JOURNAL OF BASIC SCIENCE AND ENGINEERING

REGULARIZED FEATURE SELECTION FOR IMPROVED DDOS ATTACK DETECTION A RECURSIVE VARIABLE ELIMINATION APPROACH USING LEAST ABSOLUTE SHRINKAGE AND SELECTION

[1]**Mrs. K. R. Prabha**
PhD Research Scholar, Department of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India.

[2]**Dr. B. Srinivasan**
Associate Professor, Department of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India.

Abstract

The complexity and variety of Distributed Denial of Service (DDOS) attacks make it very difficult to detect them in data collected from network traffic. Although machine learning algorithms have shown some success in this area, data that contains traits that are either strongly linked or irrelevant might limit their efficacy. This study presents a new method for regularised feature selection that uses the LASSO-RFE methodology to improve the accuracy of DDOS attack detection, which will help to alleviate this problem. LASSO-RFE is a two-stage process that combines the best features of LASSO regularisation with RFE. At first, the high-dimensional feature space is subjected to LASSO regularisation in order to determine which characteristics are most important for DDOS attack detection. As the regression coefficients of superfluous or unimportant characteristics approach zero, LASSO removes them from the model. The remaining features chosen by LASSO are then subjected to RFE in order to improve the model's performance and narrow the feature set even more. Until the ideal subset of features is found, RFE repeatedly removes the characteristics that aren't important based on their model weights. Improving the DDOS attack detection models' resilience and interpretability, the suggested technique provides a systematic and efficient way for feature selection. When compared to more conventional machine learning methods, experimental findings show that LASSO-RFE significantly improves the accuracy of DDOS attack detection. In order to improve the efficiency and accuracy of DDOS attack detection in network traffic data, LASSO-RFE reduces the dimensionality of the feature space and efficiently filters out unnecessary characteristics.

Keywords: Distributed Denial of Service, Machine Learning, Network Traffic Data, Recursive Variable Elimination

I. Introduction

Maintaining the availability and integrity of online services is crucial in light of the increasing frequency of Distributed Denial of Service (DDOS) attacks, which pose major risks to network infrastructures [1]. The widespread and organised nature of DDOS attacks makes them a formidable threat, since they can flood target systems with malicious traffic, disrupting services and even compromising security [2]. Because these threats are constantly changing, traditional signature-based detection approaches aren't always enough to stop them. Therefore, methods based on machine learning are becoming more important for DDOS attack detection, since these methods take use of algorithms' capacity to adapt to evolving attack patterns [3-4]. Nevertheless, the efficacy of DDOS attack detection systems trained using machine learning techniques relies on the features' quality and relevance [5-6]. The detection procedure is sometimes made more difficult by the abundance of characteristics found in network traffic data, some of which can be superfluous, redundant, or strongly linked. Reduced interpretability, higher computing cost, and inferior detection performance might result from inaccurate feature selection [7-8].

Based on the Least Absolute Shrinkage and Selection with Recursive Variable Elimination (LASSO-RFE) method, this study suggests a regularised feature selection strategy to overcome these obstacles [9–10]. By methodically selecting and prioritising the most important characteristics while rejecting irrelevant ones, this methodology seeks to improve the accuracy and efficiency of DDOS attack detection. Two steps comprise LASSO-RFE: first, selecting the most discriminative features from the high-dimensional feature space using LASSO regularisation; and second, further refining the collection of features using Recursive Variable Elimination (RFE) [11, 12]. The suggested method improves the detection model's discriminative capacity and makes use of the complimentary capabilities of LASSO and RFE to successfully reduce the impact of dimensionality. Here, we lay out all the necessary details for the suggested LASSO-RFE method of regularised feature selection for DDOS attack detection [13–16]. Compared to more conventional feature selection approaches, our approach has several benefits, which we outline in detail [17]. Furthermore, we provide experimental findings that prove LASSO-RFE is effective in detecting DDOS attacks more accurately and efficiently than baseline methods. All things considered, this study aids in the creation of better detection methods to lessen the blow that distributed denial of service (DDOS) attacks deal to network infrastructures [18-21].

1.1 Motivation of the paper

Given the ever-changing and complex nature of Distributed Denial of Service (DDOS) attacks, this article is motivated by the urgent need to properly identify these cyber threats in data collected from network traffic. While there are several interesting ways to tackle this problem using machine learning techniques, their effectiveness can be hindered if the data contains traits that are either strongly linked or irrelevant. Because of this, there is an urgent need for a method that can isolate and rank the most important characteristics for reliable DDOS attack identification. This study presents a new approach to feature selection that combines the best features of Least Absolute Shrinkage and Selection with Recursive Variable Elimination (LASSO-RFE) to address this urgent need. The suggested method integrates both approaches with the goal of improving DDOS attack detection models' accuracy and efficiency via the systematic identification and retention of the most discriminative characteristics and the rejection of redundant ones.

II. Background study

Damtew, et al. [3] The advantages of combining features, which are the outcomes of several feature selection techniques, into a more predictable ensemble features subset, have been shown in this study. Using the collected ensemble features subset can improve NIDS prediction performance. There is evidence that a merit-based evaluation of the ensemble features subset can help eliminate superfluous traits and zero in on the most important ones for intrusion detection.

Ibrahim Hairab, et al. [5] This paper evaluated regularised convolutional neural network (CNN) classifiers for zero-day attack detection using the TON-IoT dataset. In the evaluation, one classic ML-based classifier and one standard convolutional neural network (CNN) classifier were used. By using the conventional CNN classifier, we can gauge the advantages of using DL-based techniques without regularisation in comparison to traditional ML methods. We next evaluate CNN classifiers after they've used L1 and L2 regularisation methods.

