International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at :www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET

Date Driven Security Threat Detection in IIOT Using Random Forest

K. Sravani[1], G.V.S Ananthnath[2]

[1]MCA Student, Department of Master of Computer Applications, KMM institute of postgraduate studies, Tirupati, Annamayya (D.t), Andhra Pradesh, India

[2]Associate Professor, Department of Master of Computer Applications, KMM institute of post Graduate studies, Tirupati, Annamayya (D.t), Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted : 25 May 2025
Published: 30 May 2025

Publication Issue :
Volume 12, Issue 3
May-June-2025

Page Number :
570-576

ABSTRACT

Security threats found within IIoT systems require specialized protection measures to safeguard essential infrastructure for dealing with essential issues. The proposed solution delivers a framework that effectively deals with Cyber Threat Intelligence generation for IIoT deployment contexts. The system recognizes cyber threats by using Random Forest algorithm together with artificial intelligence methods to establish proactive threat prevention that is both effective and automatic. SMOTE serves as an approach for handling unbalanced data in IIoT anomaly detection operations to create more consistent results. Smart Sentry implements analytical procedures throughout its proactive system design for security threat scanning of IIoT activities to ensure operation continuity prior to disruptions.

Keywords: Industry IoT, Cyber Security Intelligence, Artificial Intelligence, Machine Learning, Outlier Detection, Random Forest.

INTRODUCTION

The fourth industrial revolution receives its name from IIoT because it helps businesses worldwide integrate core infrastructure systems for performance enhancement. The same way that IIoT facilitates industrial operations it also creates multiple potential cyber threats that interrupt operations. The problems listed below demonstrate how SmartSentry functions as an Intrusion Detection System which focuses on IIoT attacks while Gerry Walsh keeps his position as a well-known presence in the cybersecurity education space through extensive time spent in academics. Through a combination of Artificial Intelligence and Profound Brain Organization along with algorithms such as Arbitrary Woodland, Choice Tree, Backing Vector Machine, k-Closest Neighbor and FS based Indicators Smart Sentry achieves threat detection and elimination functions. Engineered Minority Over-examining Method (Destroyed) serves to manage data inconsistency by continuing operations against threats and information imbalance[1].

A. Objective Of The Study

SmartSentry serves as a smartphone application which aims to create an effective CTI solution that IIoT structures can implement. The application targets to boost asset security protection against cybertoxin attacks through real-time operation of new functional deep learning and machine learning algorithms. SmartSentry implements various algorithms which enable high identification accuracy of system anomalies alongside security threats through its synthetic minority over-sampling technique (SMOTE). The main objective of this project involves building automated monitoring systems that maintain IIoT operations by addressing emerging threats.[8].

B. Scope Of The Study

Security enhancement focuses on protecting against threats which affect IIoT systems operating in such networks. This framework will establish and operate intelligent artificial intelligence through machine learning and deep learning methods to enable quick threat detection online and determine suitable responses to online threats. The investigators sought to enhance the Ml model results by employing synthetic minority over-sampling technique for data set balance control. Smart Sentry offers threat checking and appraisal functions which results in discovering new security risks whereas it informs required parties. The objectives of this system focus on user interface development for process monitoring capabilities and incident management tools and security step adoption by main actors to safeguard IIoT systems. SmartSentry provides substantial protection to critical assets needed to function effectively with the increase in industrial revolution connectivity.[7].

C. Problem Statement

While the growth of the Industrial Internet of Things (IIoT) has significantly enhanced system operational efficiency, it has also introduced increased vulnerabilities in critical infrastructure, exposing systems to a higher risk of cyberattacks. Traditional security technologies often fall short in managing these IIoT-related threats due to the sheer volume and high speed of data being processed. Existing approaches, particularly those from Group 1, show limitations in anomaly detection because of highly imbalanced datasets and rapidly evolving threat landscapes, which result in weak detection accuracy and inadequate response mechanisms.

The new cybersecurity paradigm must be systemic or holistic enough to deliver the pressing requirements for the exclusive environments of IIoT. In addition, it is at this point that Cyber Threat Intelligence (CTI) architecture becomes vital feature or element for IIoT systems, as it provides the strategic structural support it needs. SmartSentry obtains its claim to superiority through the combination of machine learning and deep learning for cyber threat detection inside enterprise networks. It is intended to contribute to improving the whole security posture of IIoT infrastructures through enhancing detection of threats and their effectiveness in incident response during real time operations.

