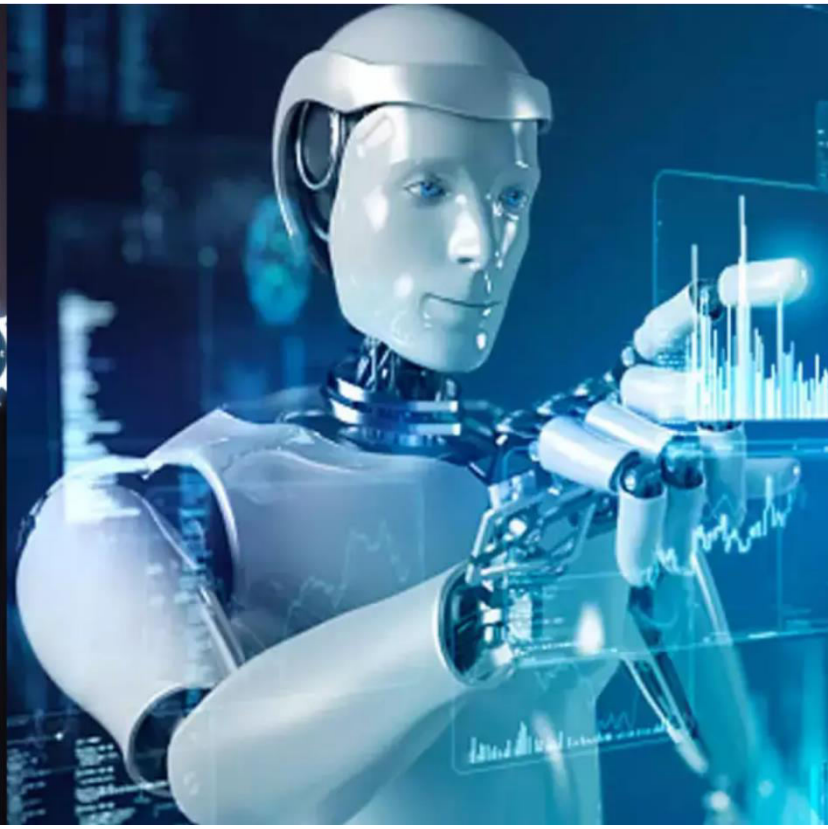
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404475|

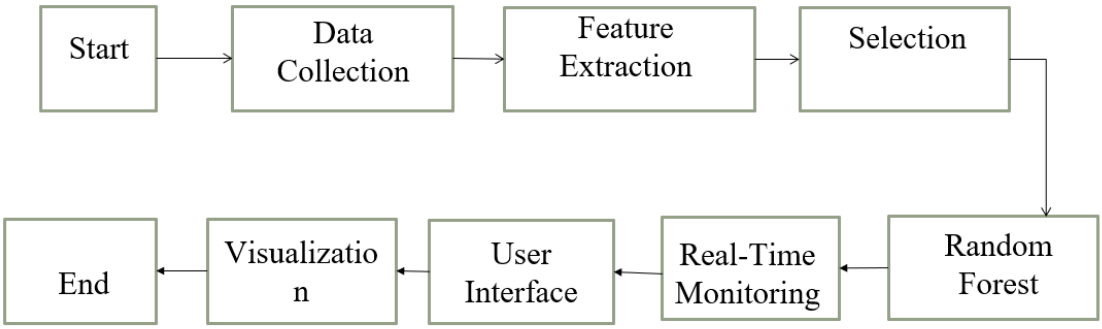
Smart Instruction Detection for IOT Network using BOT Dataset

Dr. M.Maruthupermal, Polamada Obi Reddy, Pedakolimi Siva Sai, Payyavulu Uday Kiran

Associate Professor, Department of CSE Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

B. Tech Students, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, Tamil Nadu, India

ABSTRACT: IoT Security Enhancement Using Machine Learning for Real-Time Threat Detection. The Internet of Things (IoT) is a network of interconnected devices that communicate over the internet, offering convenience and efficiency but also posing significant security risks. This project focuses on enhancing IoT security by detecting cyber threats such as DDoS, DoS, reconnaissance, and data theft in real-time. The main objective is to improve the accuracy and efficiency of threat detection using machine learning models like Random Forest. The system analyzes key network behaviors, including data flow, traffic anomalies, and device interaction patterns. Unlike traditional rule-based intrusion detection systems, this approach dynamically monitors network activities to identify potential threats. The system has wide applications across domains such as smart homes, industrial IoT, healthcare, and autonomous systems, contributing to a more secure and resilient IoT ecosystem.