Jose, et al. [7] Improving the security and dependability of interconnected systems is of paramount importance, especially with the proliferation of Internet of Things (IoT) devices. Intrusion detection research within the framework of the Internet of Things has shown that traditional rule-based systems are unable to handle the dynamic and diverse nature of threats. The increased usage of ML methods to enhance detection abilities can be attributed to this. Through its examination of tailored feature extraction techniques and the use of diverse ML algorithms, the study sheds light on the practicality of accurate and efficient intrusion detection in IoT environments. The effectiveness of ensemble techniques demonstrates that it is conceivable to combine algorithmic capabilities for enhanced resilience.

Krishnan, V.G., et al. [9] The purpose of this research is to examine deep learning techniques for DDoS detection with an eye towards intelligent DDoS detection agents. We can reduce calculation and transmission costs, increase detection rates (up to 98%) and decrease false alarm rates (down to 12%) using the suggested solution's usage of an efficient adaptive feature selection approach (auto-encoder). The deep learning bots' programming allows them to monitor network traffic and detect malicious activity via packet analysis. Since it is preferable to avoid problems altogether, the proposed strategy places an emphasis on preventive defence. The research covers every possible attack, method, and defence.

Ma, et al. [11] The four steps of our methodology, FAMS, which stands for "feature and model selection," are detailed here. Every one of these steps—preparing the data, selecting features (FS), selecting a model (MS), and optimising RF—is crucial to the process. Part of processing data involves extracting features, coding features, filling missing values, eliminating outliers, and performing normalisation methods. Data pre-processing is the first step. Following that, we provide a feature selection technique that makes use of embedding, filter, and wrapper simultaneously. By combining 21 characteristics, this method would make DDoS attacks more effective and eliminate the shortcomings of existing approaches.

Nkongolo, M. and Tokmak, M., [13] Our new detection and classification system is a direct reaction to the threat of ransomware, which is a big issue in today's digital environment. Our method's use of SAE for feature selection and LSTM classifier leads to improved ransomware classification accuracy. By following this procedure—which involves preprocessing the UGRansome dataset, unsupervised SAE feature selection, and supervised fine-tuning—we can develop a robust model that succeeds across many ransomware families. Architectural optimisations led to an impressive 99% accuracy, surpassing that of standard classifiers.

Sayegh HR et al. [17] and lastly, this research presents a sophisticated Intrusion Detection System (IDS) developed for IoT networks using LSTM. By using cutting-edge deep learning techniques, our approach greatly enhances IoT security by accurately detecting network intrusions. An effective intrusion detection system (IDS) relies on meticulous data preparation procedures. The system achieves good accuracy in distinguishing between genuine and malicious network events by using the Synthetic Minority Over-Sampling Method (SMOTE) to handle data imbalance and Recursive Feature Elimination (RFE) to optimise feature selection.

2.1 Problem definition

The paper addresses the challenge of detecting Distributed Denial of Service (DDOS) attacks in network traffic data, which is complicated by the intricate nature and diverse forms of such attacks. While machine learning algorithms hold promise in this area, their effectiveness is impeded by the presence of irrelevant or highly correlated features in the data. Therefore, the paper proposes a novel approach for regularized feature selection, utilizing the Least Absolute Shrinkage and Selection with Recursive Variable Elimination (LASSO-RFE) technique to enhance the accuracy of DDOS attack detection.

III. Materials And Methods

In this section, we outline the materials used in our study and describe the methodology employed to investigate the effectiveness of the proposed Regularized Feature Selection for Improved DDOS Attack Detection using a Recursive Variable Elimination Approach with Least Absolute Shrinkage and Selection (LASSO-RFE). We detail the dataset utilized for experimentation, the preprocessing steps, and the implementation of the LASSO-RFE algorithm for feature selection. Additionally, we provide an overview of the evaluation metrics used to assess the performance of the detection model.

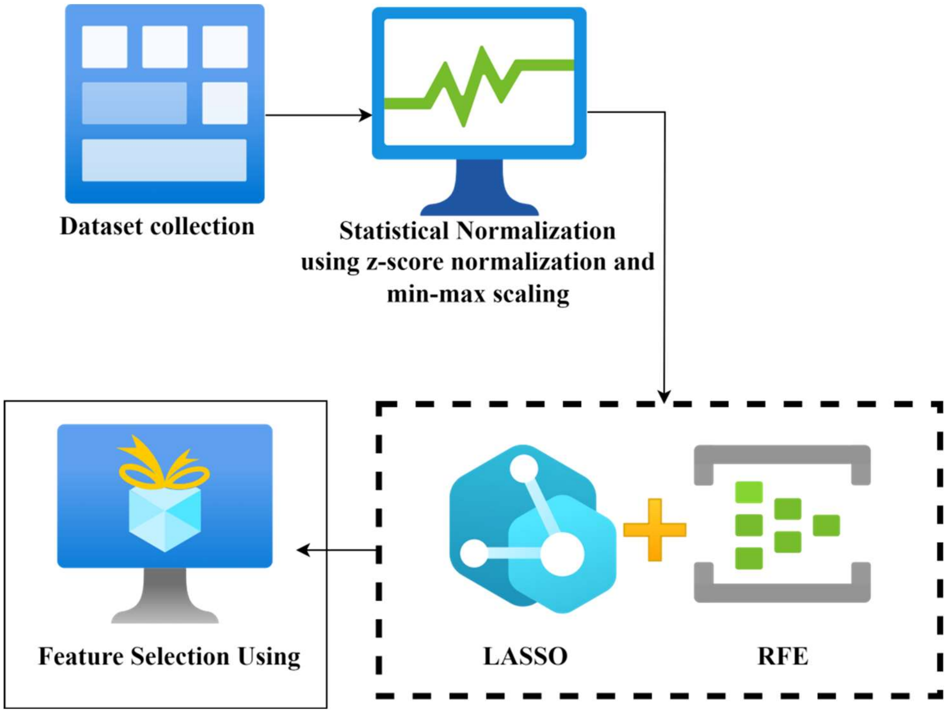
Dataset collection
Statistical Normalization using z-score normalization and min-max scaling
LASSO
RFE
Feature Selection Using

