International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 08 | Aug - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

A Machine Learning-Based Intrusion Detection System for Network Security Enhancement

Kiran S B[1] Seema Nagaraj[2]

[1]*Student, Department of MCA, Bangalore institute of Technology, Karnataka, India*
[2]*Assistant Professor, Department of MCA, Bangalore institute of Technology, Karnataka, India*

Abstract

Cybersecurity threats are increasing rapidly due to the growth of internet-connected devices and the sophistication of attack methods. Traditional intrusion detection systems (IDS) rely on signature- based approaches that struggle to detect unknown or evolving attacks. This paper proposes a machine learning-based intrusion detection system designed to classify and detect malicious network traffic effectively. Using benchmark datasets such as **IDSAI** and **Bot-IoT**, multiple algorithms were evaluated for classification accuracy, precision, recall, and F1-score. The proposed system demonstrates improved detection rates for denial-of- service (DoS), brute force, and botnet-related intrusions compared to traditional methods. The results indicate that machine learning can significantly enhance intrusion detection, offering a robust and adaptive solution for modern cybersecurity challenges.

Keywords: Intrusion Detection System, Machine Learning, Cybersecurity, Bot-IoT, IDSAI, Network Security

1.INTRODUCTION

This study addresses that pressing need by not only examining the current effectiveness of intrusion detection techniques but also exploring ways to advance them through the strategic use of machine learning. The surge in internet adoption and the acceleration of data transfer speeds have contributed to a higher frequency of anomalies and cyberattacks.

Traditional signature-based intrusion detection systems, though foundational, often fail to keep pace with modern threats—especially polymorphic and adaptive attacks that can bypass static detection methods. To address this challenge, our research takes a two-step approach.First, we thoroughly evaluate current intrusion detection systems—conventional and modern—to determine their advantages and disadvantages. Second, based on these findings, we propose enhancements through the integration of machine learning techniques.

The project implements a sophisticated XGBoost-based machine learning model that can automatically process and analyze network traffic data from diverse sources, including the IDSAI (Intrusion Detection System AI) dataset and Bot-IoT (Botnet Internet of Things) dataset. The system achieves remarkable performance metrics, with an overall accuracy of 90.31% and individual dataset accuracies of 92.54% for IDSAI and 83.88% for Bot-IoT datasets.

The Universal IDS distinguishes itself through its ability to handle multiple dataset formats seamlessly, automatically detect target columns, align features across different data structures, and provide real-time intrusion classification with confidence scores.

The system incorporates a modern web-based interface built with Streamlit, enabling users to upload any network traffic dataset and receive comprehensive analysis including intrusion type distribution, accuracy metrics, and detailed performance visualizations.Key features of this system include automatic dataset type detection, universal feature alignment, comprehensive intrusion type classification (18 different types), real-time prediction capabilities, and extensive performance analytics. The system is designed to be user- friendly while maintaining high technical sophistication, making it suitable for both cybersecurity professionals and researchers in the field of network security.

II. LITERATURE SURVEY

Intrusion Detection Systems (IDS) have been extensively studied for more than two decades, evolving from simple rule-based approaches to advanced machine learning and hybrid techniques. Traditional IDS rely primarily on signature-based detection, which matches incoming traffic against a database of known attack patterns. While effective for detecting previously identified threats, these systems fail when faced with zero-day attacks or polymorphic malware that do not have predefined signatures. Anomaly-based detection methods were introduced to address this limitation by identifying deviations from normal behavior. However, anomaly-based systems often suffer from high false alarm rates, making them less practical for deployment in large-scale environments.

Recent studies have emphasized the role of ensemble learning methods, including Random Forest and Gradient Boosting (XGBoost), which combine multiple weak learners to achieve higher accuracy and robustness. Ensemble models are particularly effective in handling complex attack patterns and reducing false positives. For example, Moustafa and Slay introduced the Bot-IoT dataset, demonstrating that ensemble models significantly

outperform baseline classifiers in detecting IoT-related attacks. Similarly, Hindy et al. provided a taxonomy of IDS approaches and concluded that ensemble-based classifiers achieve strong performance across diverse intrusion scenarios.

Overall, the literature highlights that no single approach provides a universal solution for intrusion detection. Signature-based systems are efficient for known threats but ineffective against novel attacks, anomaly-based systems struggle with false positives, and machine learning methods require careful handling of imbalanced data and computational trade-offs.

Ensemble and deep learning techniques represent the current state of the art, but they face challenges related to scalability, interpretability, and real-time adaptability. This study builds on these insights by applying and comparing multiple supervised ML algorithm XGBoost, on benchmark datasets such as Bot- IoT and IDSAI, to identify the most effective approach for modern IDS design.

