iJRASET
International Journal For Research in Applied Science and Engineering Technology



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 13 **Issue:** IV **Month of publication:** April 2025

**DOI:** https://doi.org/10.22214/ijraset.2025.69784

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 13 Issue IV Apr 2025- Available at www.ijraset.com*

# Distributed Denial of Service Attack Detection Model for Peer to Peer (P2P) Networks

Madhesh S[1], Mukunthan S[2], MohanKumar R[3], JayaPrakash G S[4], Maragathavalli P[5]
*Department of Information Technology, Puducherry Technological University, India*

***Abstract:*** ***Distributed Denial of Service (DDoS) attacks are among the most prevalent and disruptive forms of cyberattacks, aiming to make a machine or network resource unavailable to its intended users. Traditional rule-based detection systems often fail to adapt to evolving attack strategies. This paper presents a machine learning-based hybrid framework for DDoS detection using Support Vector Machines (SVM), Bidirectional Long Short-Term Memory networks (BiLSTM), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The system uses NetFlow-inspired features extracted from live traffic captured in a virtualized Mininet environment. SVM is employed for supervised classification, BiLSTM for time-series based sequence learning, and DBSCAN for unsupervised anomaly detection. The results demonstrate that this hybrid approach provides robust detection accuracy, reduced false positives, and adaptability to unknown attacks.***

***Keywords: DDoS Detection, Network Security, SVM, LSTM, DBSCAN, Machine Learning, Mininet, Anomaly Detection, Flow Features, Intrusion Detection***

## I. INTRODUCTION

In the modern digital world, online services and cloud-based platforms have become integral to daily life and critical infrastructure. This widespread adoption, however, has also increased vulnerability to cyberattacks. One of the most common and damaging types is the Distributed Denial of Service (DDoS) attack, wherein a target is overwhelmed with traffic from multiple sources, exhausting resources and rendering services inaccessible.

Conventional DDoS detection techniques typically rely on predefined signatures or threshold-based rules. While effective against known threats, these methods fall short when facing novel or distributed attack patterns. Machine learning offers a promising alternative by learning traffic behavior from data and generalizing to unseen attacks.

This paper explores a hybrid machine learning approach using three models—SVM, BiLSTM, and DBSCAN. The combination of supervised and unsupervised techniques enables both classification of known attacks and detection of new anomalies. Over 80 features are extracted from packet flows and evaluated each model's performance in terms of accuracy, precision, and detection time.

## II. LITERATURE REVIEW

The detection of Distributed Denial of Service (DDoS) attacks has been extensively explored using both supervised and unsupervised machine learning techniques. Existing works focus on improving detection accuracy, handling data imbalance, and enhancing real-time adaptability within dynamic network environments.

In [2], the authors evaluated multiple supervised machine learning algorithms—logistic regression, SVM, random forest, KNN, and XGBoost—for traffic classification in Software Defined Networks (SDN). Their experiments, conducted in an SDN framework, concluded that the random forest classifier achieved the highest accuracy of 98.97% with a False Alarm Rate (FAR) of just 0.023. However, this approach is highly sensitive to parameter tuning, suffers from scalability limitations, and shows potential vulnerability to model drift when faced with newer, unseen threats.

To address the class imbalance often observed in DDoS datasets, the work in [1] introduced a hybrid model integrating DBSCAN, SMOTE, and LSTM. By applying unsupervised clustering and oversampling strategies, the proposed model reduced validation loss significantly (from 0.1934 to 0.0428) and achieved validation accuracy of 99.50%. Despite these improvements, the approach still faced challenges such as high false positive/negative rates and high computational complexity, particularly under real-time constraints.

Similarly, [5] investigated imbalance handling techniques including SMOTE, TL, OSS, NearMiss, ROS, and RUS, alongside Convolutional Neural Networks (CNN). Their results validated SMOTE's robustness, achieving over 99% in various performance metrics. However, limitations include overfitting due to oversampling, binary classification focus, and increased training time, making it less practical for real-time systems.

