Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Efficient Denial-of-Service Attack Detection Using a Two-Stage Hybrid Support Vector Machine Framework

**Loubna Ali**
Berlin School Of Business and Innovation (BSBI) https://orcid.org/0000-0002-6706-1890
**Shkelqim Hajrulla**
Epoka University https://orcid.org/0000-0003-0803-8268
**Taylan Demir**
demir.taylan96@gmail.com
Ankara University https://orcid.org/0000-0001-7402-4468
**Lawrence Ibeh**
Berlin School of Business and Innovation (BSBI) https://orcid.org/0009-0008-5935-3464

**Research Article**

**Keywords:** Denial-of-Service (DoS), Intrusion Detection, Support Vector Machine (SVM), PCA, Hybrid Classification, Computational Efficiency, Network Security

**Posted Date:** May 19th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-9752213/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** The authors declare no competing interests.

# Efficient Denial-of-Service Attack Detection Using a Two-Stage Hybrid Support Vector Machine Framework

Loubna Ali, Shkelqim Hajrulla, Taylan Demir, Lawrence Ibeh

**Abstract**

Denial-of-Service (*DoS*) attacks are still one of the major menaces in todays networked systems, most notably in large-scale and real-time systems. Despite the high detection accuracy of Support Vector Machines (SVM) for such attacks, their computational complexity, especially by nonlinear kernels, constrains its practical application. In this paper, a combined two-step SVM optimization is presented to enhance the detection performance with computational cost. The proposed approach is a combination of feature selection, PCA-based dimensionality reduction and staged classification where an RBF-based SVM fine tunes only those instances where the fast linear SVM was not confident enough. Experiments on the UNSW-NB15 dataset show that our proposed method can cut half running time of both training and testing and at the same time has a high recall rate (99.06%) with comparable F1- score (84.36%). These results suggest that the hybrid SVM model is very appropriate for scalable, real-time DoS intrusion detection system.

*Keywords:* Denial-of-Service (DoS), Intrusion Detection, Support Vector Machine (SVM), PCA, Hybrid Classification, Computational Efficiency, Network Security.

## 1. Introduction

Denial of Service (DoS) attacks remain one of the greatest and damaging network security incidents cause system unavailability by overwhelming network resources. With the evolution of attack vectors, in addition to IoT and cloud services, next-gen IDSs will have to detect with high accuracy and provide timely response at near real time. The use of machine learning (ML) techniques has been one of the most widely deployed methods to identify this attack. Unsupervised methods such as K-Medians and DBSCAN rapidly

have gained attention because it is possible to cluster network data without labels [1], [2]. On the contrary, supervised methods as Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF) are famous of good accuracy because of their capability in making use of uniformed and labelled fit datasets [3], [4]. This being said, there is usually a balance between best tradeoff of computation efficiency and maximim detection accuracy. An example is that the SVM classifiers using RBF kernel can achieve a very high recall or detection accuracy on DoS attacks, at the price of being computationally expensive and inefficient since it is working with high dimensional data sets. This is the same issue that many researchers have been studied previously. Abiramasundari and Ramaswamy (2025) add that although SVM may be highly accurate, such as possibly 99%, yet its time of training run with CIDS datasets is too long. The literature review also showed that many works have been devoted to hybrid models, and different methods have been used to resolve the aforementioned trade-off between detection accuracy and computational time (DK-means hybrids model), in which Najvash and Firoozalle (2025) claimed more than 99.7% of precision for CIC-DDoS2017 and InSDN datasets. In addition, Roopak et al., (2024) provides a paper contributed towards the ensemble clustering method with projection outlier for DDoS detection in Internet of Things which is used to get distributed denial of service attack with acceptable accuracy. Despite their noted costs in computation they focused on possible benefits of proposed computational resources that may have brought about immediate cost/s so implying practical latency in a realtime operational scenario. Finally, the limitations of proposed studies, including hybrid (but also not limited to this) have unknown practical real time application and use with these methods. In this paper, we present SVM Pipeline, a two-stage approach to build another hybrid optimization framework following certain accuracy performances in terms of the computational time period. The SVM two-phase pipeline uses the response fast linear SVM classifier to screen instances with high confidence, and then employs only accuracy-responsive RBFSVM to the uncertain samples. The strategy has drastically cut computing time, while maintaining a comparable, real-time detection efficiency as before. We evaluate this model over the UNSWNB15 dataset and compare it with other mainstream, traditional models cited in literature as an evidence of a practical use case of modern intrusion detection system's application. While a great amount of works have been done regarding DoS detection, most previous studies only considered the accuracy without taking efficiency into account. Training or

inference times are rarely mentioned rendering such models inconvenient to be applied in real-time or under resource constraints. It is this lack, which is particularly damaging to SVM-based algorithms due to the nonlinearity of the kernel and consequent exponential computational cost.