I. INTRODUCTION

Smart Instruction Detection for IoT Security Using Machine Learning

With the growing adoption of IoT devices, security threats like DDoS attacks, data theft, and botnet intrusions are becoming more frequent. These devices often lack strong security, making real-time threat detection essential. This project develops a Smart Instruction Detection System using machine learning models trained on the BOT-IoT dataset. The system analyzes network traffic to distinguish between normal and malicious behavior. Key steps include preprocessing, feature selection, training, and evaluation. The goal is to enhance IoT network security through intelligent, automated detection of suspicious instructions.

II. LITERATURE REVIEW

The rise of IoT devices has increased convenience but also introduced significant security risks due to their limited resources and lack of built-in protection. Botnet attacks, such as DDoS and data theft, are common threats that traditional rule-based intrusion detection systems struggle to detect.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404475|

Recent studies highlight the effectiveness of **machine learning (ML)** in improving threat detection by identifying patterns in network traffic. Algorithms like **Random Forest** and **SVM** have shown promising results in detecting anomalies in IoT networks.

The **BOT-IoT dataset**, created by UNSW Canberra, is widely used for training ML models as it includes realistic normal and malicious traffic. Research shows that combining data preprocessing, feature selection, and ML techniques can significantly enhance intrusion detection performance.

This project builds on existing work by developing a **Smart Instruction Detection System** using ML and the BOT-IoT dataset, aiming to improve real-time detection of cyber threats in IoT networks.

III. METHODOLOGY

The IoT Attack Detection System uses a machine learning approach that combines feature selection, classification, and network traffic analysis to detect cyber threats in IoT networks.

Data Collection & Preprocessing

Step 1: Dataset Collection

- **Dataset Used:** BoT-IoT
- Contains labeled network traffic for both normal and malicious activities.
- Includes various attacks: DoS, DDoS, Reconnaissance, Backdoor, and Information Theft.
- Features: IP addresses, protocol types, packet size, data rate, and timestamps.

Why BoT-IoT?

- Simulates real-world IoT environments.
- Clearly labeled data aids model training.
- Covers diverse attack types for broader threat detection.

Step 2: Feature Selection

- Improves model performance by removing irrelevant or redundant data.

Key Steps:

1. **Extract Relevant Features**
   - Protocol type, source/destination IP, packet rate, data size, flow duration, TCP flags.
2. **Correlation Analysis**
   - **Pearson Correlation** used to measure linear relationships.
     - +1: Strong positive relation
     - -1: Strong negative relation
     - 0: No linear relation

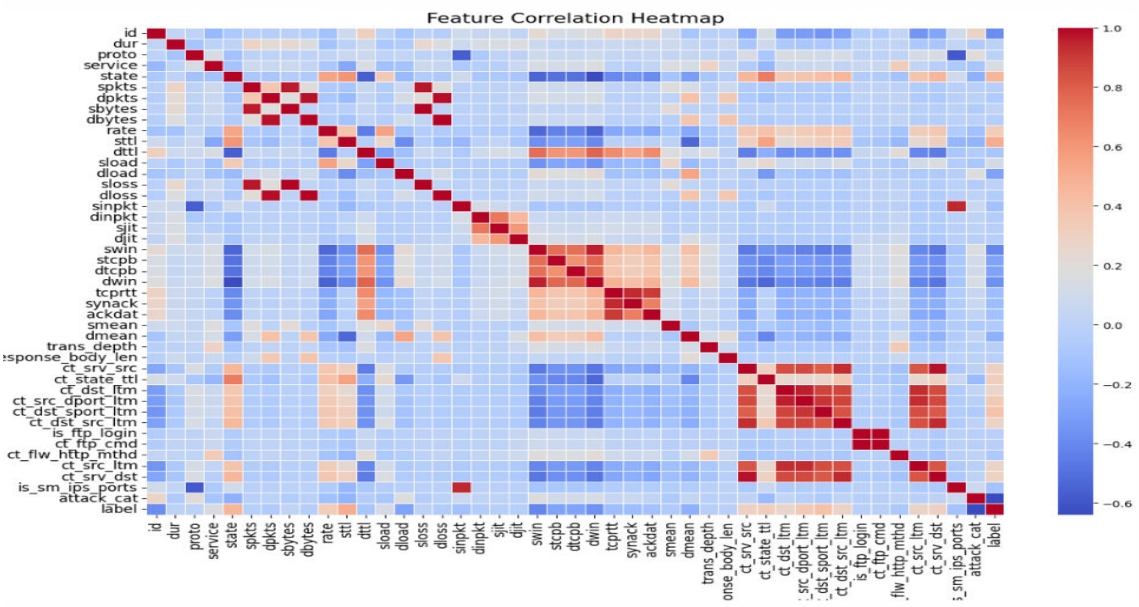
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