©IJRASET: All Rights are Reserved 

A novel architecture called ShieldRNN was proposed in [8], targeting IoT-based DDoS detection. It utilized varying sequence lengths, a seq2seq training framework, and majority voting for prediction. This method reached an F1-score of 99.919%, outperforming conventional RNN methods. Yet, the model introduced tradeoffs between sequence length and detection accuracy, increased server overhead, and scalability concerns in large IoT deployments.

The work in [11] explored the use of SVM in a Mininet-simulated SDN environment, focusing on six flow-based features. While this method achieved an average accuracy of 95.24%, it lacked support for ICMP traffic and was heavily dependent on effective feature selection and realistic traffic generation.

In contrast to the above studies, our proposed work aims to synergize the strengths of SVM, BiLSTM, and DBSCAN into a unified hybrid framework. By combining supervised, sequential, and unsupervised learning paradigms, our approach provides improved generalization to unknown attacks, efficient handling of imbalanced traffic, and real-time performance suitable for deployment in SDN and cloud environments.

## III. PROPOSED METHODOLOGY

The proposed framework integrates supervised, sequential, and unsupervised learning techniques—Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Density-Based Spatial Clustering of Applications with Noise (DBSCAN)—to form a hybrid DDoS detection system. This ensemble leverages the individual strengths of each model and improves overall robustness, especially in identifying novel and evolving attack patterns. The system is evaluated using both simulated traffic and real-world datasets, with a focus on real-time detection and alerting capabilities.Fig1.1 illustrates the End-to-End DDoS Detection Workflow

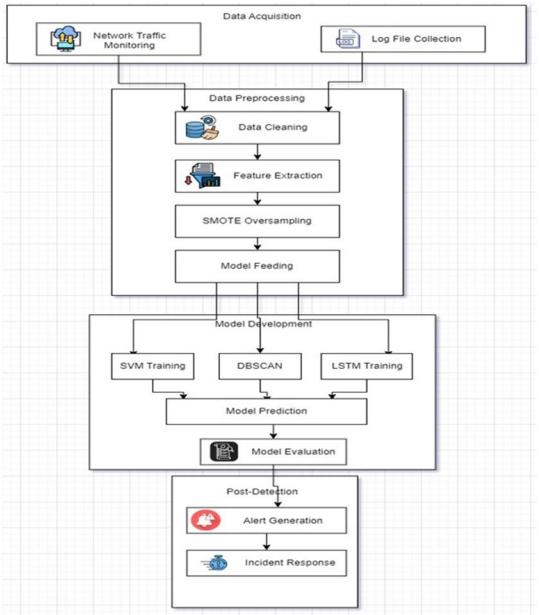
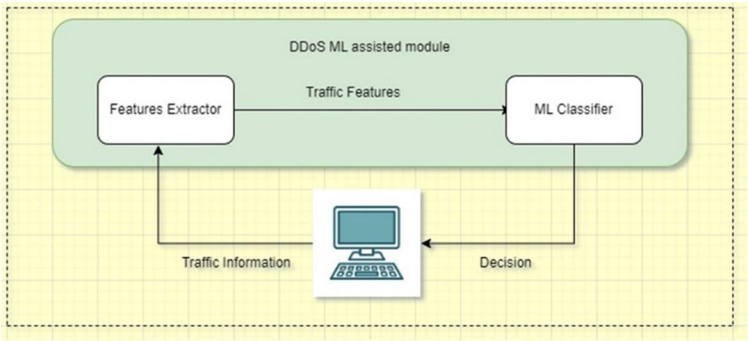



Fig 1.1 End-to-End DDoS Detection Workflow




`Fig 1.2 Individual Peer Architecture

### A. Dataset Generation and Traffic Simulation

To simulate a realistic network environment, we utilize Mininet, a lightweight virtual network emulator capable of running real-time network topologies. Traffic is generated between multiple hosts and switches, including both legitimate and malicious flows. Attack scenarios such as TCP SYN flood, UDP flood, and ICMP flood are orchestrated using tools like hping3, while normal traffic includes HTTP, DNS, and file transfer sessions.