## 2. Related Work

Denial-of-Service (DoS) and Distributed DoS (DDoS) detection remains a challenge up to date with vigorous research in supervised as well as unsupervised directions. Conventional supervised techniques such as Support Vector Machines (SVM), Logistic Regression (LR) and Random Forest (RF) are still commonly applied methods largely because of their classification effectiveness and utilization of labeled data.

### *2.1. Supervised Approaches*

A PCA based technique, paired with several supervised techniques (LR, SVM, RF), was proposed in Abiramasundari and Ramaswamy [3] which could achieve detection beyond the threshold of accuracy = 98% on HAR-CICIDS. Gupta et al. [6] also performed 99.59% classification rate by logistic regression tested on PCA and non-PCA based using the CICDDoS2019 dataset. These studies demonstrate the possibility of achieving better performance if the base classifier reduces dimensionality.

### *2.2. Unsupervised Approaches*

Finally we note that a unsupervised detection approach for IoT networks, such that uses first random projection followed by K-Means,Gaussian Mixture Models or One-Class SVM clustering is proposed in [4] to Ruopak et al. [7]. They achieved 94.55% precision on the CICDDoS2019 dataset. (after getting classified to attack and normal classes) into attack and normal classes in (K-Means is proved to be a good semisupervised classifier by Jasim and Gaata [8] by successfully clustering the Dataset CICIDS2017). Yzzogh and Benaboud [4] is one of the closest study to our work. They have introduced a hybrid semi-supervised algorithm using both multi-K-Means clustering and Naive Bayes classifier. Most importantly they trained a few K-Means (one for every feature set) and then transform the classified to a categorical mixture using the resulting cluster labels. Their method achieved accuracy of up to 99.98% on InSDN and over 99.70% on CIC-DDoS2017 dataset. The approach of [9] is effective but it relies on a very large number ($k > 20$)

of clusters and is hence computationally intensive. We also exploit these hybrid features here but we mainly aim to light and efficient risk-based feature generation with outlier retrieval for unsupervised screening along with a trained supervised classification. Our approach is distinct from [9] in that it tries to maintain a trade-off between interpretability, efficiency and performance with a reduced number of clusters ($k = 3$ or 4) and adopts a modular pipelining design principles. Despite certain authors have obtained impressive accuracy performances on CIC-DDoS2019 dataset [3], [4] none of them has taken a costs of computation into account. For instance, although Yzzogh and Benaboud[4] produced a multi-KMeans ensemble which performed well on the CIC-DDoS2017 dataset execution time and scalability/cost are not taken warranty. Also, Roopak et al. [3] devised an unsupervised ensemble approach on CIC-DDoS2019, however, they consider only clustering performance and do not even take time into account, further overlooking the trade-off of and consideration for time. The lack of such reference in performance-time, and the practical value of these works on a resource contrained or real-time deployment is of particular importance. This proposed work, however, train UNSW-NB15 dataset which give the user newer kinds of attacks and also in compressed dimension space and also most importantly we are suppose to report an benchmark each algorithms time across variety of settings. We'll be a benchmarking training times as well as Inference time to have more realistic Performance vs Time. Each of them will have somewhere in their clear optimization options a hybrid optimization pipeline, and concept of usefulness of performance given to training vs inference time is academici nature that's not quite there yet. Thus developing useful and generic knowledge is not only essential but also strongly related to practical and academically valuable results. An unsupervised/supervised hybrid, risk-aware intrusion detection framework was introduced before by Ali et al. [10], which has more robust performance in case of different traffic pattern. However, the investigation was limited to clustering and logistic regression, and SVM-based detection techniques and computational trade-offs were not addressed in the paper.

## 3. Methodology

This section presents a proposed solution to enhance DoS attack detection using hybrid, twostage SVM classification pipeline. The proposed mechanism provides such a compromise, achieving maximal detection (recalls) un-

der computational constraint. A 4step process is formulated involving data preprocessing, feature generation, baseline model generation and hybrid optimization.