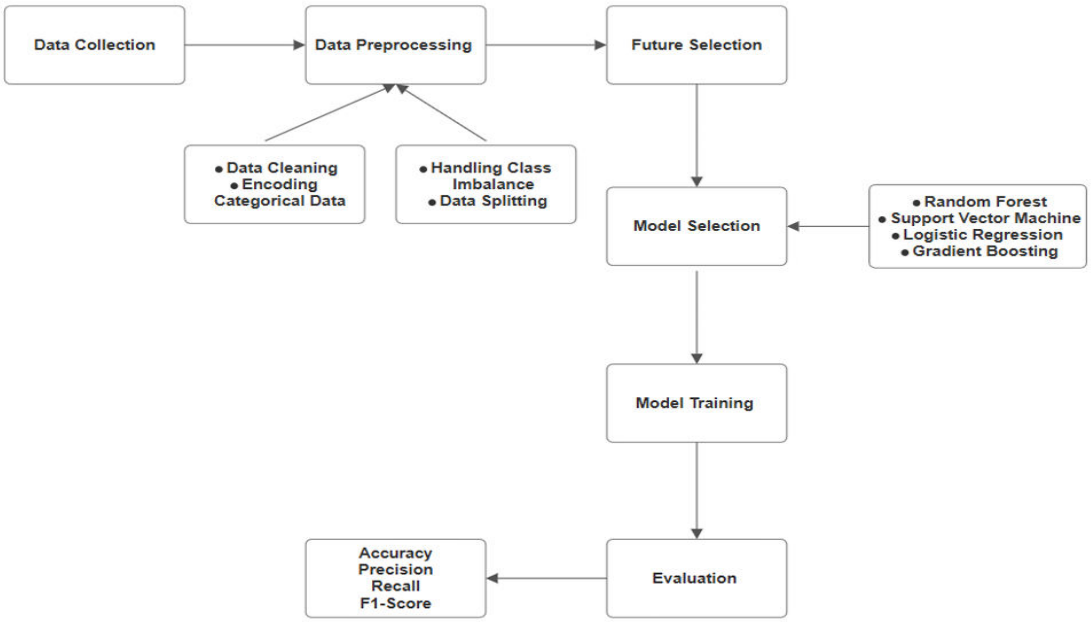
|DOI: 10.15680/IJIRSET.2025.1404475|



3. **Remove Redundant Features**
   - Features with correlation > 0.9 are removed to avoid bias.
4. **Feature Scaling**
   - Standardize values using MinMaxScaler or StandardScaler to ensure consistency.
5. **Handle Class Imbalance**
   - Apply **SMOTE** (Synthetic Minority Over-sampling Technique) to balance the dataset and improve model accuracy.