The packet data is captured using Wireshark or tcpdump at the victim node, from which flow-level statistics are derived.

In addition to simulation, datasets such as CICDDoS2019 and a Kaggle DDoS dataset (https://www.kaggle.com/datasets/devendra416/ddos-datasets ) are used to train and validate the system. This combination ensures a mix of synthetic and real-world traffic, covering a broad spectrum of DDoS variants.

### B. Network Topology and Data Collection

A simulated environment is created using **Mininet**, with a **tree topology** consisting of **100 hosts and 5 switches**. This allows emulation of a moderately complex network suitable for both benign and malicious traffic flows. Legitimate traffic (HTTP, DNS, FTP, etc.) is generated using custom scripts and tools like iperf , while DDoS attack traffic will be introduced using tools like hping3.

### C. Feature Extraction

Packet-level data is captured at the victim host using Wireshark or tcpdump and aggregated into flows using 5-tuple identifiers. For each flow, 84 NetFlow-like features are extracted, covering:

- IP and Transport Information: Source/destination IPs, ports, and protocol
- Time-based Features: Flow duration, inter-arrival times (IATs), active/idle time
- Volume Metrics: Total packets and bytes in/out, average packet size
- TCP Flags and Control Features: Count of SYN, FIN, ACK, URG, RST
- Statistical Metrics: Mean, variance, skewness, kurtosis of intervals and sizes
- Subflow Dynamics: Packet/byte counts within subflows, header sizes

The extracted feature set is normalized using MinMaxScaler for the BiLSTM and SVM models to maintain consistent scale. For DBSCAN, dimensionality is reduced using Principal Component Analysis (PCA) to minimize noise and speed up clustering.

### D. Model Architecture

#### 1) Support Vector Machine (SVM)

SVM is trained as a binary classifier to distinguish between attack and normal flows. It employs a radial basis function (RBF) kernel, with hyperparameters optimized using grid search. The normalized 84-feature input improves decision boundary modelling in high-dimensional space. This model is effective for detecting clearly separated attack patterns but may struggle with time-dependent features.

#### 2) Bidirectional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM is employed to capture both past and future contextual dependencies within traffic flows, making it especially powerful for sequential data where attack patterns are influenced by both earlier and upcoming packets. Each input sequence is a series of 84-feature flow records processed in both forward and backward directions. The architecture includes a Bidirectional LSTM layer, followed by dropout for regularization, and a dense layer with softmax activation for binary classification. This enables the model to learn temporal relationships from both directions of the flow timeline, thus improving detection accuracy in complex and evasive attack patterns. The model is trained over multiple epochs with early stopping based on validation loss. The input features are normalized using MinMaxScaler to improve training stability and convergence speed.

#### 3) DBSCAN

DBSCAN is utilized to detect anomalies in an unsupervised manner. The PCA-reduced feature set enhances clustering performance and reduces noise. DBSCAN does not require prior labels, making it effective in identifying previously unseen (zero-day) attacks based on low-density clusters in the feature space. Parameters such as eps (neighborhood radius) and min_samples (minimum points per cluster) are fine-tuned using visual inspection and silhouette scores.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

### E. Ensemble Strategy and Decision Logic

The final detection system leverages a voting mechanism that combines the predictions of the three models. In case of disagreement, priority is given to BiLSTM for temporal context, followed by DBSCAN anomaly flagging. This strategy reduces false positives and allows early-stage detection of emerging threats.

### F. Deployment in Real-Time Environment

The complete detection system is deployed as a live monitoring agent on the victim host. Packet capture and feature extraction run continuously, with predictions performed in near real-time and results logged for administrative action.