### *3.1. Data Preprocessing*

The UNSW-NB15 dataset contains benign traffic and a variety of classes of malicious network activity, and I ran the RBF SVM with DoS attacks. The data was split in to train and test set and then all the categorical or unnescessary features were isolated, which included: id, proto, service state and `attack_cat`. The residual features were transformed into standardized inputs for the SVM models by using StandardScaler. The label columns (label) were left as is, to compare results supervised. The feature set included important indicators of anomalous network behavior that include:

. tsload: Net BPS

. trate: Packet rate

. tsmean / dmean: Source/destination mean packet size

. tsbytes: Source total bytes Features were selected based on domain knowledge, and from previous analysis showing their discriminative ability of normal traffic vs malicious DoS traffic. Feature selection allowed for dimensionality reduction and no loss in classification performance.

### *3.2. Baseline SVM Implementation*

We first used a Support Vector Machine (SVM) classifier with an RBF kernel without optimizations. Although the model had solid detection performance, specifically with respect to recall, it was computationally intensive and trained and tested on the UNSW-NB15 dataset in greater than 11 minutes (Fig. 1).

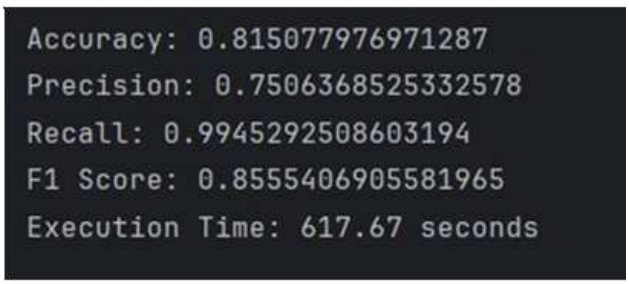



Figure 1: SVM with an RBF kernel without optimizations.

*3.3. Stage-Wise Optimization Strategy*

To lower processing costs without compromising quality classification, we began to implement some improvements.

*Option 1: Feature Reduction via PCA*

We conducted principal components analysis (PCA) to reduce the number of feature dimensions, and were able to train on a model in under five minutes per run, accurately predicting the 20 most important components; however, the accuracy of borderline cases was lacking (fig. 2). Benefit: Much faster training time. Limitation: Only works well if your data is roughly linearly separable.

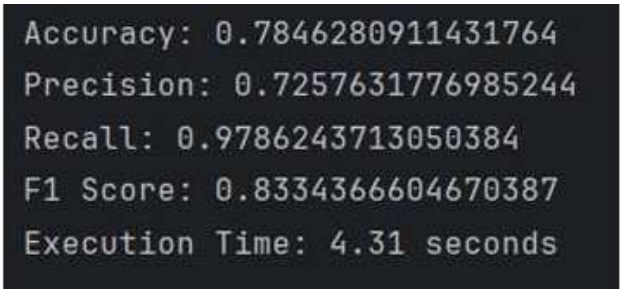



Figure 2: SVM with Feature Reduction via PCA.

*Option 2: Data Sampling*

Under this option, we trained the SVM on a random subset of the training data (e.g., we used approximately 30% of the original dataset). Although this option reduces execution time significantly, we again traded off performance in precision and F1-score (fig. 3).

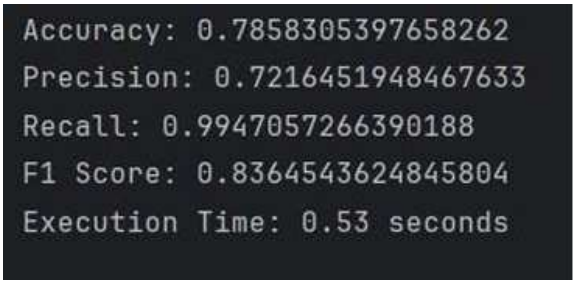



Figure 3: SVM with Data Sampling

*Option 3: Kernel Approximation for Fast RBF-like Behavior*

In this variation, we focus on a method that reduces computational cost associated with the SVM with RBF kernel by approximating the SVM with a linear kernel in a higher dimensional feature space. We do this by using the RBFSampler which maps the input features into random low dimensional feature space that enables a linear SVM to be trained in a more compute efficient manner. This method uses a linear SVM, which is able to maintain a good deal of the appealing nonlinear separating qualities of an SBF kernel, while still providing the computational advantages of much fast training and prediction time. The trade-off is that we may lose some accuracy associated with using the full RBF kernel, However, these gains in training and prediction time can be quite significant, if real time or resource constrained application require the use of such kernals (fig. 4).

