

Anomaly Detection in Industrial IOT Sensor Data in MATLAB

Ms.Somoshi Kavita Mahendra.*¹, Prof. R. S. Bansode.*²

***¹** ME Student, Department Of Electronics and Telecommunication Engineering g (VLSI & Embedded System), SharadChandra Pawar College of Engineering, Otur, India.

***²** Project Guide, Professor, Sharadchandra Pawar College Of Engineering, Otur, India.

ABSTRACT

The rapid adoption of the Internet of Things (IoT) in industrial environments has transformed conventional monitoring systems into intelligent and data-driven infrastructures. Industrial IoT (IIoT) sensors continuously generate large volumes of operational data, including temperature, pressure, vibration, and humidity measurements, enabling real-time supervision of industrial equipment and processes. However, manually identifying abnormal behavior within such high-dimensional and continuously streaming data is both time-consuming and inefficient. Therefore, the integration of machine learning techniques for automated anomaly detection has become essential for ensuring operational reliability and minimizing equipment failures. This research presents a comparative analysis of two supervised machine learning algorithms—Support Vector Classifier (SVC) and Random Forest—for anomaly detection in Industrial IoT sensor data using the MATLAB environment. Initially, the collected sensor dataset undergoes preprocessing procedures such as data cleaning, handling missing values, normalization, and feature scaling to improve model performance and reliability. The processed dataset is subsequently divided into training and testing subsets for effective model evaluation. The Support Vector Classifier model utilizes a Radial Basis Function (RBF) kernel to identify complex nonlinear relationships within sensor data and classify operational states as either normal or anomalous. In parallel, the Random Forest algorithm employs an ensemble learning strategy by integrating multiple decision trees to enhance classification accuracy and improve robustness against noisy and inconsistent sensor readings. The performance of both models is assessed using evaluation metrics including accuracy score, confusion matrix, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curve analysis. Experimental findings indicate that both machine learning models effectively detect anomalies in industrial sensor datasets. However, the Random Forest classifier demonstrates comparatively superior performance due to its enhanced generalization capability, resistance to overfitting, and efficient handling of multidimensional data. The proposed approach highlights the practical significance of machine learning-based anomaly detection in Industrial IoT systems for predictive maintenance, fault diagnosis, operational safety, and reduction of unplanned industrial downtime. This study confirms that intelligent anomaly detection frameworks can substantially improve the efficiency, reliability, and sustainability of modern industrial operations.

I. INTRODUCTION

1.1 Overview

The Industrial Internet of Things (IIoT) has transformed traditional industries into smart and interconnected systems through the use of intelligent sensors and data-driven technologies. These sensors continuously monitor parameters such as temperature, pressure, vibration, humidity, current, and voltage, generating large volumes of operational data. Analyzing this data helps industries improve equipment reliability, operational efficiency, and system performance. One major challenge in IIoT environments is detecting anomalies or abnormal patterns in sensor data. Such anomalies may occur due to equipment faults, sensor failures, environmental changes, or

human errors. Early anomaly detection is essential to prevent machine damage, reduce downtime, and enhance industrial safety. Traditional threshold-based monitoring systems are often inadequate for handling complex and nonlinear industrial data. To overcome these limitations, this project proposes an anomaly detection system using machine learning techniques in MATLAB. The system utilizes Support Vector Classifier (SVC) and Random Forest algorithms to analyze historical IIoT sensor datasets and classify system behavior as normal or abnormal. The SVC model identifies complex nonlinear patterns, while the Random Forest model improves detection accuracy through ensemble learning. The project demonstrates how machine learning-based anomaly detection can support predictive maintenance, improve industrial reliability, and enable intelligent decision-making in modern IIoT systems.