## IV. EVALUATION METRICS

To evaluate the performance of the proposed hybrid DDoS detection system, multiple metrics are employed. These metrics provide insights not only into the accuracy of the predictions but also into the system's reliability, robustness, and ability to handle class imbalance — a common issue in DDoS detection.

### A. Accuracy

Accuracy is the most straightforward metric, measuring the ratio of correctly predicted instances (both attacks and benign flows) to the total number of instances:

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

Where:

- TP = True Positives (correctly identified DDoS attacks)
- TN = True Negatives (correctly identified benign traffic)
- FP = False Positives (benign traffic misclassified as attack)
- FN = False Negatives (attacks misclassified as benign)

### B. Precision

Precision measures the proportion of predicted DDoS attacks that were actually DDoS attacks. It reflects the model's ability to minimize false alarms.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

### C. Recall (Sensitivity)

Recall measures the model's ability to detect all actual DDoS attacks. A high recall implies fewer missed attacks (i.e., low false negatives), which is essential in cybersecurity.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

### D. F1-Score

F1-Score is the harmonic mean of precision and recall, providing a single metric that balances both false positives and false negatives. It is especially useful in imbalanced datasets where accuracy alone may be misleading.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

### E. Confusion Matrix

The confusion matrix provides a granular view of classification performance by visualizing true/false positives and negatives. It helps identify whether a model has a tendency to over-predict one class over another.

### F. ROC Curve

The Receiver Operating Characteristic Curve score reflects the model's ability to distinguish between attack and benign traffic across different threshold settings. A score closer to 1 indicates excellent separability.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

### *G. Precision – Recall Curve*

The Precision-Recall Curve evaluates a model's ability to correctly detect attacks while minimizing false alarms. Precision measures the accuracy of positive predictions, while Recall measures how many actual attacks were detected. A higher and more right-skewed PR curve indicates better performance, especially important for imbalanced datasets like DDoS detection.

## V. RESULTS AND ANALYSIS

The models were evaluated using standard classification metrics: Accuracy, Precision, Recall, and F1-Score. We also measured Detection Time per Flow to assess real-time applicability.

Table I : Model Evaluation Metrics

| Model | Accuracy | Precision | Recall | F1-Score | Detection Time (ms) |
|---|---|---|---|---|---|
| SVM | 99.89% | 99.88% | 100% | 99.94% | 0.021995 |
| BiLSTM | 98.42% | 98.40% | 100% | 99.19% | 0.249360 |
| DBSCAN | 96.95% | 96.95% | 100% | 98.45% | 0.508962 |
| Ensemble | 98.41% | 98.39% | 100% | 99.19% | 18346.541343 |