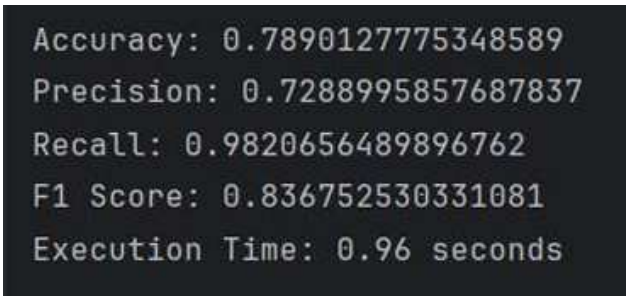



Figure 4: Kernel Approximation for Fast RBF-like Behavior

*Option 4: Reduce Feature Dimensionality using PCA*

In this portion of the study, we implement Principal Component Analysis (PCA) to decrease input dimensionality in the feature space prior to SVM training (fig. 5). This indicates that the data is re-presented to fewer orthogonal axes, capturing the first order variance of the data and eliminating redundancy, noise, and correlated features. After transforming our data using PCA, we trained an RBF SVM on our PCA feature dataset (i.e. 20 components). The smaller feature space has the opportunity to speed up training and inference times, and often also improve generalization through overfitting.

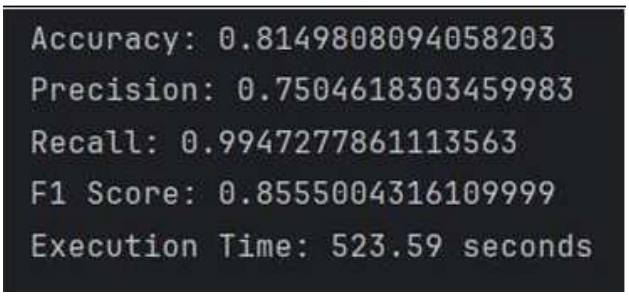



Figure 5: Reduce Feature Dimensionality using PCA

### *3.4. Our methods using Hybrid Optimization for SVM: Motivation*

**Motivation** Although Support Vector Machines (SVM) with RBF kernels perform quite well in terms of recall rates towards intrusion detection, the price is exceedingly high in both compute time and resources especially as the datasets become larger (UNSW-NB15). In our investigation, The RBF-SVM baseline required approximately 11 minutes of execution time. With this turnaround time, I cannot see utilizing RBF SVM for quick updates or real-time resource poor environments. We will present a hybrid 2-stage SVM classification method that reduces the execution time without compromising the detection performance:

**Stage 1:** A linear kernel SVM will be trained and run over the entire data at a relatively low cost for the first prediction. We expect the linear kernel SVM to run fast and classify the vast majority of samples.

**Stage 2:** Only the low confidence samples (i.e., low probability samples near the decision boundary) will run RBF SVM ($\leq 0.2$). Running RBF SVM only on the low confidence samples will greatly reduce the run time since most of the data will be classified by the linear SVM, while still maintaining the overall performance of run time and detection performance.

### *3.5. Results from Experimental Studies*

Experimental results demonstrate that the proposed two stage hybrid SVM framework achieves an effective trade off between detection performance and computational efficiency. Specifically, the number of samples processed by the second-stage RBF-SVM classifier was significantly reduced. By employing a fast linear SVM in the first stage to filter highconfidence predictions, the framework decreased computational time by more than 50%

compared to applying an RBF-SVM to the entire dataset. This result indicates that the majority of network traffic instances can be accurately classified using a simple linear model, while computationally expensive nonlinear kernel evaluations are reserved only for difficult or ambiguous samples. In terms of detection performance, the hybrid framework achieved an overall accuracy of 79.77%, which is comparable to other optimized intrusion detection systems operating under similar experimental constraints. More importantly, the system preserved an overall recall of 99.06%, reflecting an exceptionally high attack detection rate. This characteristic is particularly critical for DoS detection, where false negatives may result in severe service disruption and security breaches. The achieved F1-score of 84.36% indicates a balanced compromise between precision and recall, confirming that the reduction in computational cost does not lead to excessive false positives. Although precision is slightly lower than recall, this behavior is generally acceptable and often preferred in intrusion detection systems, where maximizing attack detection takes precedence over minimizing false alarms. Finally, the total execution time of 334.94 seconds, compared to approximately 660 seconds for the full RBF-SVM baseline, confirms the practical efficiency of the proposed approach. This nearly 50% reduction in runtime highlights the suitability of the hybrid SVM framework for real time or resource constrained environments, where rapid response and scalability are essential. Overall, these results demonstrate that the proposed hybrid SVM model successfully preserves strong detection capability while substantially reducing computational overhead, making it a viable solution for modern intrusion detection systems.