RELATED WORK

Research into Industrial Internet of Things (IIoT) security has gained interest because IT threats are increasing in number. Research has produced some works about CTI frameworks based on ML and DL approaches designed to find anomalies in IIoT systems. [2]

The two main IDS tools - Snort and Suricata provide limited protection because their signature-based detection systems fail to stop zero-day threats. Computer programs based on anomaly detection methods currently provide solutions to overcome these existing limitations in the field. RF and DT serve as popular choices for IIoT threat detection since they provide effective interpretation along with threat identification abilities. The datasets have an

imbalanced ratio of magnitude which results in a high number of false negative outcomes.[3].

Scientists examine deep learning models starting with Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) organizations and autoencoders to boost detection precision of complex attacks in IoT systems. The dataset undergoes adjustment through the Manufactured Minority Over-testing Method (Destroyed) to enhance classifier efficiency[4].Scientists have combined SVM, KNN with the Extra Tree classifier (ETC) to enhance detection rates. The SmartSentry system implements real-time information analysis in addition to anomaly detection mechanisms to embed cyber danger into IIoT infrastructure.[5].

The path of audit enabled by blockchain has been implemented in several works for providing secure threat intelligence data that is difficult to modify. The development of federated learning serves as a new solution to train security models across IIoT devices to protect privacy frameworks [10]. The proposed SmartSentry solution achieves enhanced operational capabilities as well as improved resistance against .Parcelable cyber attacks in the IIoT environment.[6]

1. Proposed System Workflow

A. Problem Definition

Industrial automation systems see more widespread implementation because organizations continue expanding their investment in IIoT solutions. The SmartSentry system uses ML and DL concepts to deliver specific solutions against cyber threats found within Industrial IoT (IIoT) networks. The main goals of the system consist of two essential points.

1. Measures that can be taken to minimize false positive and false negative results in cyber threats detection are:
2. Meaningful real-time protection from threats in order to maintain the integrity and resilience of IIoT systems..

B. Data Collection

The system collects **real-time and historical cybersecurity data** from IIoT devices, including:

- **Network traffic logs**
- **Device access logs**
- **Intrusion detection system (IDS) alerts**
- **System vulnerability reports**

To improve model generalizability, data is sourced from **varied environments** with different attack patterns, ensuring diverse and representative training datasets.

C. Data Preprocessing

Data is **cleaned and structured** by:

- **Eliminating duplicate records** and irrelevant features.
- **Handling missing values** through imputation.
- **Standardizing log formats** across different IIoT devices and protocols.

D. Normalization

To ensure consistency across all data sources, features are **normalized or standardized** using techniques such as **Min-Max scaling and Z-score normalization**.

- **Categorical features**, such as attack types, are encoded using **one-hot encoding** or **label encoding**.
- **Log transformations** are applied to numerical attributes to improve model interpretability.

E. Model Development

Smart Sentry implements a **hybrid approach** combining machine learning and deep learning models for **cyber threat detection. Model execution alludes to the most common way of characterizing and setting up an AI or profound learning model engineering that will be utilized to make forecasts or tackle a particular undertaking (like grouping, relapse, or item recognition). With regards to profound learning, the model is commonly made out of layers of fake neurons (frequently called hubs) associated by loads and predispositions that are changed during preparing**

F. Model Evaluation

The dataset is split into **training and testing sets** using an **80-20 split**. Model performance is assessed using:

- **Accuracy, Precision, Recall, F1-score** to measure classification performance.
- **ROC-AUC Curve** to evaluate anomaly detection efficiency.
- **Cross-validation** to ensure model stability and robustness.

By integrating **real-time analysis, automated anomaly detection, and adaptive learning**, SmartSentry ensures a **proactive and resilient cybersecurity framework for IIoT environments**.