### *A. Performance Metrics*

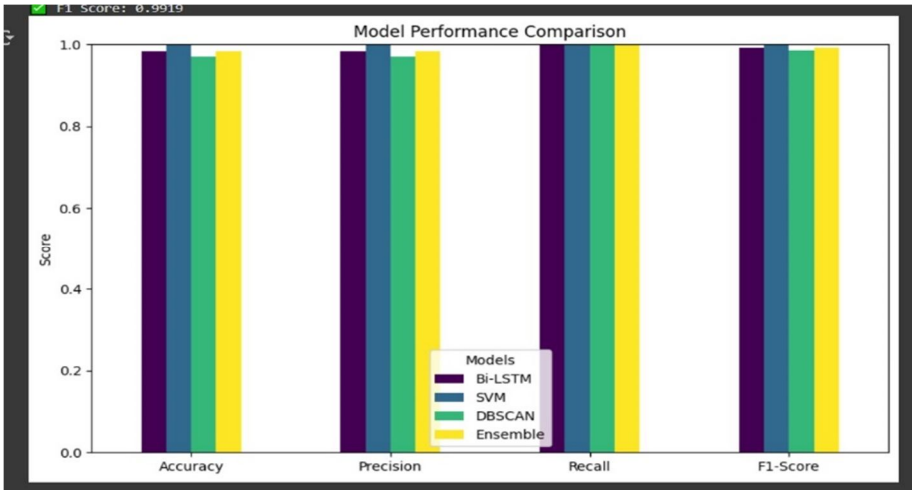
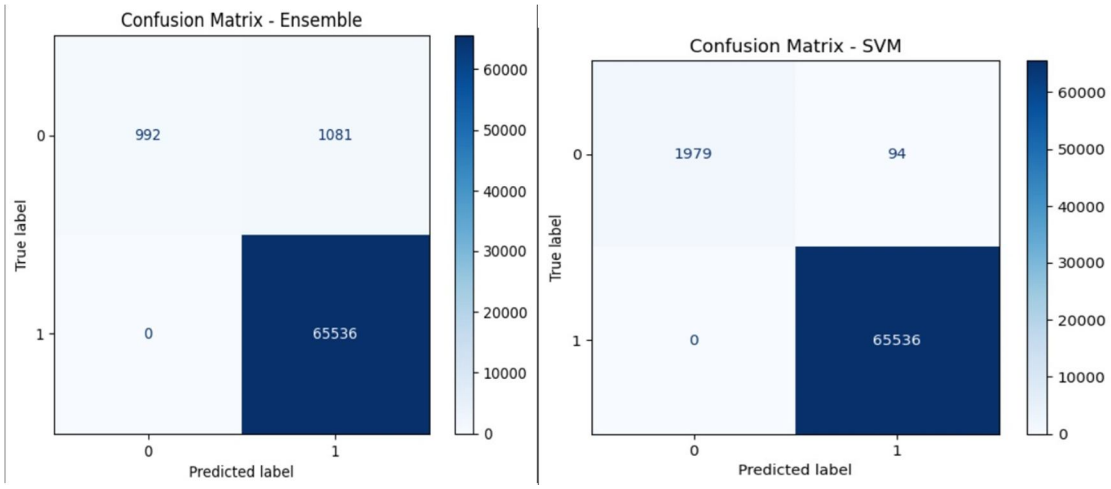