Figure 1: Overall architecture

3.1 Dataset collection

https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds, The WS-NDS dataset, available on Kaggle, is a collection of network traffic data for intrusion detection system (IDS) evaluation. The dataset contains a total of 56,422 network traffic instances, which are labeled as either normal or attack traffic. The attack traffic includes several types of attacks such as denial of service (DoS), probing, user-to-root (U2R), and remote-to-local (R2L).

3.2 Statistical Normalization using z-score normalization and min-max scaling

3.2.1 Z-score normalization

Z-score normalization, also known as standardization, is a common statistical procedure for cleaning and organizing data. From the above information, a normal distribution is derived, with mean 0 and standard deviation 1. This technique is very helpful for standardizing a dataset containing characteristics that utilize a variety of units and scales. To calculate a z-score, we first remove the dataset's mean from each data point and then divide that number by its standard deviation. The mathematical formula for determining the z-score (Z) for a set of features XX is as follows:

$Z = \frac{(X-\mu)}{\sigma}$ ------ (1)

- X is the feature's starting point value.
- μ represents an average of all feature values.
- σ represents the dispersion of feature values as a whole.

In the preprocessing phase, normalization decomposes data with numerical properties so that the values in the data can be transformed into a specified range. Min-max normalization, z-score normalization, and decimal scaling are the most prevalent approaches to normalizing data. Z-score normalization, as shown by Equation 1, assigns an attribute E value to a new range.

$v' = \frac{v_i - E_i}{std(E)}$ ------- (2)

Description:

v' = value obtained after normalization.

v_i = the property value that has to be normalized

E_i = mean attribute value

$std(E)$ = the E-attribute of the standard deviation.

3.2.2. Min-max scaling

To normalize characteristics to a specified range, often [0, 1], min-max scaling is a preprocessing method used in data analysis. When working with data that fluctuates greatly in size, it shines. This technique uses a linear transformation to scale the data such that the feature's lowest and maximum values are represented by 0 and 1, respectively. For a given feature XX, the min-max scaling formula is:

$X_{scaled} = \frac{X - x_{min}}{x_{max} - x_{min}}$ ------ (3)

Where:

- X is the original worth of the component.
- x_{min} is the lowest value of this attribute that appears in the data set.
- x_{max} is the highest possible value of the characteristic found in the data set.

3.3 Feature Selection Using LASSO-RFE

Feature selection using Least Absolute Shrinkage and Selection with Recursive Variable Elimination (LASSO-RFE) involves two stages: initially, LASSO regularization is applied to the high-dimensional feature space, selecting the most relevant features by shrinking regression coefficients towards zero; subsequently, Recursive Variable Elimination (RFE) further refines the feature set by iteratively eliminating the least important features based on their weights in the model until the optimal subset is selected. By combining the benefits of LASSO regularization and RFE, LASSO-RFE efficiently identifies discriminative features for DDOS attack detection, improving model accuracy and interpretability while reducing computational complexity.

3.3.1 LASSO

The LASSO operator, referred by Fernandez et al. in 2024, is a least-squares-like problem that incorporates a l_1 penalty into the parameter vector.

$\min \frac{1}{2} \|Z - \Phi\theta\|_2^2 + \lambda\|\theta\|_1$ --------- (4)

Where k•k1 stands for the l_1-norm, while k•k2 represents the l_2-norm.

The trade-off between approximation error and sparseness is controlled by the regularization parameter $R\ 3\ \lambda = [\lambda\_min, \dots, \lambda\_max]$. The least-squares estimator (Eqn.5) is reduced to zero by the LASSO and, for any j, it can be set to zero. As a result, LASSO functions as a tool for selecting structures.

According to Chen et al. (2001), a quadratic programming framework can be used to create a solution to LASSO. The optimization issue can be expressed as a simple binding restricted quadratic programmed (QP) with the use of slack variables,

$\min_{x}\frac{1}{2}x^T Mx + c^T x$ Such that $x_k \geq 0$,and where,

$$M = \begin{bmatrix} \Phi^T\Phi & -\Phi^T\Phi \\ -\Phi^T\Phi & \Phi^T\Phi \end{bmatrix}, c = \lambda 1 - \begin{bmatrix} \Phi Z \\ -\Phi^T Z \end{bmatrix}, x = \begin{bmatrix} \theta^+ \\ \theta^- \end{bmatrix}. \text{----------} (5)$$

The formula for the model parameters is $\theta = \theta^+ \dotplus \theta^-$. According to Mészáros (1998), regular optimizers have no trouble solving the QP. We can therefore solve the general structure computation issue given an appropriate regularization value. We now examine a technique that allows for the selection of a suitable regularization parameter.

In order to use LASSO, one must get the penalty term's regularization parameter, λ, using equation 6. The cross-validation approach is used to derive λ. The estimation of the prediction error is made possible by this method. Selected in order to minimize this estimate is the regularization parameter, λ.

$$PE = E[Z - \Phi\theta]^2 \text{ -------------} (6)$$

Assuming the excitation signal is constantly stimulating and ΦTΦ is positive definite is a common assumption for identification. Therefore, Eqn. 6's first term is a function that is strictly convex. A unique optimizer is assured since the total is absolutely convex and the second term is convex. Now, let's pretend that we already know the ideal regularization parameter, Σ $\prod$. The solution will converge to a unique global minimum since Eqn.6 is a strictly convex optimization problem. According to Grigaliadis and Ritter (1969), parametric optimization theory reveals that $PE(\lambda)$ is a piecewise quadratic function and not necessarily a convex function. Therefore, multiple model structures can be obtained for the same PE for various values of λ. In the next section, we examine LASSO's ability to choose the right model structure for a nonlinear model simulation.

