To appear in the Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2022.

TABLE I
PERFORMANCE EVALUATION OF DIFFERENT MODELS ON CICDDOS2019 DATASET WITH AND WITHOUT THE PROPOSED FS.

Model	Without Proposed EFS			Proposed EFS with 34 Features			Proposed EFS with 23 Features			Proposed EFS with 14 Features		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
LR	**96.99**	87.29	91.58	96.52	98.69	**97.58**	95.90	98.96	97.38	95.66	**99.26**	97.39
RF	51.19	93.10	48.63	96.98	**99.53**	98.22	**97.01**	99.50	**98.22**	96.47	99.51	97.94
SVM	**94.39**	92.78	93.57	93.30	99.52	**96.20**	93.28	**99.52**	96.19	93.28	99.51	96.19
NB	51.32	93.70	49.27	51.25	93.14	48.95	91.70	**99.81**	95.39	**92.83**	99.66	**95.99**
MLP	97.48	95.31	96.37	**97.64**	**98.57**	**98.10**	97.28	97.69	97.49	96.55	95.20	95.87

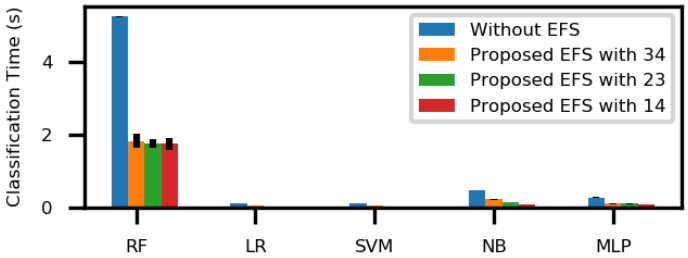

Fig. 2. Comparison of prediction time. It illustrates the mean values and standard deviation of the models classification time.

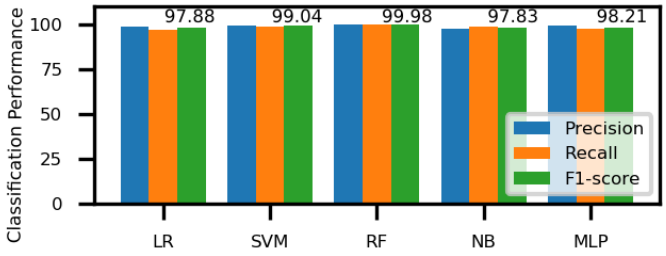

Fig. 3. Analyzing performance evaluation of the proposed EFS method in the InSDN test dataset. The figure illustrates F1-score values higher than 97% for all ML approaches on the 30% unseen InSDN test dataset.

TABLE II
PERFORMANCE OF LR MODEL FOR A NEW ATTACK IN CICDDOS2019 TEST DATASET

FS Method	Test Attacks	precision	recall	F1-score
Without Proposed EFS	Portmap	96.27	87.27	90.80
Proposed EFS and 23 Features	Portmap	99.56	98.94	99.24

dataset that cannot help RF split the node during the training process. MLP has the highest F1-score without applying the proposed EFS because it consists of seven layers. However, there is an improvement in the F1-score of MLP when the proposed EFS is applied. It can demonstrate that the necessary information for the classification task is kept when the feature space is reduced utilizing the proposed EFS.

Another influencing factor of a DDoS detector is the classification time. The measured results of classification time are determined by running the evaluation on a system with an Intel Core i9-10900K CPU, DDR4 memory and 10 runs for each model. Fig. 2 illustrates the observed classification time and its standard deviation. As the standard deviation is very low, the results can be considered to be statistically significant. It is noteworthy that applying EFS shortens the prediction time enormously for some of the classifiers. For instance, the mean of LR classification time is 0.1114 (s), while it reduces to 0.0313 (s) when applying the proposed EFS. Therefore, according to the performance results in Table I and the prediction time in Fig.2, LR with 23 selected features using the proposed EFS can be a good candidate for a practical DDoS detector. In addition, Table II demonstrates the performance evaluation of the LR model for the "PortMap" attack in the CICDDoS2019 test dataset, which is not available in the training data. Therefore, improving the classification performance leveraging the proposed EFS can show the robustness of 23 selected features.

Further, we applied the proposed EFS method on another new dataset called InSDN to investigate whether the performance results in Table I are biased to a specific dataset or not. The InSDN dataset includes benign, and seven completely different attack flows that can occur in the different elements of the SDN platform. We select the most relevant features of this dataset leveraging the proposed EFS.

According to the results of Fig. 3, all the ML approaches obtain the F1-score higher than 97% when the proposed EFS selects 23 relevant features of the InSDN dataset. In this case, we demonstrate that our proposed EFS method does not work only for one DDoS attack dataset but also obtains outstanding results on another new dataset collected in a different network.