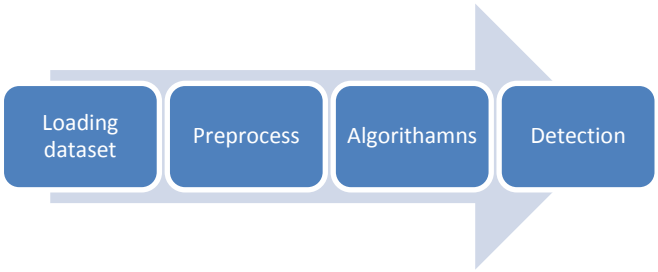

Fig 1: System Architecture of Analysis

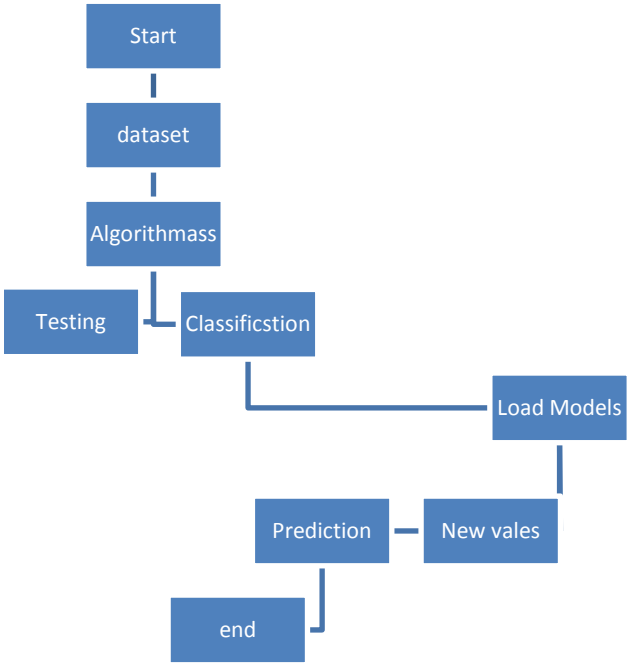

Fig 2: Testing Flow chart of Analysis

Methodology

1) Random Forest:

The high-speed growth of the Industrial Internet of Things (IIoT) has brought about a miraculous improvement in operational efficiency systems, but it has also come with a slew of new vulnerabilities, rendering critical infrastructure increasingly at risk of cyber threats. Traditional security mechanisms often struggle to manage these threats effectively, largely due to the high velocity and vast volume of data processed within IIoT systems. Existing methods—particularly those categorized under Group 1—exhibit notable shortcomings in anomaly detection, primarily because of heavily imbalanced datasets and the dynamic nature of modern threat environments. These issues contribute to reduced detection accuracy and delayed response times.

To overcome these limitations, there is a clear need for a more resilient and adaptive cybersecurity framework specifically designed for IIoT contexts. Cyber Threat Intelligence (CTI) architecture offers a promising solution by providing a structured and strategic approach to support the evolving security needs of IIoT systems. One such implementation is SmartSentry, which addresses these challenges by incorporating both machine learning and deep learning algorithms to detect and respond to cyber threats in enterprise networks. The project is focused on strengthening the security of IIoT infrastructures, enhancing real-time threat detection, and optimizing incident response to support continuous and secure industrial operations.

The Random Forest is a proven decision tree type that possesses a few outstanding perfectionistic values such as power for efficient classification and good regression performances. One of its key advantages is its superior accuracy compared to the traditional Decision Tree algorithm. It is particularly effective in handling datasets with missing values, offering robust performance even when some data entries are incomplete. Unlike many other algorithms, Random Forest does not require extensive hyperparameter tuning to achieve good prediction results, which makes it user-friendly and practical for real-world applications.

Another significant strength of the Random Forest algorithm is its ability to address overfitting, a common issue in Decision Trees. By constructing multiple trees and aggregating their outputs, the model reduces variance and enhances generalization. At every node split, a random feature subset is chosen by the algorithm while constructing individual trees in the forest. This randomness increases diversity among trees and improves robustness in the model as well as its performance.

The Random Forest algorithm is built upon the foundation of decision trees, with nodes serving as the starting points for their construction. A decision tree acts as a structured framework for organizing decisions in a tree-based model, and understanding how it functions offers valuable insight into the inner workings of a Random Forest. A typical decision tree is composed of three main components: root nodes, decision nodes, and leaf nodes. The tree begins at the root node and systematically splits the training data into subsets. Each split forms a branch that can lead to further subdivisions, continuing until a terminal point is reached this is known as the leaf node, which cannot be divided any further. At each node, a decision is made that predicts the outcome based on the data at that point. These decision points collectively form a flowchart that guides the data through the tree toward a final classification or prediction. The final outcome is determined at the leaf node, which represents the conclusion of that particular decision path. While this structure makes decision trees interpretable and easy to visualize, it also introduces complexity when dealing with deeper trees and large datasets. A diagram accompanying this explanation typically highlights the structure of the tree and categorizes the three main node types, helping to clarify the roles each plays within the overall model.

RESULT

RandomForestClassifier Models Results

```
Accuracy Score of RandomForest=  0.9971428571428571
=====================================================
f1 score score fo RandomForest = 0.9971629505314473
=====================================================
precision_score of RandomForest is= 0.99707863914336
=====================================================
recall_score of RandomForest is = 0.9972536874407818
=====================================================
coonfusion matrxi of RandomForest =
 [[429   0   0]
 [  0 484   2]
 [  2   0 483]]
```

Fig 5: Metrics-Accuracy, F1 score etc random Forest

The confusion matrix provides a clear overview of the Decision Tree model's performance across three distinct classes. It reveals that the model correctly predicted 429 instances for class 0, 484 for class 1, and 483 for class 2—demonstrating strong classification accuracy for each category. Only a small number of misclassifications occurred, with just two false positives and two false negatives recorded. These minimal errors suggest that the model performs with a high degree of precision and reliability in distinguishing between the three classes.