Fig 1.3 Accuracy, Precision, Recall, F1-Score for SVM, BiLSTM, DBSCAN,Ensemble



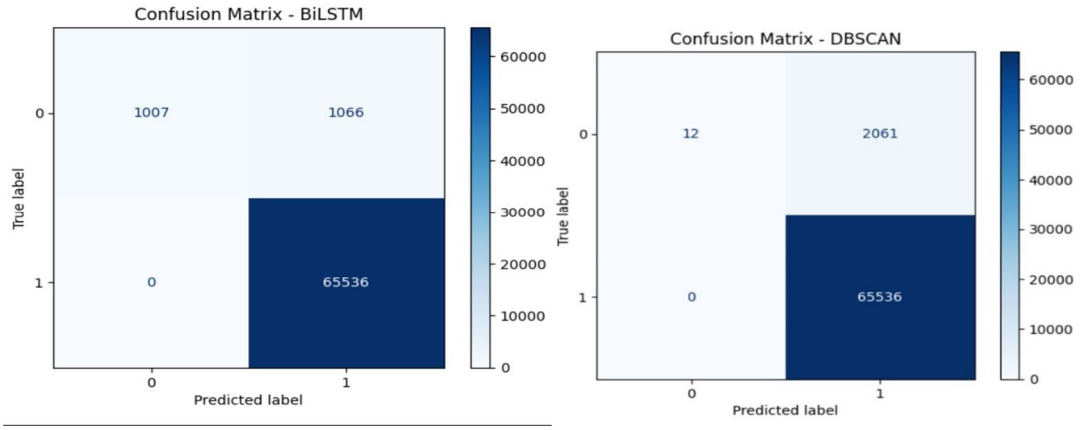



Fig 1.4 Confusion Matrix for Ensemble, SVM, BiLSTM, DBSCAN Model

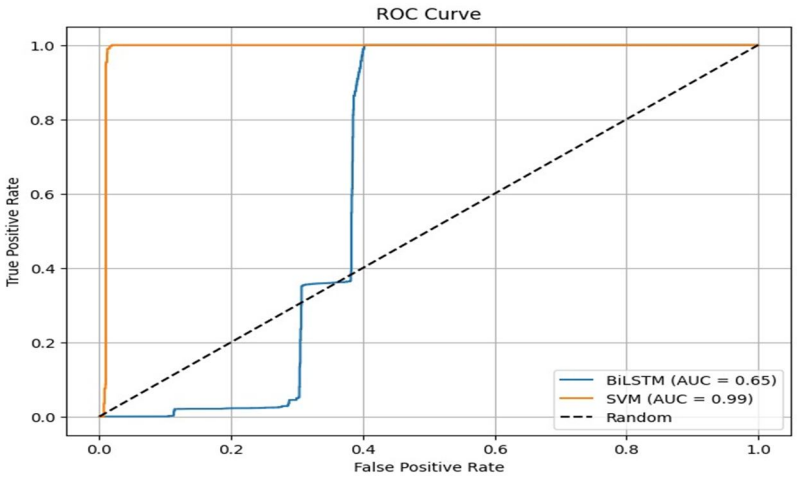



Fig 1.5 ROC Curve for BiLSTM,SVM

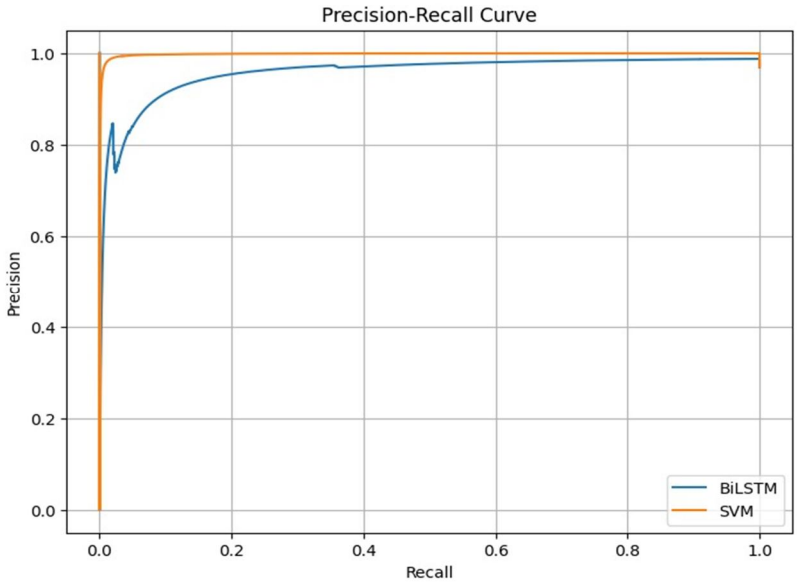



Fig 1.6 Precision-Recall Curve for BiLSTM , SVM

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

### *B. Comparative Analysis*

Table II : Comparison with other papers

| Paper | Methodology Used | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|---|
| Paper 1 | DBSCAN + SMOTE + LSTM | 96.12 | 93.6 | 96.2 | 98.3 |
| Paper 2 | Random Forest (SDN) | 98.97 | 98.33 | 96.37 | 97.34 |
| Paper 8 | ShieldRNN (Seq2Seq + Voting) | 97.998 | 97.992 | 97.983 | 97.993 |
| Paper 11 | SVM (Mininet + 6 features) | 95.24 | 93.20 | 96.10 | 94.63 |
| Our Proposed model | SVM + LSTM + DBSCAN (Hybrid) | 98.41 | 98.39 | 100.0 | 99.19 |

- SVM outperformed other models with high accuracy and low detection latency.
- LSTM showed strong temporal pattern recognition, especially useful in bursty or sustained attacks.
- DBSCAN effectively identified outliers but had limitations in precision due to unsupervised nature.
- All models performed well in detecting SYN and UDP floods; slight drop in performance observed for ICMP floods due to their similarity with benign ping traffic.
- Real-time timeout and alert mechanism worked as intended across all simulated scenarios.