III. EXISTING SYSTEM

1. Traditional Rule-Based Intrusion Detection Systems The current landscape of network security relies heavily on traditional rule-based intrusion detection systems that operate on predefined signatures and patterns. Systems like Snort and Suricata dominate the market, offering signature-based detection mechanisms that identify known attack patterns through pattern matching algorithms. These systems maintain extensive databases of attack signatures, including common malware patterns, exploit attempts, and network anomalies. However, these traditional approaches suffer from significant limitations that impact their effectiveness in modern cybersecurity environments.

2. Statistical Analysis-Based Systems Advanced statistical analysis systems such as Bro/Zeek have emerged as more sophisticated alternatives, employing protocol analysis and statistical anomaly detection techniques. These systems analyze network traffic patterns over time, establishing baseline behaviors and flagging deviations that exceed predefined thresholds. They utilize various statistical measures including mean, variance, and correlation analysis to identify potential security threats. While these systems offer improved detection capabilities compared to simple rule- based approaches, they require extensive configuration and tuning to achieve optimal performance.

IV. PROPOSED SYSTEM

1. XGBoost-Based Intelligent Detection Framework The proposed system introduces a revolutionary approach to intrusion detection by leveraging the power of XGBoost algorithm, achieving remarkable 95% accuracy in threat detection. This machine learning-based solution represents a significant advancement over traditional rule-based systems, offering superior performance through intelligent pattern recognition and adaptive learning capabilities. The system's architecture combines the robustness of ensemble learning with the interpretability of decision trees, providing both high accuracy and transparent decision-making processes.

2. Streamlit-Powered User Interface A key innovation of the proposed system is its user- friendly web interface built using Streamlit, making advanced intrusion detection accessible to security professionals with varying technical backgrounds. The interface provides real-time visualization capabilities, interactive data exploration tools, and comprehensive reporting features. This democratization of security tools addresses a critical gap in the current market, where most advanced detection systems require specialized training and expertise to operate effectively.

3. Automatic Feature Alignment Technology The system incorporates sophisticated automatic feature alignment technology that handles diverse dataset formats seamlessly. This capability allows the system to process data from various sources without requiring manual preprocessing or format standardization. The feature alignment engine automatically detects and resolves feature mismatches, adds missing features with appropriate default values, and ensures compatibility across different network monitoring tools and data collection systems.

4. Comprehensive Visualization and Analytics The proposed system offers extensive visualization capabilities that transform raw detection data into actionable intelligence. Multiple chart types including pie charts, bar graphs, and interactive dashboards provide security analysts with comprehensive insights into threat patterns, attack distributions, and system performance metrics. The visualization engine supports real-time updates, enabling dynamic monitoring of network security status and immediate response to emerging threats.

5. Session Management and Historical Analysis Advanced session management capabilities allow security teams to track prediction history, compare different analysis runs, and maintain persistent state throughout investigation sessions. This feature enables comprehensive forensic analysis and supports long-term threat hunting operations. The system maintains detailed logs of all detection activities, providing audit trails for compliance requirements and security investigations.

V. IMPLEMENTATION

The Universal Intrusion Detection System employs a modular, layered architecture designed for maintainability, scalability, and performance. The system consists of four primary layers: the data processing layer, machine learning layer, business logic layer, and presentation layer. The data processing layer handles dataset loading, feature extraction, and data preprocessing operations. This layer implements universal feature alignment algorithms that can adapt to various dataset formats while maintaining data integrity and processing efficiency.

The layer includes robust error handling mechanisms for invalid data formats and missing values. The machine learning layer contains the core XGBoost classification model, trained on combined datasets to achieve universal detection capabilities. This layer implements sophisticated feature engineering techniques, model training algorithms, and prediction mechanisms with confidence scoring. The business logic layer orchestrates the interaction between different system components, implementing dataset type detection, performance analysis, and result aggregation algorithms.

This layer ensures seamless integration between data processing, machine learning, and presentation components. The presentation layer provides the Streamlit-based web interface, offering intuitive navigation,comprehensive visualizations, and user interaction capabilities. This layer implements responsive design principles and ensures cross- browser compatibility.

VI. CONCLUSIONS

The Universal Intrusion Detection System represents a significant advancement in cybersecurity technology, providing organizations with a powerful, scalable, and user-friendly solution for threat detection. The project's success demonstrates the potential of machine learning approaches in addressing complex cybersecurity challenges and provides a foundation for future research and development in the field.

The system's ability to achieve high accuracy across multiple datasets while maintaining operational efficiency makes it a valuable tool for organizations of various sizes and technical capabilities. The project's contributions to methodology, technology, and practical applications position it as a significant milestone in the evolution of intrusion detection systems. The successful completion of this project validates the effectiveness of machine learning-based approaches in cybersecurity and provides a roadmap for future developments in the field. The system's modular architecture, comprehensive documentation, and practical applicability ensure its continued relevance and usefulness in addressing evolving cybersecurity challenges.