1.2 Introduction

The Industrial Internet of Things (IIoT) is a major component of Industry 4.0, enabling industries to collect and analyze real-time data through interconnected sensors and devices. Modern industrial systems continuously generate sensor data related to temperature, pressure, vibration, voltage, and other operational parameters. This data provides valuable insights into equipment condition and production performance. Detecting anomalies within this large amount of data is critical for maintaining system reliability and preventing failures. Anomalies may indicate equipment degradation, sensor malfunction, or unexpected operational behavior. Traditional rule-based monitoring systems are limited because they cannot effectively handle dynamic and nonlinear relationships among multiple sensor variables. Machine learning techniques offer a more intelligent and adaptive solution for anomaly detection. By learning from historical sensor data, machine learning models can identify hidden patterns and automatically detect abnormal conditions. This project focuses on implementing an anomaly detection system using Support Vector Classifier (SVC) and Random Forest algorithms in MATLAB. Historical IIoT sensor data is preprocessed, normalized, and used to train and test the models. The SVC algorithm is effective for nonlinear classification, while Random Forest provides robustness and higher accuracy through multiple decision trees. The proposed system aims to improve predictive maintenance, reduce operational downtime, and enhance industrial safety through efficient anomaly detection.

1.3 History and Background

Anomaly detection in industrial systems originated from traditional fault monitoring techniques used during the early stages of industrial automation. Initially, industries depended on manual observation and simple threshold-based alarms to identify abnormal machine behavior. Although effective for basic systems, these approaches became insufficient as industrial processes grew more complex. During the 1990s, statistical methods such as Statistical Process Control (SPC) and Principal Component Analysis (PCA) were introduced to improve fault detection capabilities. Later, the emergence of IoT and cloud computing enabled industries to deploy large-scale sensor networks capable of real-time monitoring and data collection. The rapid growth of IIoT generated massive amounts of sensor data, creating the need for intelligent data analysis techniques. As a result, machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, Neural Networks, and Random Forest gained popularity in industrial applications. Support Vector Classifier (SVC) became widely used for accurate classification of normal and faulty conditions, especially in nonlinear datasets. Random Forest further improved detection performance by combining multiple decision trees to increase robustness and reduce overfitting. Today, machine learning-based anomaly detection plays an important role in predictive maintenance and smart manufacturing by helping industries detect faults early, improve reliability, and reduce maintenance costs.

1.4 Theme

The main theme of this project is the development of an intelligent anomaly detection system for Industrial IoT environments using machine learning algorithms. The project focuses on analyzing historical sensor data using Support Vector Classifier (SVC) and Random Forest models to automatically identify abnormal conditions. The system replaces traditional rule-based monitoring methods with data-driven and self-learning techniques capable of handling complex industrial data patterns. This approach supports predictive maintenance, reduces equipment downtime, and improves operational efficiency. The project also aligns with Industry 4.0 concepts by integrating machine learning, industrial automation, and data analytics to create smarter and more reliable industrial systems.

1.5 Objectives

1. To develop an anomaly detection system using SVC and Random Forest algorithms in MATLAB.
2. To use historical IIoT sensor datasets for training and testing machine learning models.
3. To preprocess sensor data through cleaning, normalization, and feature extraction techniques.
4. To classify industrial data into normal and abnormal operational conditions.
5. To evaluate model performance using accuracy, precision, recall, and F1-score.
6. To compare the performance of SVC and Random Forest algorithms for anomaly detection.
7. To support predictive maintenance by enabling early fault detection.
8. To improve industrial reliability, safety, and operational efficiency using machine learning techniques.

II. SYSTEM DEVELOPMENT

Data Source and Pre-processing

In this project, historical Industrial IoT (IIoT) sensor data is used for anomaly detection. The dataset contains multivariate sensor readings collected from industrial equipment over time. Parameters such as temperature, pressure, vibration, humidity, and rotational speed are included, as they are important indicators of machine condition and system performance.

The data is imported into MATLAB using database connectivity tools or through CSV/Excel files. Before training the machine learning models, the dataset undergoes several preprocessing steps to improve data quality and model performance.

Pre-processing Steps

1. Data Cleaning - Missing values, duplicate records, and inconsistent sensor readings are removed from the dataset. Faulty readings caused by communication issues or sensor errors are filtered out. Missing data is handled using mean substitution or interpolation techniques.

2. Noise Reduction - Industrial sensor data may contain noise due to environmental disturbances or electrical interference. Smoothing techniques such as moving average filtering are applied to reduce noise and improve signal quality.