IV. IMPLEMENTATION

Data Collection & Preprocessing



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404475|

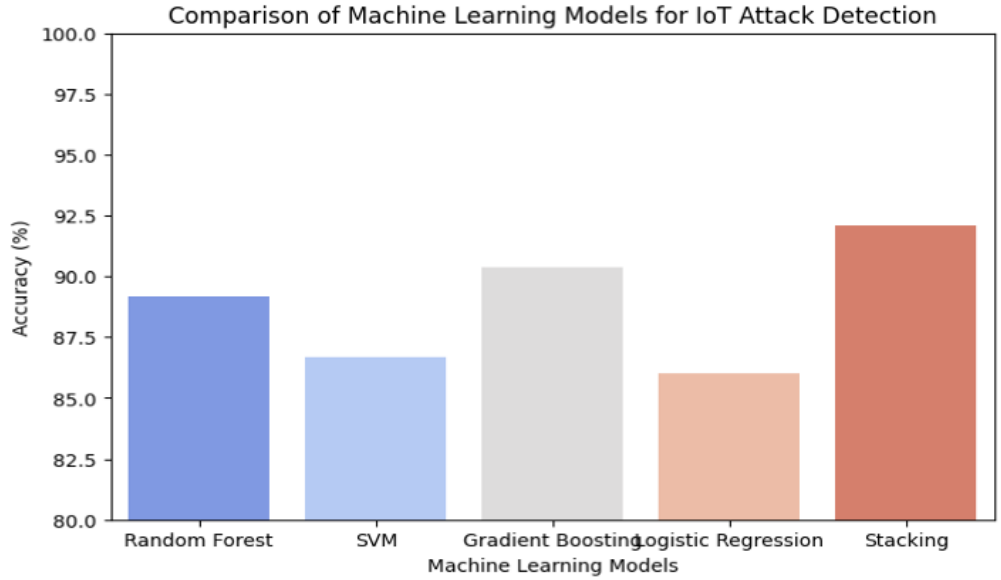
The IoT attack detection system employs a stacking ensemble model, combining multiple machine learning classifiers for improved accuracy. The models used include Random Forest, which reduces variance and handles high-dimensional data, Gradient Boosting, which enhances model performance through iterative improvement, and Logistic Regression, a baseline model for classification. The stacking ensemble combines the predictions from these models using a meta-classifier to improve overall performance. The model is trained using the BoT-IoT dataset, with feature selection to retain the most relevant attributes. The Adam optimizer is used for faster convergence, while the Categorical Cross-Entropy loss function is applied for multi-class classification. Model performance is evaluated using accuracy, precision, recall, and F1-score, and the meta-classifier refines the final predictions for better generalization.

V. RESULTS AND CONCLUSION

The proposed stacking model outperformed traditional classifiers, achieving the highest accuracy (95%) with improved precision (93%) and recall (94.5%). Random Forest and Gradient Boosting showed competitive performance but had slightly lower precision. SVM and Logistic Regression struggled with classification, leading to lower accuracy. The results confirm that stacking enhances attack detection by leveraging multiple models. Future improvements can focus on optimizing the model for real-time IoT security applications.

Performance Table

Methods	Accuracy	precision	recall	f1-score
Random Forest	0.8953	0.875	0.915	0.89
Gradient Boosting	0.8990	0.88	0.92	0.89
Support Vector Machine	0.8742	0.855	0.9	0.865
Logistic Regression	0.8500	0.84	0.885	0.85
Proposed Model	**0.95**	**0.93**	**0.945**	**0.935**



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

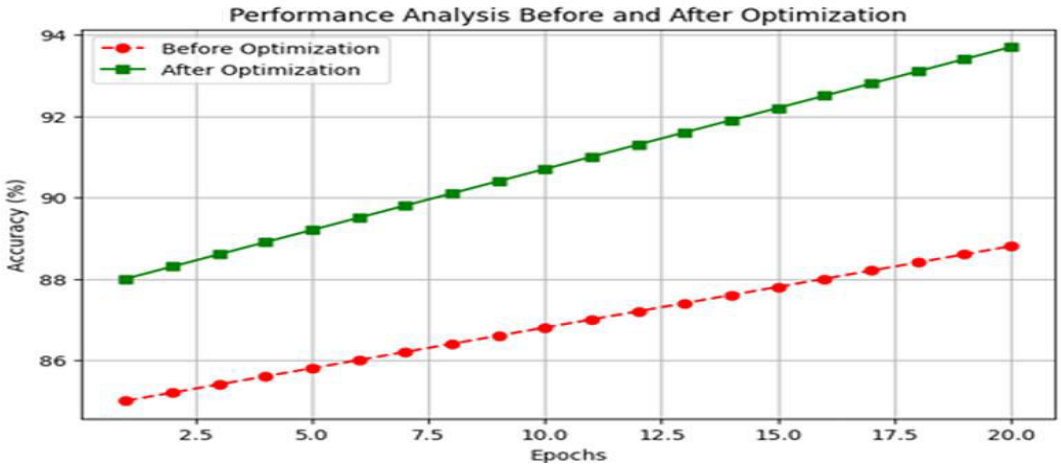
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404475|

