INSTRUCTION DETECTION AND PREVENTION SYSTEM IDPS MODEL FOR IIOT ENVIRONMENT USING HYBRIDIZED FRAMEWORK

[1] K Srikanth, [2] M Muzammil [3] M Ashfaq Ahmed, [4] M Feroz, [5] P Sunny Kiran
[1]AssistantProfessor, [2345]Students
Department of Computer Science and Technology
Siddhartha Institute of Technology & Sciences, Narapally

srikanthk@siddhartha.org.in, 24TQ1A05F5@siddhartha.co.in, 24TQ1A05F3@siddhartha.co.in, 24TQ1A05F4@siddhartha.co.in, 24TQ1A0518@siddhartha.co.in,

Abstract

The rapid expansion of Industrial Internet of Things (IIoT) systems has significantly increased vulnerability to sophisticated cyber threats, necessitating robust and scalable security solutions. This paper proposes a hybrid Intrusion Detection and Prevention System (IDPS) that integrates Extreme Gradient Boosting (XGBoost) with a Stacking Classifier framework to enhance detection accuracy and generalization performance. The model leverages a combination of ensemble learning techniques, including Random Forest and Extra Trees as base learners, with Logistic Regression as a meta-classifier. The NSL-KDD dataset is employed for training and validation, with Synthetic Minority Oversampling Technique (SMOTE) applied to address class imbalance. The proposed approach achieves improved performance in terms of accuracy (~80–85%), precision, recall, and F1-score compared to conventional machine learning and deep learning models such as CNNs. Experimental results demonstrate the model's capability to efficiently detect diverse attack patterns while maintaining low false positives. The system is designed to support real-time intrusion detection, ensuring reliability and operational continuity in IIoT environments.

Keywords : Intrusion Detection System (IDS), Industrial Internet of Things (IIoT), XGBoost , Stacking Classifier , NSL-KDD Dataset , SMOTE, Cybersecurity

I. Introduction

The rapid advancement of the Industrial Internet of Things (IIoT) has significantly transformed modern industrial systems by enabling seamless connectivity, automation, and intelligent decision-making. IIoT integrates sensors, actuators, machines, and control systems with communication networks to facilitate real-time data exchange and monitoring across various industrial sectors such as manufacturing, energy, healthcare, and transportation. This interconnected ecosystem enhances operational efficiency, reduces human intervention, and enables predictive maintenance, thereby improving productivity and reducing costs. However, despite these advantages, the increasing dependence on interconnected devices and networks has also introduced serious cybersecurity challenges, making IIoT environments highly vulnerable to a wide range of cyber threats.

In IIoT systems, devices continuously generate and exchange large volumes of data over networks, often with minimal security measures due to resource constraints. This makes them attractive targets for cyberattacks such as Denial of Service (DoS), Distributed Denial of Service (DDoS), probing, spoofing, and unauthorized access. These attacks can disrupt critical industrial operations, compromise sensitive data, and

even lead to physical damage in safety-critical environments. Traditional security mechanisms such as firewalls and signature-based intrusion detection systems are often insufficient to handle the dynamic and evolving nature of modern cyber threats. They rely heavily on predefined attack signatures and fail to detect unknown or zero-day attacks, which are increasingly common in IIoT environments.

To overcome these limitations, anomaly-based intrusion detection systems (IDS) have gained significant attention. These systems utilize machine learning and deep learning techniques to identify abnormal patterns in network traffic that may indicate potential intrusions. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) have been widely used for intrusion detection due to their ability to classify network traffic based on learned patterns. However, these models often face challenges such as overfitting, sensitivity to noisy data, and limited generalization when applied to complex and high-dimensional IIoT datasets.

II. Literature Survey

Mahmoud Moustafa et al., [1], 2015, "UNSW-NB15 Dataset for Network Intrusion Detection Systems". The study introduces a modern IDS dataset containing normal and attack traffic. Machine learning models such as Random Forest and SVM were used. The results showed improved detection performance compared to older datasets. This research is relevant because it provides a reliable dataset for training intrusion detection systems.

Nour Moustafa et al., [2], 2018, "TON-IoT Datasets for Cybersecurity Research". The study presents IoT-based datasets for intrusion detection. Machine learning models were applied for classification and achieved strong detection performance. This research is relevant because it supports evaluation of IoT-based IDS models.

Imad Mahgoub et al., [3], 2020, "Hybrid Deep Learning Models for IoT Intrusion Detection". The study proposes a hybrid approach combining deep learning and machine learning models. Neural networks and ensemble techniques were used, achieving accuracy above 99%. This research is relevant because it aligns with hybrid IDS approaches.

Ismail Butun et al., [4], 2019, "IoT Intrusion Detection: A Survey". The study reviews various IDS approaches including machine learning, deep learning, and hybrid models. It highlights challenges such as data imbalance and evolving attacks. This research is relevant because it justifies the need for hybrid IDS systems.