3. Normalization - Sensor parameters are scaled to a common range to eliminate bias caused by different measurement units. Normalization improves the efficiency and accuracy of machine learning algorithms.

4. Feature Extraction - Important statistical features such as mean, variance, standard deviation, skewness, and kurtosis are extracted from raw sensor data. These features help represent system behavior more effectively for anomaly detection.

After preprocessing, the dataset is divided into training and testing sets for model development and performance evaluation.

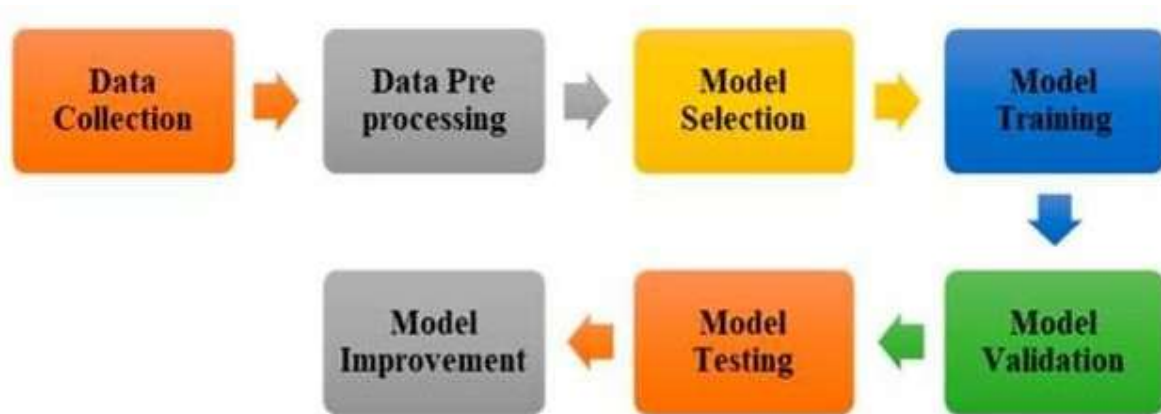


Fig.1.1 Proposed Anomaly detection method

Model 1: Support Vector Classifier (SVC)

The Support Vector Classifier (SVC) is a supervised machine learning algorithm used to classify data into normal and anomalous categories. It creates an optimal decision boundary that separates different classes with maximum margin.

Implementation Steps

1. Import preprocessed sensor data into MATLAB.
2. Train the SVC model using the `fitcsvm()` function with an RBF kernel to handle nonlinear patterns.
3. Optimize parameters such as kernel scale and box constraint for better accuracy.
4. Test the model using unseen data and evaluate performance using accuracy, precision, recall, and F1-score.
5. Visualize classification results using MATLAB plots and confusion matrices.

The SVC model performs effectively for nonlinear industrial datasets where abnormal conditions are difficult to separate using traditional methods.

Model 2: Random Forest Classifier

Random Forest is an ensemble machine learning algorithm that combines multiple decision trees to improve classification accuracy and robustness. It is highly effective for handling noisy and high-dimensional industrial data.

Implementation Steps

1. Train the Random Forest model using MATLAB's `TreeBagger()` or `fitcensemble()` functions.
2. Generate multiple decision trees using random subsets of features and data samples.
3. Evaluate the model using confusion matrix, accuracy, precision, recall, and F1-score.
4. Analyze feature importance to identify the most influential sensor parameters.
5. Visualize prediction results and feature contributions using graphs and plots.

The Random Forest model provides strong generalization capability and reduces overfitting, making it suitable for industrial anomaly detection applications.

Model Comparison

Both SVC and Random Forest models are evaluated and compared based on detection accuracy, robustness, and computational performance. The comparison helps identify the most suitable algorithm for anomaly detection in Industrial IoT sensor data.

III. FLOW CHART & ALGORITHM

MATLAB Workflow

The project implementation follows a systematic workflow in MATLAB for efficient anomaly detection using machine learning techniques.