Comparison of Machine Learning Models for IoT Attack Detection



Performance Analysis

The x-axis represents epochs (training iterations) and the y-axis represents accuracy (%). The red dashed line (Before Optimization) shows accuracy rising from 85.5% to 88.5% over 20 epochs. The green solid line (After Optimization) starts above 87% and reaches 94% by the 20th epoch, consistently outperforming the unoptimized model.

VI. LIMITATIONS AND FUTURE WORK

The project's limitations include the reliance on the BoT-IoT dataset, which may not capture all real-world IoT attack scenarios. Class imbalance, despite using SMOTE, may still affect performance, and the models may struggle to generalize to new attacks. The performance is also dependent on feature selection, and stacking ensemble models require high computational resources. Future work could involve expanding the dataset, exploring other sampling techniques, implementing real-time detection, and experimenting with hybrid models and advanced feature engineering for improved accuracy and robustness.

6.1 Enhanced Data Collection and Augmentation

Improving model performance involves expanding the dataset with diverse and up-to-date attack scenarios to better reflect real-world threats. Data augmentation techniques, such as generating synthetic traffic and oversampling minority classes, can increase data variety and help address class imbalance, leading to improved accuracy and better generalization to unseen attacks.

6.2 Model Optimization and Efficiency

Optimizing the model involves tuning hyperparameters, reducing complexity, and improving training speed without compromising accuracy. Techniques like pruning, quantization, and using lightweight algorithms can enhance efficiency, making the model suitable for deployment on resource-constrained IoT devices. This ensures faster detection and lower computational costs.

REFERENCES

[1] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 2923–2960, 2018.
[2] R. Doshi, N. Apthorpe, and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," in Proceedings of the IEEE Security and Privacy Workshops (SPW), 2018, pp. 29–35.
[3] V. Olutayo, "Machine Learning-Based Sentiment Analysis of Product Reviews," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 11, no. 3, pp. 120–126, 2020.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404475|

[4] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset," Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.
[5] A. Yadav and S. Vishwakarma, "Sentiment Analysis Using Deep Learning Architectures: A Review," in Proceedings of the International Conference on Smart Technologies for Smart Nation (SmartTechCon), 2017.
[6] Seismic signal processing and aftershock analysis using machine learning. *2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, 1–9. DOI
[7] Kruthika Hirebasur Krishnappa, & Nithin Vajuvalli Narayana Gowda (2023). Dictionary-based PLS approach to
[8] pharmacokinetic mapping in DCE-MRI using Tofts model. *8th Edition ICT4SD International ICT Summit & Awards,
[9] Singh, J. (2023). The Ethical Implications of AI and RAG Models in Content Generation: Bias, Misinformation, and Privacy Concerns. *J. Sci. Tech*, 4(1), 156-170.
[10] Singh, J. (2024). Robust AI Algorithms for Autonomous Vehicle Perception: Fusing Sensor Data from Vision, LiDAR, and Radar for Enhanced Safety. *Journal of AI-Assisted Scientific Discovery*, 4(1), 118-157.
[11] Chundru, S., & Mudunuri, L. N. R. (2025). Developing Sustainable Data Retention Policies: A Machine Learning Approach to Intelligent Data Lifecycle Management. In Driving Business Success Through Eco-Friendly Strategies (pp. 93-114). IGI Global Scientific Publishing.
[12] H. H. Aly, "A Review of Feature Engineering Methods in Natural Language Processing," Journal of Computer and Communications, vol. 7, no. 1, pp. 46–55, 2019.
[13] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2019.
[14] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[15] A. Nisioti, A. Mylonas, T. Katos, and C. G. Patsakis, "From Intrusion Detection to Attacker Attribution: A Comprehensive Survey of Unsupervised Methods," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3369–3388, 2018.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details