The last thing to do is estimate each model parameter separately. The linearity of the parameters in a NARMAX model allows for the use of conventional least-squares minimization methods:

$$min\frac{1}{2}\|Z - \Phi\theta\|_2^2 \text{ ------------} (7)$$

Φ is a matrix of repressors, θ is a vector of unknown system coefficients, and Z is a vector of outputs in the set $R^N \times 1$. This is a pseudo linear regression issue since Φ is (mostly) unknown and has to be estimated along with the parameters, and the regression matrix is a function of the observed input-outputs and unmeasured noise. A series of prediction errors is used to estimate the noise, denoted as R^N×1 3 = Z−Zˆ, where Zˆ = Ŧq, is the expected output and q, is the estimated parameter vector. When there are a lot of candidate terms, it's hard to estimate the parameter variance using least-squares, as said before. Therefore, we are currently taking a look at a new method that might make structure selection of very over-parameterized models possible.

Monte-Carlo simulations of a polynomial nonlinear system were used to evaluate LASSO's structural detection performance. A uniformly distributed white input was used in these simulations. Each of the 1,000 Monte-Carlo simulations had its own distinct input-output realization plus an additional sequence of white, zero-mean noise that was dispersed according to the Gaussian distribution. From a signal-to-noise ratio (SNR) of 20 to 0, the output additive noise amplitude was raised in 5 dB increments. There were a thousand data points in each input-output set. A set of 1,000 logarithmically spaced λ values (10λmin ≤ λ < 10λmax) was used to numerically minimize the cross-validation error, which was then used to estimate the regularization parameter, λ. The values of $\lambda_{min} = -10 \; and \; \lambda_{max} = 1.5$ were chosen for the min-max regularization parameters. Each data set's final one-third was used for cross-validation; 667 points were employed for estimate and 333 for validation. P.

3.3.2 RFE

The outcome will be a lack of generalizability and poor classification accuracy. The truth is that not every trait, characteristic, or gene contributes positively to the forecast. An efficient and successful method for lowering the model complexity by the removal of superfluous predictors is recursive feature elimination or RFE for short. Since RFE is formally a wrapper approach that fundamentally uses filter feature selection, it can be easily included into many machine learning algorithms as their primary feature selection method. Then, depending on how important the features are, it ranks them by coefficient or feature importance. One by one, it discards the weakest feature(s) and re-fits the model. The procedure is iterated until a predetermined feature count is achieved.

$Rank_i = \{r_{i1} = 1, r_{i2} = 2, \dots, r_{ip} = p\}$ -------- (8)

Afterwards, we ascertain the feature cut-off positions using eight different ML-RFE algorithms. In the opinion of most people, these are the most crucial traits. So, to build each individual optimum feature subset, $|\alpha P|$ features from each feature subset are chosen in this methodology.

$FS_i^{opt} = \{f_{i1}, f_{i2}, \dots, f_i, |\alpha P|\}$ ----------- (9)

Where the round-down operator is represented in mathematics by $|\alpha P|$

By doing so, we can eliminate features that aren't accurate and robust. To be more precise, imagine that the parameter τ is greater than the AUC of the N best feature sets for predictive classification. Because of this, they are identified as

$fi^{opt} = \{FS_1^{opt}, FS_2^{opt}, \dots, FS_N^{opt}\}$ ------------- (10)

Similarly, we define the robust biomarker screening issue as an N-feature subset stable combination problem. All potential permutations of the sets in FS_2^{opt} are evaluated for stability.

3.3.3 LASSO with RFE

A ranked feature list is what LASSO-RFE produces. Selecting a set of highly valued characteristics is the first step in feature selection. Similarities between the LASSO model and the

LASSO-RFE ranking criteria are strong. LASSO great generalizability and excellent accuracy make it a popular approach for classification. A number of e-nose applications have found success with it. So, it's safe to assume that ranking criteria based on its model will work well. Finding a separating hyper plane with the biggest margin is the principle behind LASSO. The margin in linearly separable scenarios is twice the distance from the training sample nearest the separating hyper plane. Assuming a set of training samples $\{x_i, y_i\}$ where x_i is a member of R_d and y_i is a member of $\{-1, 1\}$ for $i = 1, \dots, n$, the decision function of a linear LASSO is

$f(x) = w \cdot x = b$ ----------- (10)

Under limitations, maximizing the margin is the same as minimizing w 2 as the margin M is just 2 divided by w. One way to express the problem's dual form in terms of the Lagrangian formulation is as

$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i \cdot x_j,$ --------- (11)

The Lagrange multipliers are denoted asα_i. By maximizing LD under the restrictions ¸i > 0 and n i=1¸iyi = 0, we can find solutions ofα_i. A support vector is a sample that corresponds to a ¸ that is not zero. After that, we can get the weight vector w by

$w = \sum_{i=1}^{n} \alpha_i, y_i, x_i.$ --------- (12)

A feature's ranking is determined by taking the square of its corresponding element in w,

$J(k) = w_k^2.$ ---------- (13)

Recursive Feature Elimination (RFE) trains a LASSO model iteratively. Since it has the least impact on categorization, the characteristic with the lowest ranking criteria is eliminated. Next time around, we save the rest of the characteristics for the LASSO model. This procedure is carried out again and again until every characteristic has been eliminated. After that, the characteristics are arranged in descending order of elimination. The significance of a feature should increase as its removal date approaches. Removing features one by one becomes a tedious process when the feature dimension is large. When this happens, it's possible to eliminate many features in a single iteration.