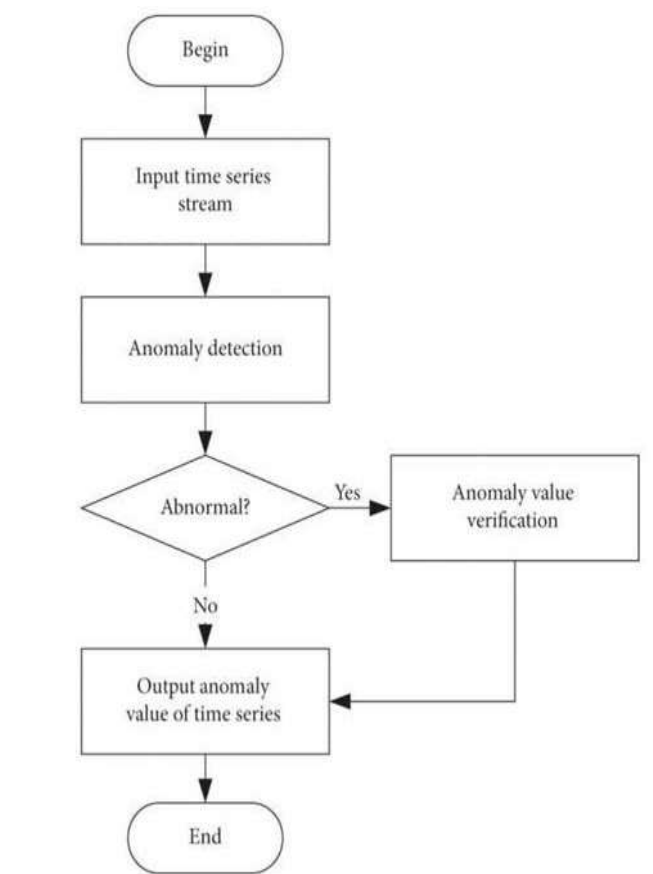
1. Data Import - Historical IIoT sensor data is imported into MATLAB from databases, CSV files, or Excel sheets and converted into a suitable format for analysis.
2. Data Pre-processing - The dataset undergoes preprocessing steps such as data cleaning, normalization, noise reduction, and feature extraction to improve data quality and model performance.
3. Model Training - The preprocessed dataset is used to train both Support Vector Classifier (SVC) and Random Forest models for anomaly detection.
4. Testing and Validation - The trained models are tested using unseen data to evaluate their ability to correctly identify normal and anomalous conditions.

5. Performance Comparison - The performance of SVC and Random Forest classifiers is compared using metrics such as accuracy, precision, recall, F1-score, and confusion matrix analysis.

6. Visualization and Reporting - MATLAB visualization tools are used to display classification results, anomaly patterns, feature importance, and model performance graphs. These visual outputs assist in understanding system behavior and support effective industrial maintenance decisions.

Algorithm

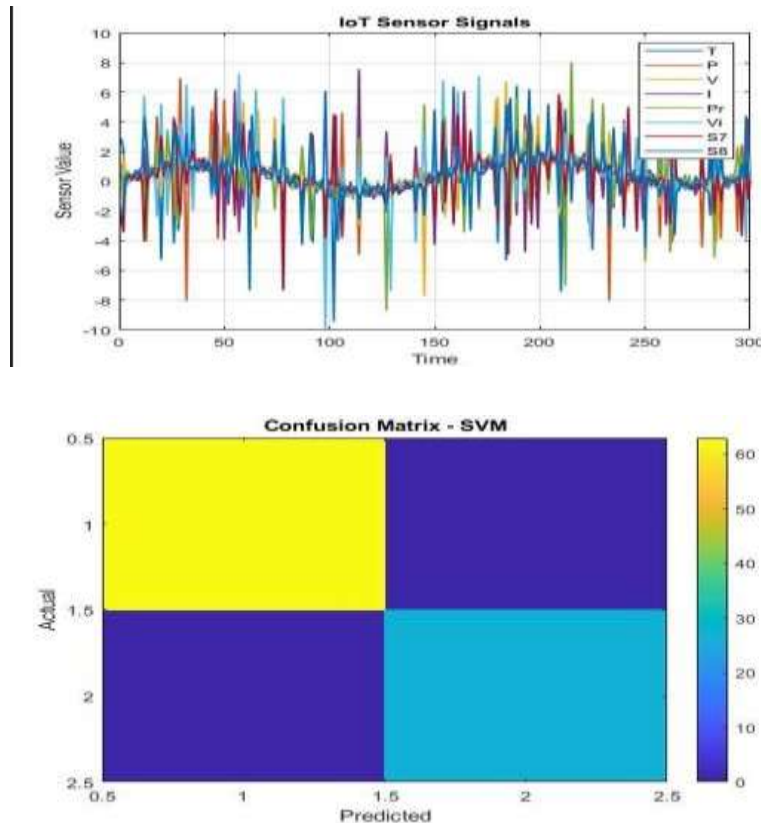
1. Begin
2. Input: Receive the time series stream.
3. Detect: Run the anomaly detection process.
4. Check: IF an abnormality is detected:
 - a. Proceed to Anomaly value verification.
- b. After verification, jump to Output. ELSE (No abnormality): a. Proceed directly to Output.
5. Output: Provide the final (and verified) anomaly value (or a "no anomaly found" status).
6. End