## VI. CONCLUSION AND FUTURE SCOPE

This research highlights the effectiveness of a multi-model framework for DDoS detection. By leveraging the strengths of supervised (SVM), deep learning (LSTM), and unsupervised (DBSCAN) models, we achieve a balance between accuracy and adaptability. The extracted flow-based features capture essential traffic behavior and enable real-time detection.

Future work includes deploying the system in a live Software Defined Network (SDN) environment and integrating an automated mitigation module to block malicious IPs dynamically. Enhancing feature extraction for encrypted traffic and exploring federated learning models are also potential directions.

## REFERENCES

[1] Efendi, R., Wahyono, T., & Widiasari, I. R. (2024). "DBSCAN SMOTE LSTM: Effective Strategies for Distributed Denial of Service Detection in Imbalanced Network Environments", Big Data and Cognitive Computing, 8(9), 118. https://doi.org/10.3390/bdcc8090118

[2] Hirsi, A., Audah, L., Salh, A., Alhartomi, M. A., & Ahmed, S. (2024). Detecting DDoS Threats using Supervised Machine Learning for Traffic Classification in Software Defined Networking. IEEE Access, 12, 166675–166702. https://doi.org/10.1109/access.2024.3486034

[3] Alfatemi, A., Rahouti, M., Amin, R., ALJamal, S., Xiong, K., & Xin, Y. (2024). "Advancing DDoS Attack Detection: A Synergistic Approach Using Deep Residual Neural Networks and Synthetic Oversampling", arXiv preprint arXiv:2401.03116. https://arxiv.org/abs/2401.03116

[4] Qing, Y., Liu, X., & Du, Y. (2024). "Mitigating Data Imbalance to Improve the Generalizability in IoT DDoS Detection Tasks", The Journal of Supercomputing, 80, pp. 9935–9960. https://doi.org/10.1007/s11227-023-05829-5

[5] Joloudari, J. H., Marefat, A., Nematollahi, M. A., Oyelere, S. S., & Hussain, S. (2023). "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks." Applied Sciences, 13(6), 4006. https://doi.org/10.3390/app13064006

[6] Silivery, A. K., Rao, K. R. M., & Suresh Kumar, L. K. (2023). "An Effective Deep Learning Based Multi-Class Classification of DoS and DDoS Attack Detection." arXiv preprint arXiv:2308.08803. https://arxiv.org/abs/2308.08803

[7] Abdelkhalek, A., & Mashaly, M. (2023). "Addressing the Class Imbalance Problem in Network Intrusion Detection Systems Using Data Resampling and Deep Learning." The Journal of Supercomputing, 79, 10611–10644. https://doi.org/10.1007/s11227-023-05073-x