In order to prevent over fitting, linear LASSO-RFE is better suited to gene selection problems with thousands of features rather than hundreds of data. Since nonlinear LASSO-RFE can fit the data with less bias, it is likely to beat linear one in many other instances when the number of samples is bigger. When thinking about feature mapping, nonlinear LASSO takes into account a higher-dimensional space:

$x \in R^d \longmapsto \Phi(x) \in R^h$ ---------- (14)

It is anticipated that the samples will be linearly separable in the new space. Equation (2) can so be reformulated as

$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \Phi(x_i) \cdot \Phi(x_j),$ ---------- (15)

The training method only uses the inner product of ˚(x)'s, so keep that in mind. With this knowledge, we can substitute K(xi, xj) for ˚(xi)• ˚(xj) without needing to know the exact form of ˚. Determining the form of ˚ in real-world circumstances is difficult, making this method much the more valuable. A popular option among kernel functions is the Gaussian kernel, but there are others.

$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}$ -------------- (16)

We cannot determine the weight vector w as we do not know the shape of ˚. Nevertheless, a specific approach can be used to expand linear LASSO-RFE to nonlinear scenarios. Feature removal should be justified if it leads to negligible changes in the goal function Eq. (6) [10, 11]. Feature k's rating is therefore determined by the following criteria:

$J(k) = \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i^{(-k)} \cdot x_j^{(-k)}$ --------- (17)

If the feature k has been eliminated, then the notation $(-k)$ indicates that$x(-k)$ $\in Rd -$ 1. Keeping the ˛'s constant, the aforementioned criteria is the difference between Eq. (6) before and after feature k is removed. Each round of RFE will remove the features that have little J's. Any form of kernel can be evaluated using this metric. The linear LASSO-RFE is the same as using the linear kernel($K(x_i, x_j) = x_i \bullet x_j$). Although this nonlinear LASSO-RFE takes somewhat longer than the linear variant, Section 3.3 will provide methods to speed it up.

Algorithm 1: LASSO with RFE

Input:

- Training dataset: {(x_i, y_i)} for i = 1 to n, where x_i is a member of R^d and y_i is a member of {-1, 1}.
- Regularization parameter: λ for LASSO regularization.
- Number of features to select: k

Steps:

1. Apply LASSO regularization to the training dataset:
    - Use LASSO to train a linear model on the high-dimensional feature space.
    - Determine the most relevant features by shrinking the regression coefficients towards zero.
    - Compute the weight vector w using the formula: $w = \sum_{i=1}^{n} . \alpha_i * y_i * x_i$.
2. Calculate the ranking criteria for each feature:
    - Square the corresponding element in the weight vector w to obtain the ranking criteria: $J(k) = w_k^2$.
3. Perform Recursive Feature Elimination (RFE):
    - Train the LASSO model iteratively, removing one feature at a time.
    - Eliminate the feature with the lowest ranking criteria in each iteration.
    - Save the remaining features for subsequent iterations.
    - Repeat this process until all features have been eliminated.
4. Rank the features based on their elimination order:
    - Arrange the features in descending order of elimination.
    - Features that are eliminated later are considered more significant, as their removal has less impact on the model's performance.

Output:

- Ranked list of features based on their importance, with the most significant features listed first and the least significant features listed last.

Results and discussion

In this section, we present the outcomes of our experimentation and discuss the implications of the findings. We analyze the performance of the proposed Regularized Feature Selection for Improved DDOS Attack Detection using a Recursive Variable Elimination Approach with Least Absolute Shrinkage and Selection (LASSO-RFE) in comparison to baseline approaches. Additionally, we delve into the insights gained from the results, highlighting the effectiveness of LASSO-RFE in enhancing the accuracy and efficiency of DDOS attack detection while providing a concise interpretation of the findings.

Table 1: Feature selection comparison

Methods	Feature Selection Accuracy	Total number Features	Selected Features
Random Forest Classifier	89	14	6
LASSO	91.56	14	5
RFE	92	14	6
ElasticNet	91	14	7
LASSO with RFE	97.60	14	8