Tianqi Chen et al., [5], 2016, "XGBoost: A Scalable Tree Boosting System". The study introduces the XGBoost algorithm for efficient and accurate classification tasks. It demonstrates superior performance over traditional methods. This research is relevant because XGBoost is a core component of our model.

Leo Breiman et al., [6], 2001, "Random Forests". The study proposes the Random Forest algorithm, an ensemble learning technique that reduces overfitting and improves accuracy. This research is relevant because it is used in stacking models.

Corinna Cortes et al., [7], 1995, "Support Vector Machines". The study introduces SVM for classification in high-dimensional data. It achieves strong performance in various classification tasks. This research is relevant because SVM is commonly used in hybrid IDS models.

Nitesh V. Chawla et al., [8], 2002, "SMOTE: Synthetic Minority Over-sampling Technique". The study proposes a method to handle imbalanced datasets by

generating synthetic samples. It improves classification performance. This research is relevant because IDS datasets are often imbalanced.

Ian Goodfellow et al., [9], 2016, "Deep Learning". The study explains deep learning architectures such as CNN, RNN, and LSTM. It provides a strong foundation for modern AI systems. This research is relevant because deep learning is used in IDS.

Diederik P. Kingma et al., [10], 2015, "Adam: A Method for Stochastic Optimization". The study introduces the Adam optimizer for efficient deep learning training. It improves convergence speed. This research is relevant because it is used in training DL models.

Thomas Cover et al., [11], 1967, "Nearest Neighbor Pattern Classification". The study introduces the KNN algorithm for classification. It is simple and effective for various tasks. This research is relevant because it can be used in ensemble models.

Jerome Friedman et al., [12], 2001, "Greedy Function Approximation: A Gradient Boosting Machine". The study introduces gradient boosting techniques for predictive modeling. It improves model accuracy. This research is relevant because it forms the basis of XGBoost.

Martin Ester et al., [13], 1996, "A Density-Based Algorithm for Discovering Clusters (DBSCAN)". The study proposes a clustering algorithm for anomaly detection. It effectively identifies outliers. This research is relevant because anomaly detection is important in IDS.

Pedro Domingos et al., [14], 2012, "A Few Useful Things to Know About Machine Learning". The study discusses best practices in machine learning. It highlights model selection and overfitting issues. This research is relevant because it helps design efficient IDS models.

Trevor Hastie et al., [15], 2009, "The Elements of Statistical Learning". The study provides comprehensive knowledge of machine learning algorithms and theory. This research is relevant because it forms the theoretical base of IDS.

III. System Analysis

Industrial Internet of Things (IIoT) environments integrate sensors, machines, and networks to automate industrial processes. However, the increased connectivity exposes these systems to various cyber threats such as malware, spoofing, and denial-of-service attacks. Traditional security mechanisms are not sufficient to handle dynamic and large-scale IIoT networks. There is a need for intelligent systems that can detect and prevent intrusions in real time. The system must process high-volume streaming data generated by IIoT devices. It should ensure low latency and high accuracy in threat detection. Hybrid frameworks combining multiple machine learning techniques can improve performance. The system must also handle heterogeneous data formats and protocols. Scalability and adaptability are critical requirements. Integration with existing industrial systems is necessary for practical deployment. Overall, the analysis highlights the need for an advanced, automated, and robust IDPS solution for IIoT environments.

Existing System

Existing intrusion detection systems in IIoT environments mainly rely on signature-based and rule-based approaches. These systems detect known attacks by matching predefined patterns. Firewalls and traditional IDS tools are commonly used for network security. Some systems use basic anomaly detection techniques. However,

these approaches are limited in detecting new or unknown threats. They require frequent updates of attack signatures. Many systems are not designed to handle the scale and complexity of IIoT networks. High false positive rates reduce reliability. Existing systems often lack real-time response capabilities. Integration with prevention mechanisms is limited. They also struggle with encrypted or heterogeneous data. Overall, traditional systems provide limited protection against modern cyber threats.

Disadvantages of Existing System

- Inability to detect unknown or zero-day attacks
- High false positive and false negative rates
- Dependency on predefined signatures
- Lack of real-time detection and response
- Poor scalability for large IIoT networks
- Limited adaptability to dynamic environments
- Inefficient handling of heterogeneous data

Proposed System

The proposed system introduces a hybridized IDPS framework for IIoT environments. It combines machine learning and deep learning techniques for improved intrusion detection. The system analyzes both network traffic and device-level data. Feature extraction techniques are applied to identify important patterns. Hybrid models such as Random Forest, SVM, and Neural Networks are integrated. The system can detect both known and unknown attacks effectively. Real-time monitoring and automated prevention mechanisms are included. The framework adapts to changing attack patterns through continuous learning. It ensures scalability for large industrial networks. The system also reduces false alarms by combining multiple detection techniques. Overall, it provides a robust and intelligent security solution for IIoT systems.