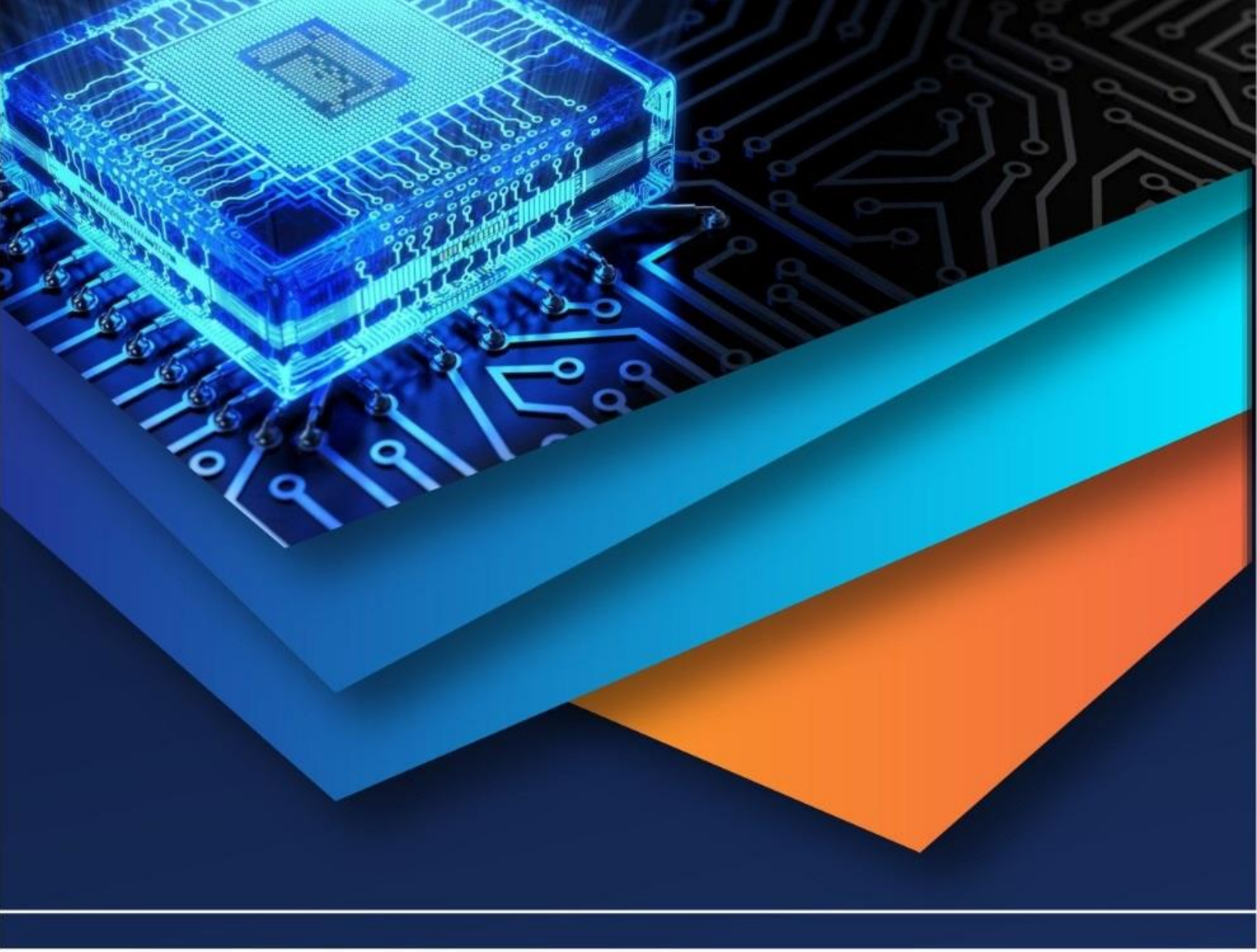
[8] Alasmary, F., Alraddadi, S., Al-Ahmadi, S., & Al-Muhtadi, J. (2022). ShieldRNN: a distributed Flow-Based DDOS detection solution for IoT using sequence majority voting. IEEE Access, 10, 88263–88275. https://doi.org/10.1109/access.2022.3200477

[9] Shafin, S. S., Prottoy, S. A., Abbas, S., Hakim, S. B., Chowdhury, A., & Rashid, M. M. (2021). "Distributed Denial of Service Attack Detection Using Machine Learning and Class Oversampling." In Applied Intelligence and Informatics (pp. 247–259). Springer, Cham. https://doi.org/10.1007/978-3-030-82269-9_19

[10] Calvert, C. L., & Khoshgoftaar, T. M. (2019). "Impact of Class Distribution on the Detection of Slow HTTP DoS Attacks Using Big Data." Journal of Big Data, 6, Article 67. https://doi.org/10.1186/s40537-019-0230-3

[11] Ye, J., Cheng, X., Zhu, J., Feng, L., & Song, L. (2018). A DDOS attack detection method based on SVM in software defined network. Security and Communication Networks, 2018, 1–8. https://doi.org/10.1155/2018/9804061

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)