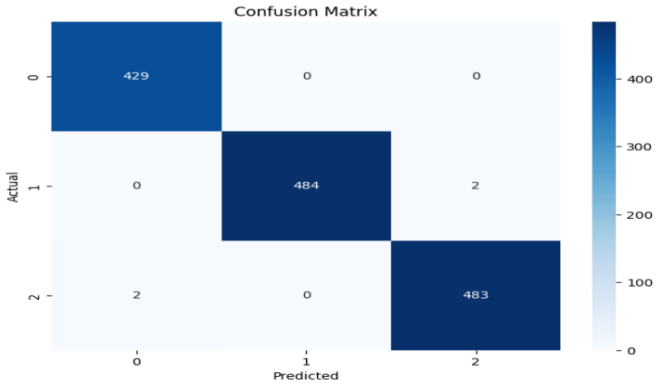

Fig 6: Metrics-Confusion Metrix For random Forest

Key Observations:

The confusion matrix offers a comprehensive view of the Decision Tree model's classification performance across three classes. The model accurately predicted 429 instances for class 0, 484 for class 1, and 483 for class 2, indicating strong performance across all categories. Misclassifications were minimal, with only

two false positives and two false negatives reported. These low error rates reflect the model's high precision and reliability, demonstrating its effectiveness in accurately distinguishing between the three target class.

Discussion

This paper examines machine learning methods for cyber threat detection in Industrial Internet of Things (IIoT) systems, especially focusing on Random Forest. During the evaluation, Random Forest has reached impressive accuracy of 99.7%, which efficiently identifies complex patterns in the dataset. The model's performance manifests high detection accuracy and is therefore suitable for the real-time monitoring of IIoT environments that are constantly exposed to serious cyber threats.

The highlighted characteristic in the performance of random forests is the ensemble approach whereby multiple decision trees are constructed and combined for making predictions that are very accurate and reliable. This has the feature of upliftment in the reliability of predictions but it also tends to avoid overfitting, which is often common in machine learning, ensuring better performance of the model over new unseen data. Thus, the Random Forest algorithm is an efficient solution to cyber defense upliftment in an IIoT environment by offering better threat detection accuracy and consistency.

This paper candidly presents the benefits of practically upscaling the Random Forest algorithm with the Synthetic Minority Over-sampling Technique (SMOTE) for cyber-terrestrial threat detection on handling the data imbalance problems. The enhancement is done through SMOTE, by generating synthetic examples of the minority classes, which enable the model to recognize the subtle signature of attack patterns at all levels of threat diversity. Conclusively, in combining these two techniques, the security system promises much more reliability and coverage at a lower false-negative ratio. As industrial environments become ever more connected with the IIoT, there will be a new pressure on their secure operation. It is now agreed that intelligent defense systems, such as the one presented in this study, can effectively secure and adapt to evolving cyber threats with greater precision and consistency by problems being encountered by their clients.

Conclusion

The study has confirmed Random Forest as the best solution to detect security threats and anomalies in IIoT infrastructure systems. The model achieves excellent abilities to detect intricate dataset patterns with 99.7% accuracy which enables it to uncover vital cyber threats from subtle alterations in the data. Random Forest achieves great generalization capabilities through multiple decision trees that combine forces to reduce false alerts and security breaches which makes it useful for industrial safety operations requiring uninterrupted uptime.

Random Forest showcases superiority as an evaluation model due to its enhanced stability together with increased precision rates. The machine learning solution needs additional improvements that can appear through ongoing parameter adjustments and instant data stream monitoring. The implementation of SMOTE enabled the model to locate rare instances of minority classes that indicate cyber threats through improved handling of class imbalances. A system's data characteristics determine what kind of proper algorithms must be implemented according to the research findings. SmartSentry which relies on Random Forest technology delivers organizations a smart protective system that defends IIoT environments from cyberattacks.

Future Enhancement

Consequently, the analysis of the different models of machine learning shows that there is a significant difference in the classification results. The Decision

Tree and Random Forest models topped the performance list with accuracies of 99.7% proving that the models can learn intricate patterns and relationship set in the data set. These ensemble techniques use the results of multiple decision trees to improve the predictive power, which makes such trees optimal for problems, where the influence of overfitting is critical.