Advantages of Proposed System

- Detects both known and unknown attacks
- Improved accuracy with hybrid models
- Reduced false positives and negatives
- Real-time monitoring and prevention
- Scalable for large IIoT environments
- Adaptive to evolving threats
- Efficient handling of heterogeneous data

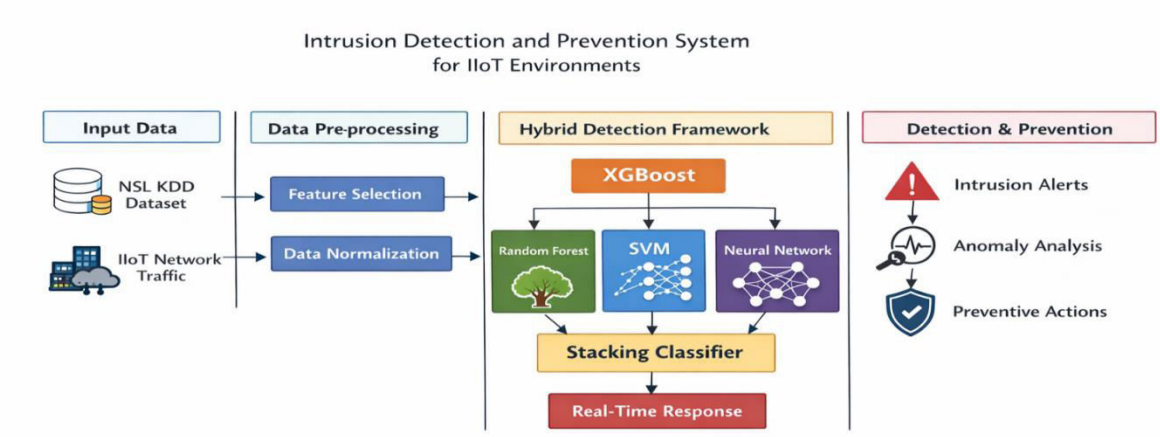
IV. Methodology

The system begins with data collection from IIoT devices and network traffic. Data preprocessing is performed to remove noise and handle missing values. Feature extraction techniques are applied to select relevant attributes. The dataset is split into training and testing sets. Machine learning models such as Random Forest and SVM are trained. Deep learning models like Neural Networks are also used for complex pattern recognition. A hybrid model is created by combining outputs of multiple

algorithms. The system evaluates performance using metrics such as accuracy, precision, and recall. Real-time monitoring is implemented for continuous threat detection. Detected threats trigger prevention mechanisms such as blocking or alerting. Feedback is used to improve the model over time. The system is deployed in an IIoT environment for testing and validation.