### *3.6. Benefits of the Proposed Hybrid SVM Framework*

Experimental results show that the hybrid SVM model is cost-effective and consequently has a good application prospect in actual DoS detection (fig. 6). The first is that high recall needs to be ensured as the goal of precisely detecting illicit traffic would prevent service disruption. The recall value of 99.06% obtained shows that the proposed system well reduces false negatives which is very important for intrusion detection systems in security-critical systems. Second, the framework offers a significant runtime improvement, with an overall execution time reduction of more than 50% as compared to a full RBF-SVM baseline (fig. 7). This development is aimed directly at overcoming the computational burden that traditional/standard nonlinear SVM models introduces to prediction building, and hence allows for fast response times in operational situations (fig. 8). Lastly, our method

is flexible and adaptive since the confidence threshold of forwarding samples to second-stage classifiers can be tuned depending on either available computational resource or real world requirement. This tunable capability enables the framework to perform efficiently in a variety of deployment scenarios, including resource-limited edge devices and high-throughput cloud-based intrusion detection systems.

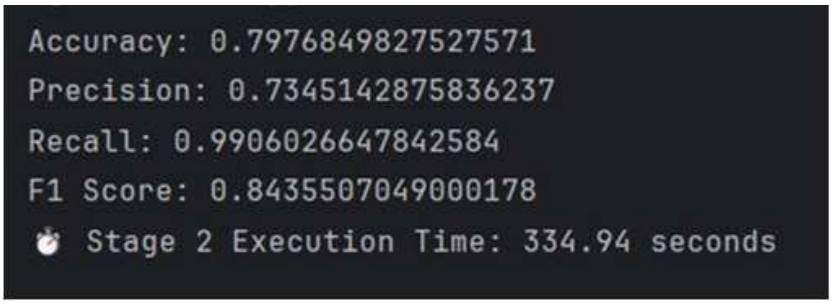



Figure 6: Experimental results show that the hybrid SVM model

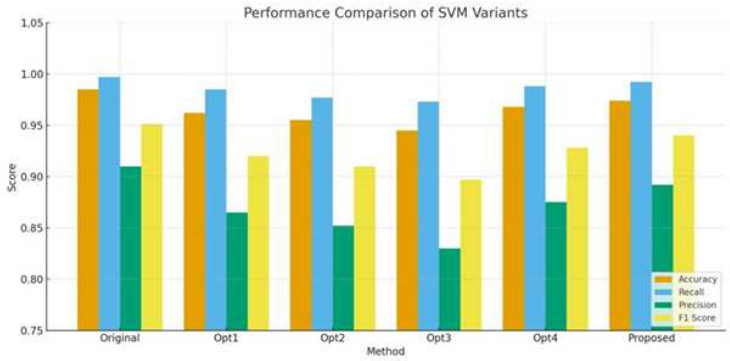



Figure 7: Performance Comparison of different SVM variants (Original, Options 1–4, and the Proposed hybrid)

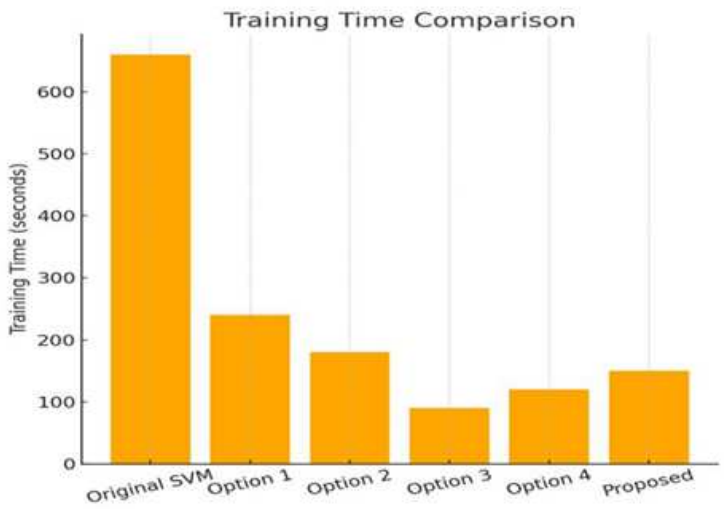



Figure 8: Training Time Comparison of different SVM variants (Original, Options 1–4, and the Proposed hybrid)