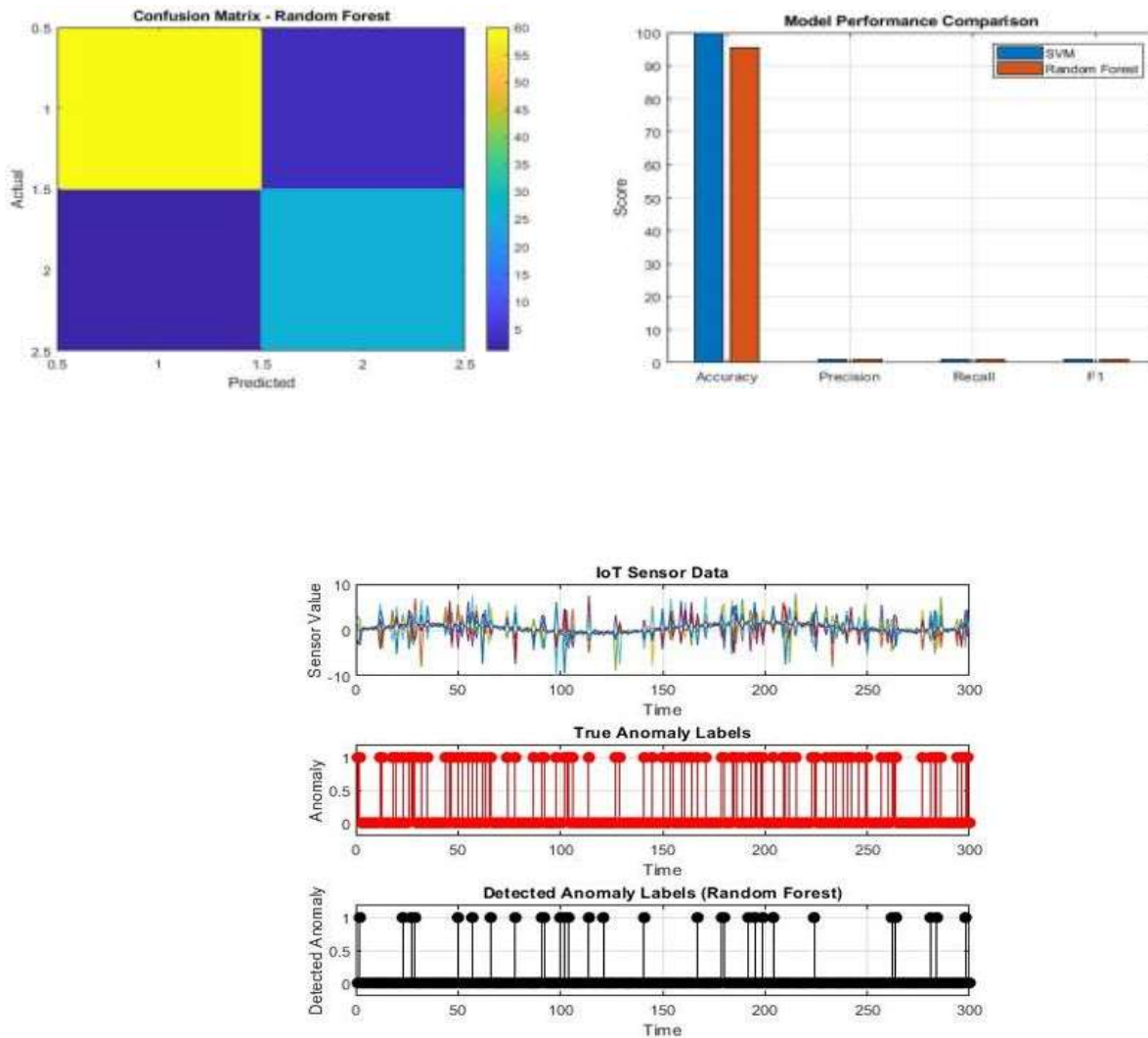
IV. RESULTS

The experimental results demonstrate the effectiveness of machine learning techniques in detecting anomalies within Industrial IoT sensor data. The collected sensor signals exhibited both normal operational patterns and abnormal fluctuations across parameters such as temperature, pressure, vibration, and voltage. These variations indicate the presence of potential anomalous conditions in the industrial environment.

The confusion matrices generated for both the Support Vector Machine (SVM) and Random Forest models show strong classification performance with a high number of correctly predicted normal and anomalous instances. Both models achieved high detection accuracy with very few misclassifications, confirming their capability to distinguish between normal and faulty operational states. From the performance comparison graph, the SVM model achieved slightly higher overall accuracy, precision, recall, and F1-score compared to the Random Forest model. This indicates that the SVM classifier was more effective in identifying nonlinear patterns within the sensor dataset. However, the Random Forest model also delivered reliable results with strong robustness and stable predictions due to its ensemble learning mechanism.



The anomaly detection visualization further confirms that the trained models successfully identified abnormal events present in the sensor data. The detected anomaly labels closely matched the true anomaly labels, demonstrating the reliability of the proposed system for industrial fault monitoring and predictive maintenance applications. Overall, the results validate that machine learning-based anomaly detection using SVM and Random Forest in MATLAB provides an efficient and accurate solution for monitoring Industrial IoT systems. The proposed approach can help industries reduce unexpected equipment failures, minimize downtime, and improve operational safety and reliability.



V. CONCLUSION