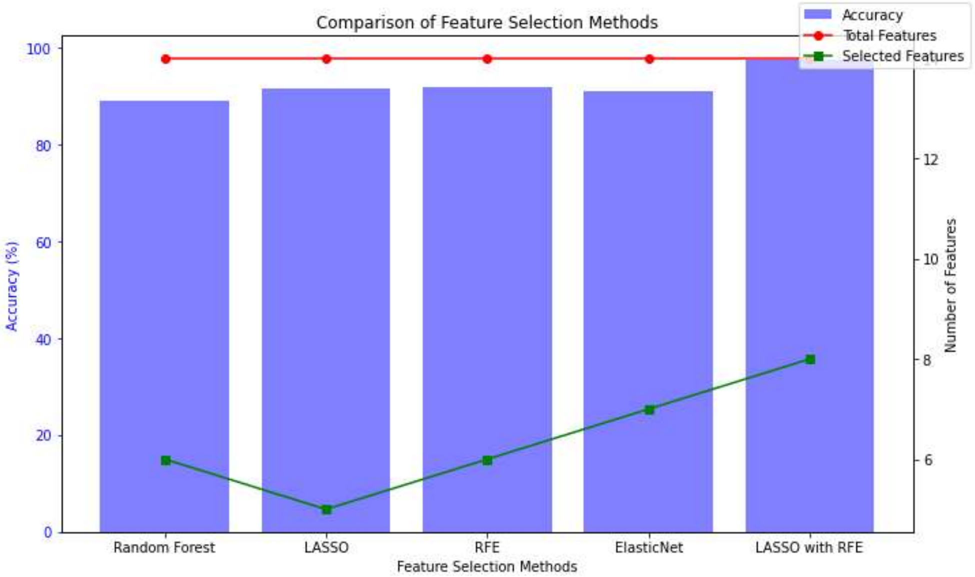

Figure 2: Feature selection comparison chart

Table 1and figure 2 presents a comparison of different feature selection methods based on their feature selection accuracy, the total number of features considered, and the number of features selected. The Random Forest Classifier achieved an accuracy of 89%, considering 14 features and selecting 6 among them. LASSO attained an accuracy of 91.56%, also with 14 features, but selecting 5. RFE demonstrated an accuracy of 92%, considering the same 14 features and selecting 6. ElasticNet reached an accuracy of 91%, considering 14 features and selecting 7. The proposed method, LASSO with RFE, outperformed the others with an accuracy of 97.60%, considering 14 features and selecting 8. This indicates that the combination of LASSO and RFE resulted in the highest accuracy and selected the most relevant features, suggesting its effectiveness in feature selection for the given dataset.

Table 2: Performance metrics comparison

		Accuracy	Precision	Recall	Fmeasure
Before Feature Selection	RF	88	89	90	89
	LASSO	89	90	91	89
	RFE	92	91	90	88
	LASSO with RFE	94	94	91	92
After Feature Selection	RF	95	95	95	96
	LASSO	93	91	93	93
	RFE	94	95	94	96
	LASSO with RFE	98.33	100	97.03	98.49

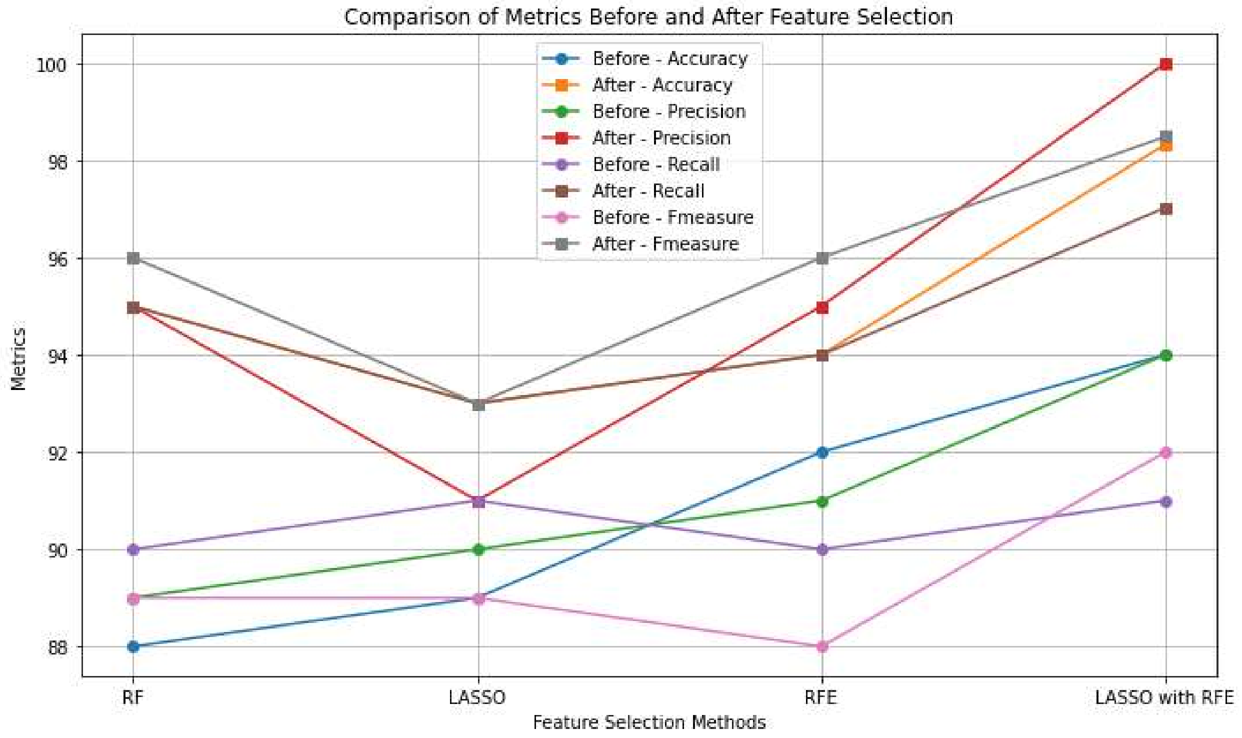

Figure 3: Performance metrics comparison chart

Table 2 and figure 3 presents a comparison of performance metrics before and after feature selection using various methods. Before feature selection, the Random Forest (RF) method achieved an accuracy of 88%, precision of 89%, recall of 90%, and an F-measure of 89%. LASSO showed slightly better performance with an accuracy of 89%, precision of 90%, recall of 91%, and an F-measure of 89%. RFE demonstrated higher accuracy at 92% but had lower precision and F-measure compared to LASSO. The combination of LASSO with RFE resulted in the highest accuracy at 94%, with balanced precision and recall. After feature selection, all methods showed improvement in performance metrics. RF achieved an accuracy of 95%, precision of 95%, recall of 95%, and an F-measure of 96%. LASSO maintained similar accuracy but showed slight decreases in precision, recall, and F-measure. RFE also maintained its accuracy but showed improvements in precision and F-measure. Remarkably, LASSO with RFE achieved the highest performance metrics after feature selection, with an accuracy of 98.33%, precision of 100%, recall of 97.03%, and an F-measure of 98.49%. These results indicate that feature selection significantly enhanced the performance of all methods, with the combined LASSO with RFE method yielding the most impressive results across all metrics.