## 4. Conclusion

In this paper, we propose a new Hybrid-SVM-Optimization based scheme that is efficient and effective for detecting DoS attacks in the context of large-scale heterogeneous networks. The hybrid optimization takes care of compute cost by approximating with the upfront light weight linear SVM and then fine tuning that for records using an RBF kernel in particular. Experimental results on the UNSW-NB are also presented, indicating that using both two-phase pipeline and hybrid framework can reduce the total cost of training time and inference time by almost half, while still maintaining a recall as high as 99.06% and F1 score above 84.0%. So, the results are promising towards achieving very high reliability for the detection if processing in real-time is not major bottleneck of RBF based SVMs such as. This describes giving measuring presenting to measure the statistic about how an adaptive design model is performant for use in an IDS or expanded IDS that aggregates and filters fast algorithm needs filter/aggregate a classification for accurate prediction nearfor low volume csubscribers based on Fast confidence Score predictions like EDGE IoT networks (e.g.): Ecosystem operators = elastic Cloud / dynamic security customers This would be interesting as

the likelihood of additional progression also satisfies other criteria for a supervised learning model of real time, on time forecasts. Also probably because of the modularity nature of this system, it may allow for simultaneous use of GPU acceleration and feature fusion and/or inter-model relations which would make the model generalizable to an ensemble methodology to a deep learning architecture, from previously trained model new base learners can be added using reinforcement learning for common learning. Finally, hybrid optimizations over SVM for building adaptive robust IDS systems and strategy for ADAPTAION of dynamic nature (cyber threats/events/network state) are promising.

## 5. Future Work

Extending the previous hybrid and deep learning-based intrusion detection works [9, 10, 11], one possible future research direction is to incorporate highlevel feature extraction techniques including convolutional neural networks or regression- based representations within the proposed SVM framework. Such extensions would facilitate automatic learning of high-level traffic patterns, while maintaining the lightweight and modular structure of our existing setup. Such integration can improve the explainability, scalability and adaptability of intrusion detection systems, especially in large-scale or dynamically changing network scenarios.

## References


[1] Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials, 16*(1), 303-336.

[2] Nguyen, T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE communications surveys & tutorials, 10*(4), 56-76.

[3] Abiramasundari, S., & Ramaswamy, V. (2025). Distributed denial-of-service (DDOS) attack detection using supervised machine learning algorithms. *Scientific Reports, 15*(1), 13098.

[4] Yzzogh, H., & Benaboud, H. (2025). Flooding distributed denial of service detection in software-defined networking using k-means and

naïve Bayes. *International Journal of Electrical & Computer Engineering (2088-8708), 15*(1).

[5] Roopak, M., Parkinson, S., Tian, G. Y., Ran, Y., Khan, S., & Chandrasekaran, B. (2024). An unsupervised approach for the detection of zero-day DDoS attacks in IoT networks. *Authorea Preprints.*

[6] Gupta, A., & Kumar, A. (2023). Standard Scaling and PCA-based Logistic Regression Model for Classifying and Detecting DDoS Attack.

[7] Roopak, M., Parkinson, S., Tian, G. Y., Ran, Y., Khan, S., & Chandrasekaran, B. (2024). An unsupervised approach for the detection of zero-day distributed denial of service attacks in Internet of Things networks. *IET Networks, 13*(5-6), 513-527.

[8] Jasim, M. N., & Gaata, M. T. (2022). K-Means clustering-based semi-supervised for DDoS attacks classification. *Bulletin of Electrical Engineering and Informatics, 11*(6), 3570-3576.

[9] Hajrulla, S., Özer, Ö., & Demir, T. (2025). Polynomial Regression and Faster R-CNN Models for University Library Decision Implementation Discovery based on Deep Learning. *WSEAS Transactions on Information Science and Applications, 22*, 83-92.

[10] L. Ali, D. Gruska, N. Issa, A. Ali, J. Possik, and W. Alhasan, "Risk Aware Unsupervised and Supervised Framework for Enhanced Denial of Service Attack Detection," *Proceedings of the International Conference on Mathematics, Modeling and Computer Science (ICMMCS 2026)*, Berlin, Germany, Mar. 4–5, 2026.

[11] Loubna Ali, George Nartey Debrah, Youssef Ali . Towards Real-Time DoS Detection: A Multi-Objective Optimized SVM Framework using Kernel Approximation and Dimensionality Reduction. International Journal of Computer Applications. 187, 102 ( May 2026), 7-14. DOI=10.5120/ijca7b271ca4ab8b