V. CONCLUSION

In this work, we proposed a multi-aspect Ensemble Feature Selection (EFS) in order to reduce the feature space and computation time while improving the classification performance. The proposed EFS consists of various statistical filtering and ML approaches to extract a robust feature set and avoid losing useful information. In order to discriminate between DDoS attacks and benign flows, we evaluate the performance of five machine learning models on the CICDDoS2019 and the InSDN datasets. The results demonstrate that the feature space reduction by the proposed EFS improves the classification performance while reducing the classification time. In future works, we will focus on the extension of the presented EFS approach and on analyzing more new attacks which are generated using Generative Adversarial Networks.

ACKNOWLEDGMENT

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany within the CELTIC-NEXT Flagship Project AI-NET-PROTECT and by the German Research Foundation (DFG) within the Collaborative Research Center (CRC) 1053 MAKI.

Pegah Golchin, Ralf Kundel, Tim Steuer, Rhaban Hark, Ralf Steinmetz. *Improving DDoS Attack Detection Leveraging a Multi-aspect Ensemble Feature Selection*
To appear in the Proceedings of the IEEE/IFIP Network Operations and Management Symposium (NOMS), 2022.

REFERENCES

[1] "Netscout report," https://www.netscout.com/report/, [online; accessed on 2021-08-08].

[2] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks," *IEEE communications surveys & tutorials*, vol. 15, no. 4, pp. 2046–2069, 2013.

[3] K. S. Sahoo, B. K. Tripathy, K. Naik, S. Ramasubbareddy, B. Balusamy, M. Khari, and D. Burgos, "An evolutionary svm model for ddos attack detection in software defined networks," *IEEE Access*, vol. 8, pp. 132 502–132 513, 2020.

[4] R. Santos, D. Souza, W. Santo, A. Ribeiro, and E. Moreno, "Machine learning algorithms to detect ddos attacks in sdn," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 16, p. e5402, 2020.

[5] M. H. Aghdam, P. Kabiri *et al.*, "Feature selection for intrusion detection system using ant colony optimization." *Int. J. Netw. Secur.*, vol. 18, no. 3, pp. 420–432, 2016.

[6] A. R. Yusof, N. I. Udzir, A. Selamat, H. Hamdan, and M. T. Abdullah, "Adaptive feature selection for denial of services (dos) attack," in *2017 IEEE Conference on Application, Information and Network Security (AINS)*. IEEE, 2017, pp. 81–84.

[7] A. H. Sung and S. Mukkamala, "Identifying important features for intrusion detection using support vector machines and neural networks," in *2003 Symposium on Applications and the Internet, 2003. Proceedings.* IEEE, 2003, pp. 209–216.

[8] I. Sharafaldin, A. H. Lashkari, S. Hakak, and A. A. Ghorbani, "Developing realistic distributed denial of service (ddos) attack dataset and taxonomy," in *2019 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2019, pp. 1–8.

[9] M. S. Elsayed, N.-A. Le-Khac, and A. D. Jurcut, "Insdn: A novel sdn intrusion dataset," *IEEE Access*, vol. 8, pp. 165 263–165 284, 2020.

[10] R. Kundel, T. Meuser, T. Koppe, R. Hark, and R. Steinmetz, "User plane hardware acceleration in access networks: Experiences in offloading network functions in real 5g deployments," in *Proceedings of the 55th Hawaii International Conference on System Sciences*. Computer Society Press, 2022, p. 1–10.

[11] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 science and information conference*. IEEE, 2014, pp. 372–378.

[12] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[13] O. Osanaiye, H. Cai, K.-K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for ddos detection in cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–10, 2016.

[14] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Information Fusion*, vol. 52, pp. 1–12, 2019.

[15] D. J. Prathyusha and G. Kannayaram, "A cognitive mechanism for mitigating ddos attacks using the artificial immune system in a cloud environment," *Evolutionary Intelligence*, vol. 14, no. 2, pp. 607–618, 2021.

[16] D. Kshirsagar and S. Kumar, "A feature reduction based reflected and exploited ddos attacks detection system," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2021.

[17] D. C. Can, H. Q. Le, and Q. T. Ha, "Detection of distributed denial of service attacks using automatic feature selection with enhancement for imbalance dataset," *ACIIDS 2021*, 2021.

[18] T. Li, B. Jing, N. Ying, and X. Yu, "Adaptive scaling," *arXiv preprint arXiv:1709.00566*, 2017.

[19] P. Schober, C. Boer, and L. A. Schwarte, "Correlation coefficients: appropriate use and interpretation," *Anesthesia & Analgesia*, vol. 126, no. 5, pp. 1763–1768, 2018.

[20] X. Chen, G. Zhou, Y. Chen, G. Shao, and Y. Gu, "Supervised multiview feature selection exploring homogeneity and heterogeneity with $l_{1,2}$ -norm and automatic view generation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 2074–2088, 2017.

[21] S. Guo, D. Guo, L. Chen, and Q. Jiang, "A l1-regularized feature selection method for local dimension reduction on microarray data," *Computational biology and chemistry*, vol. 67, pp. 92–101, 2017.

[22] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, no. 1, pp. 1–16, 2009.