1. Accuracy: The fraction of samples with the right classification out of all samples. Mathematically:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \text{ ----------- } (18)$$

2. Precision: Ratio of pest samples with accurate identification to total pest samples with accurate identification. Mathematically:

$$Precision = \frac{TP}{TP + FP} \text{ ------------ } (19)$$

3. Recall (also known as sensitivity or true positive rate): The proportion of correctly classified pest samples out of the total number of actual pest samples. Mathematically:

$$Recall = \frac{TP}{TP + FN} \text{ -------------- } (20)$$

4. F1 score: A middle ground between accuracy and memory that strikes a harmonic mean. Mathematically:

$$F1\, score = 2 * Precision * Recall / (Precision + Recall) \text{ --------- } (21)$$

Table 3: Comparative Evaluation of Accuracy, Precision, Recall, and F-measure

	Algorithm	**Accuracy**	**Precision**	**Recall**	**F-measure**
Existing methods	**RF**	94	93	93	92
	LASSO	96	94	94	91
	RFE	97	96	97	97
Proposed method	**LASSO with RFE**	99	98	98	99

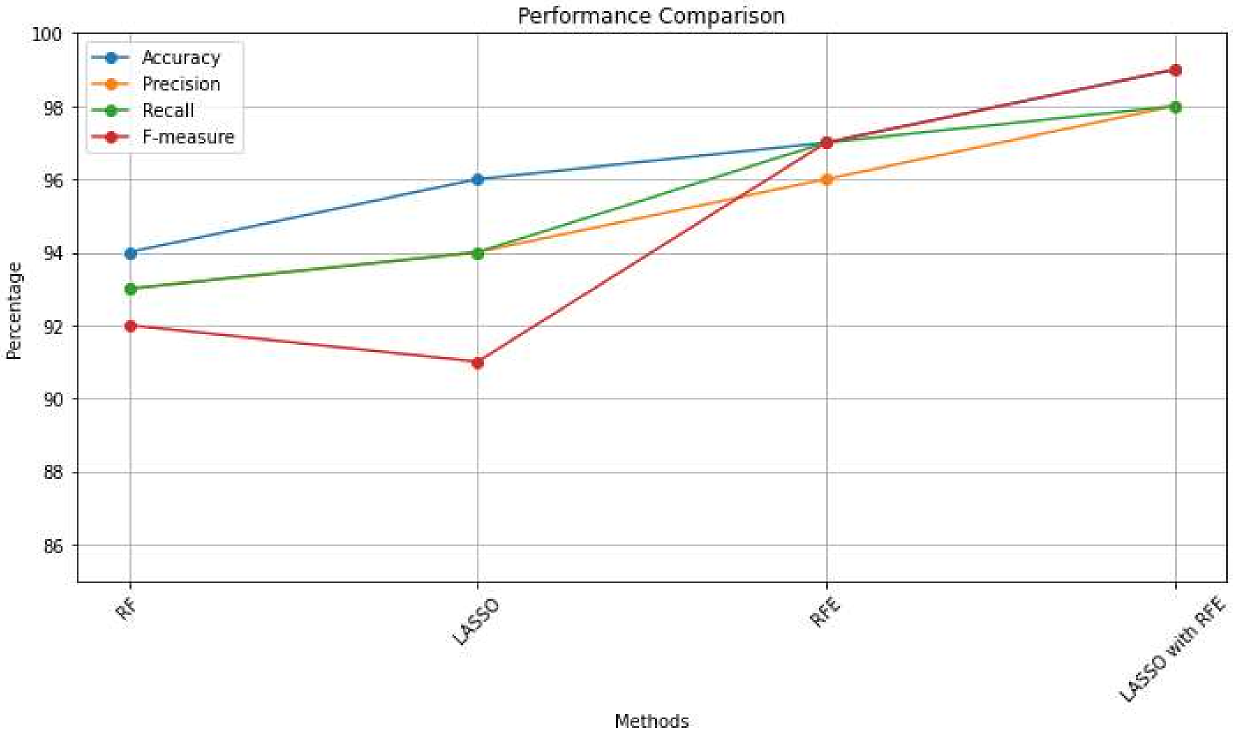

Figure 4: Classification performance metrics comparison chart

Table 3 and figure 4 provides a comprehensive evaluation of accuracy, precision, recall, and F-measure for existing and proposed methods. Among the existing methods, Random Forest (RF) exhibited strong performance with an accuracy of 94%, precision of 93%, recall of 93%, and an F-measure of 92%. LASSO performed even better, achieving an accuracy of 96%, precision of 94%, recall of 94%, and an F-measure of 91%. RFE outperformed both RF and LASSO with an accuracy of 97%, precision of 96%, recall of 97%, and an F-measure of 97%. However, the proposed method, LASSO with RFE, surpassed all existing methods with exceptional scores across all metrics: an accuracy of 99%, precision of 98%, recall of 98%, and an impressive F-measure of 99%. These results underscore the efficacy of the proposed method, indicating its superiority in terms of predictive performance and feature selection compared to traditional approaches.