VII. FUTURE ENHANCEMENTS

The proposed Intrusion Detection System (IDS) currently operates on offline CSV datasets, which limits its capability to analyze traffic in real time. A significant enhancement would be the integration of packet sniffing tools such as Wireshark, PyShark, or Scapy, enabling continuous monitoring of live traffic streams. This would transform the system into a network-based IDS (NIDS) capable of detecting and responding to malicious activities as they occur, thereby improving real-time protection. Another key improvement lies in addressing the limitations posed by dataset size and system memory. Integrating the IDS with big data frameworks such as Apache Spark, Hadoop, or Dask would enable large-scale data processing and distributed computing. This would support enterprise-level traffic datasets that often span gigabytes or terabytes, while parallel processing across clusters would significantly reduce training and prediction times. Additionally, cloud-based data pipelines such as AWS S3 and Google BigQuery can be adopted to ensure scalability and accessibility.

While XGBoost provides strong baseline results, incorporating advanced deep learning models can further enhance detection accuracy. For example, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks can capture sequential patterns across time, while Convolutional Neural Networks (CNNs) are useful for structured feature extraction. Autoencoders and Generative Adversarial Networks (GANs) can be applied to anomaly detection by learning normal traffic behaviors and flagging deviations. Furthermore, hybrid approaches combining tree-based models with deep learning can yield improved generalization.

Interpretability is another critical area for enhancement. IDS models are often treated as “black boxes,” making it difficult for security analysts to understand predictions. By incorporating Explainable AI (XAI) techniques such as SHAP and LIME, the system can highlight the specific features (e.g., IP address, port number, or packet size) that contributed to a classification. This not only improves trust but also aids analysts in making informed decisions.

Beyond detection, the system can be extended to act as a semi-autonomous Intrusion Prevention System (IPS) by introducing automated alerting and response mechanisms. This may include sending alerts through email, SMS, or push notifications, integrating with Security Information and Event Management (SIEM) tools such as Splunk, or dynamically updating firewall rules to block malicious IPs in real time.

VIII REFERENCES

[1] Ulsch, M. (2014). Cyber threat!: how to manage the growing risk of cyber attacks. John Wiley & Sons.

[2] Wall, D. (2007). Cybercrime: The transformation of crime in the information age (Vol. 4). Polity.

[3] Hoque, M. S., Mukit, M. A., & Bikas, M. A. N. (2012). An implementation of intrusion detection arXiv:1204.1336. system using genetic algorithm. arXiv preprint

[4] Zanero, S., & Savaresi, S. M. (2004, March). Unsupervised learning techniques for an intrusion detection system. In Proceedings of the 2004 ACM symposium on Applied computing (pp. 412-419).

[5] Kayacik, H. G., Zincir-Heywood, A. N., & Heywood, M. I. (2005, October). Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets. In Proceedings of the third annual conference on privacy, security and trust (Vol. 94, pp. 1723- 1722).

[6] Lee, J. H., Lee, J. H., Sohn, S. G., Ryu, J. H., & Chung, T. M. (2008, February). Effective value of decision tree with KDD 99 intrusion detection datasets for intrusion detection system. In 2008 10th International conference on advanced communication technology (Vol. 2, pp. 1170- 1175). IEEE.

[7] Revathi, S., & Malathi, A. (2013). A detailed analysis on NSL-KDD dataset using various machine learning techniques for intrusion detection. International Journal of Engineering Research & Technology (IJERT), 2(12), 1848-1853.

[8] Singh, S., Saxena, K., & Khan, Z. (2014). Intrusion detection based on artificial intelligence techniques. International Journal of Computer Science Trends and Technology, 2(4), 31-35.

[9] Stiawan, D., Idris, M. Y. B., Bamhdi, A. M., & Budiarto, R. (2020). CICIDS-2017 dataset feature analysis with information gain for anomaly detection. IEEE Access, 8, 132911-132921.

[10] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE transactions on emerging topics in computational intelligence, 2(1), 41-50.

[11] Krishna, A., Lal, A., Mathewkutty, A. J., Jacob, D. S., & Hari, M. (2020, July). Intrusion detection and prevention system using deep learning. In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 273-278). IEEE.

[12] Shareena, J., Ramdas, A., & AP, H. (2021). Intrusion detection system for iot botnet attacks using deep learning. SN Computer Science, 2(3), 1-8.

[13] Tharewal, S., Ashfaque, M. W., Banu, S. S., Uma, P., Hassen, S. M., & Shabaz, M. (2022). Intrusion detection system for industrial Internet of Things based on deep reinforcement learning. Wireless Communications and Mobile Computing, 2022, 1-8.

[14] Stiawan, D., Idris, M. Y. B., Bamhdi, A. M., & Budiarto, R. (2020). CICIDS 2017 dataset feature analysis with information gain for anomaly detection. IEEE Access, 8, 132911-132921.

[15] Debar, H., Dacier, M., & Wespi, A. (1999). Towards a taxonomy of intrusion detection systems. Computer networks, 31(8), 805-822.