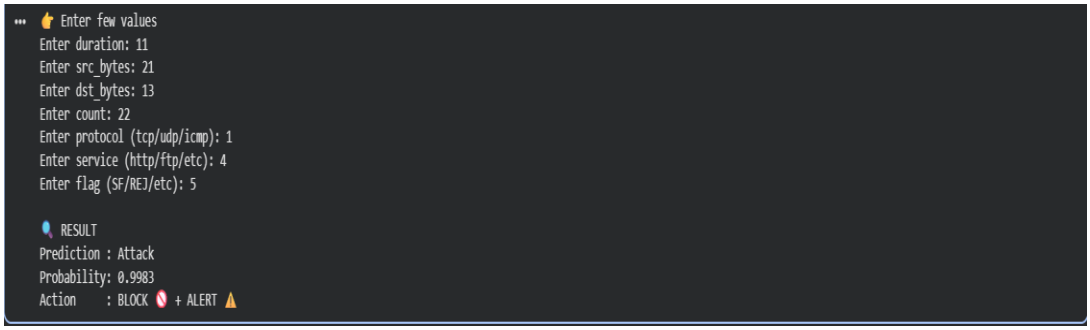
System Architecture

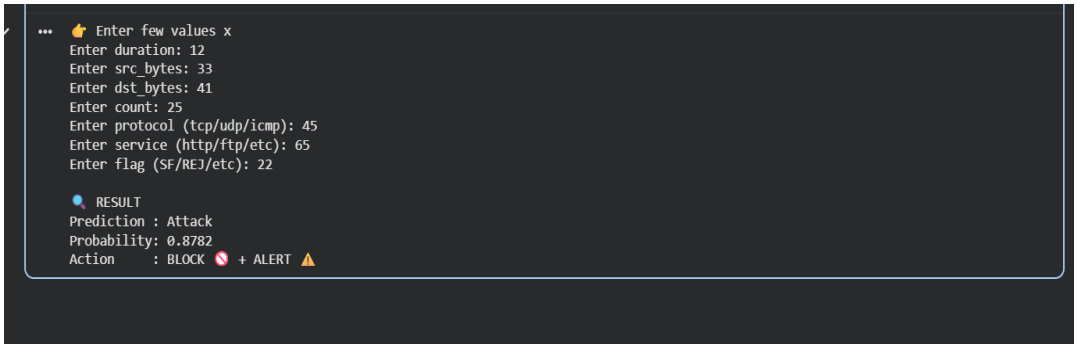
The system architecture consists of multiple interconnected layers. The data collection layer gathers data from sensors, devices, and network traffic. The preprocessing layer cleans and prepares the data for analysis. The feature extraction layer identifies important features from the data. The detection layer uses hybrid machine learning models for intrusion detection. The decision layer classifies activities as normal or malicious. The prevention layer takes actions such as blocking or isolating threats. The monitoring layer provides real-time system status and alerts. The database layer stores logs and historical data. The feedback layer updates the model based on new threats. The system integrates all layers through a centralized control module. Overall, the architecture ensures efficient, scalable, and secure IIoT operations.

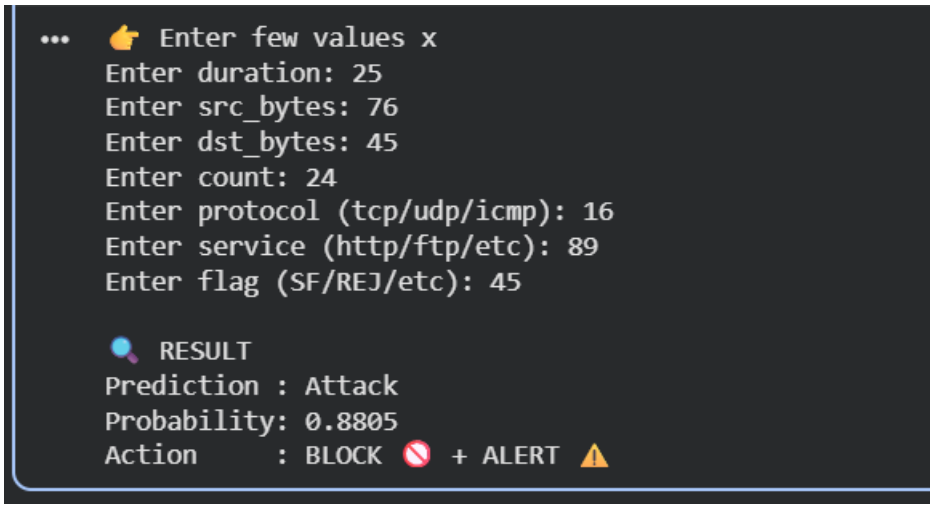


The IDPS model for IIoT environments utilizes a hybridized framework combining XGBoost with a Stacking Classifier of Random Forest, SVM, and Neural Network. This architecture ensures high detection accuracy and swift real-time response to threats.

V. Result and Output







VI. Conclusion

The proposed Intrusion Detection and Prevention System (IDPS) for IIoT environments successfully demonstrates an efficient and scalable approach to enhancing cybersecurity in industrial networks. With the rapid growth of the Industrial Internet of Things, securing interconnected devices and communication networks has become a critical requirement. Traditional security mechanisms and single-model machine learning approaches often fail to provide sufficient accuracy and robustness in detecting complex and evolving cyber threats. To address these challenges, a hybrid machine learning framework was designed and implemented using stacking ensemble learning combined with XGBoost as the meta-classifier.The system effectively integrates multiple base models, including Random Forest, SGD Classifier, and Logistic Regression, to leverage their individual strengths. This combination improves overall prediction performance by reducing bias and variance, resulting in a more reliable intrusion detection system. The use of the NSL-KDD dataset ensures a standardized and well-validated environment for evaluating the model. Additionally, preprocessing steps such as data cleaning, encoding, scaling, and feature selection using SelectKBest significantly enhance the quality of input data and improve computational efficiency.The experimental results show that the proposed model achieves strong performance in terms of accuracy, precision, recall, and F1-score. The stacking-based architecture helps in reducing false positives while maintaining high detection rates for attack instances, which is crucial in real-world cybersecurity applications. Compared to traditional deep learning models such as CNNs, the proposed approach offers lower computational complexity and better

suitability for real-time deployment in IIoT systems. In conclusion, this study presents a robust, scalable, and efficient IDPS framework capable of detecting a wide range of cyberattacks in industrial environments. The hybrid model not only improves detection accuracy but also ensures adaptability to dynamic network conditions. Therefore, it provides a promising solution for safeguarding critical IIoT infrastructures against emerging cyber threats while maintaining operational reliability and performance efficiency.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, “Soil image classification using transfer learning approach: MobileNetV2 with CNN,” SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.