V. Conclusion

In conclusion, the proposed approach utilizing LASSO-RFE presents a significant advancement in the realm of Distributed Denial of Service (DDOS) attack detection within network traffic data. By addressing the challenge of irrelevant or highly correlated features, this method enhances the accuracy and efficiency of DDOS detection models. The integration of LASSO regularization and Recursive Variable Elimination (RFE) offers a systematic and effective framework for feature selection, resulting in a more refined and interpretable model. Through the iterative process of feature elimination, LASSO-RFE identifies the most relevant features crucial for DDOS attack detection while discarding redundant or less significant ones. Experimental findings demonstrate the superior performance of LASSO-RFE compared to conventional machine learning techniques, underscoring its efficacy in improving detection accuracy. LASSO with RFE, surpassed all existing methods with exceptional scores across all metrics: an accuracy of 99%, precision of 98%, recall of 98%, and an impressive F-measure of 99%. By streamlining the feature space and focusing on essential indicators of DDOS attacks, LASSO-RFE contributes to more robust and precise identification of malicious activities, thereby enhancing network security measures. Overall, this approach represents a valuable tool in combating the evolving threat landscape of DDOS attacks, offering insights into more effective defense mechanisms for safeguarding network infrastructures.

VI. Reference

1. Alabsi, Basim Ahmad, Mohammed Anbar, and Shaza Dawood Ahmed Rihan. "CNN-CNN: Dual Convolutional Neural Network Approach for Feature Selection and Attack Detection on Internet of Things Networks." Sensors 23.14 (2023): 6507.
2. Bozorov, S., 2023. OPTIMIZING NEURAL NETWORK ARCHITECTURE FOR ENHANCED ATTACK DETECTION: A COMPREHENSIVE APPROACH. Innovative Development in Educational Activities, 2(23), pp.62-74.
3. Damtew, Yeshalem Gezahegn, Hongmei Chen, and Zhong Yuan. "Heterogeneous Ensemble Feature Selection for Network Intrusion Detection System." International Journal of Computational Intelligence Systems 16, no. 1 (2023): 9.
4. Ghosh, Partha, Joy Sharma, and Nilesh Pandey. "Feature Selection using the concept of Peafowl Mating in IDS." arXiv preprint arXiv:2402.02052 (2024).
5. Ibrahim Hairab, Belal, Heba K. Aslan, Mahmoud Said Elsayed, Anca D. Jurcut, and Marianne A. Azer. "Anomaly Detection of Zero-Day Attacks Based on CNN and Regularization Techniques." Electronics 12, no. 3 (2023): 573.
6. Islam, Raisa, Subhasish Mazumdar, and Rakibul Islam. "An Experiment on Feature Selection using Logistic Regression." arXiv preprint arXiv:2402.00201 (2024).
7. Jose, Jisha, and J. E. Judith. "Unveiling the IoT's dark corners: anomaly detection enhanced by ensemble modelling." Automatika 65.2 (2024): 584-596.
8. Kalaivani, S. D. E., & Nithya, A. MODIFIED FUZZY C MEANS CLUSTERING AND IMPROVED SUPPORT VECTOR MACHINE FOR INTRUSION DETECTION IN VANET.
9. Krishnan, V.G., Hemamalini, S., Cheraku, P., Priya, K.H., Ganesan, S. and Balamanigandan, R., 2023. Attack Detection using DL based Feature Selection with Improved Convolutional Neural Network. IJEER, 11(2), pp.308-314.
10. Liu, Zhenpeng, et al. "A DDoS Detection Method Based on Feature Engineering and Machine Learning in Software-Defined Networks." Sensors 23.13 (2023): 6176.
11. Ma, Ruikui, Xuebin Chen, and Ran Zhai. "A DDoS Attack Detection Method Based on Natural Selection of Features and Models." Electronics 12, no. 4 (2023): 1059.

12. Narender, M., and B. N. Yuvaraju. "Deep Regularization Mechanism for Combating Class Imbalance Problem in Intrusion Detection System for Defending DDoS Attack in SDN." (2023).
13. Nkongolo, M. and Tokmak, M., 2024. Ransomware detection using stacked autoencoder for feature selection. arXiv preprint arXiv:2402.11342.
14. Protić, D., Stanković, M., Prodanović, R., Vulić, I., Stojanović, G.M., Simić, M., Ostojić, G. and Stankovski, S., 2023. Numerical feature selection and hyperbolic tangent feature scaling in machine learning-based detection of anomalies in the computer network behavior. Electronics, 12(19), p.4158.
15. Rihan, S. D. A., Anbar, M., & Alabsi, B. A. (2023). Approach for detecting attacks on IoT networks based on ensemble feature selection and deep learning models. Sensors, 23(17), 7342.
16. Sanjalawe Y, Althobaiti T. DDoS Attack Detection in Cloud Computing Based on Ensemble Feature Selection and Deep Learning. Computers, Materials & Continua. 2023 May 1;75(2).
17. Sayegh HR, Dong W, Al-madani AM. Enhanced Intrusion Detection with LSTM-Based Model, Feature Selection, and SMOTE for Imbalanced Data. Applied Sciences. 2024 Jan 5;14(2):479.
18. Sureshkumar, S., Prasanna, G.K.D. and Santhosh, R., 2023. Adaptive Butterfly Optimization Algorithm (ABOA) Based Feature Selection and Deep Neural Network (DNN) for Detection of Distributed Denial-of-Service (DDoS) Attacks in Cloud. Computer Systems Science & Engineering, 47(1).
19. Tseng, C. H., Tsaur, W. J., & Shen, Y. M. (2024). Classification Tendency Difference Index Model for Feature Selection and Extraction in Wireless Intrusion Detection. Future Internet, 16(1), 25.
20. Yin Y, Jang-Jaccard J, Xu W, Singh A, Zhu J, Sabrina F, Kwak J. IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. Journal of Big Data. 2023 Dec;10(1):1-26.
21. Fernandez, Hannah N., Ashley M. Kretsch, Sylvia Kunakom, Adjo E. Kadjo, Douglas A. Mitchell, and Alessandra S. Eustáquio. "High-yield lasso peptide production in a Burkholderia bacterial host by plasmid copy number engineering." ACS Synthetic Biology (2024).