The project, “Anomaly Detection in Industrial IoT Sensor Data Using SVC and Random Forest in MATLAB,” presents an effective machine learning-based approach for identifying abnormal behavior in industrial systems using historical IIoT sensor data. The system analyzes important operational parameters such as temperature, pressure, vibration, and humidity to detect potential anomalies and equipment faults. Both Support Vector Classifier (SVC) and Random Forest algorithms successfully classified normal and abnormal operating conditions. The SVC model effectively handled nonlinear data patterns, while the Random Forest model provided higher robustness and improved prediction accuracy through ensemble learning. Data preprocessing techniques including cleaning, normalization, noise reduction, and feature extraction significantly improved the quality of the dataset and enhanced model performance. Evaluation results based on accuracy, precision, recall, and F1-score demonstrated the reliability of the proposed anomaly detection system.

Overall, the project proves that machine learning techniques can greatly improve predictive maintenance, fault detection, and operational safety in Industrial IoT environments. By utilizing historical sensor data, industries can minimize downtime, reduce maintenance costs, and improve overall system efficiency and reliability.

VI. REFERENCES

- [1] Min Yang, Jiajie Zhang, —Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges,|| Department of Information Engineering, Shandong Communication & Media College, China.
- [2] A. Mamatha, Ch. Poojitha, T. Maruthi Megansh, K. Shiva Kumar, Amol Rathod, —Machine Learning Based Anomaly Detection in IoT Sensor Data,|| Sree Dattha Group of Institutions, Hyderabad, India.
- [3] Muhammad Fahim, Alberto Sillitti, —Anomaly Detection, Analysis, and Prediction Techniques in IoT Environment: A Systematic Literature Review,|| Innopolis University, Russia.
- [4] Shibzan Shahanas, Afnaj Akthar, Saanna Anand, Rakshitha, Dr. Amirthavalli M, —Anomaly Detection in Time Series Data in IoT Environment,|| Mangalore Institute of Technology & Engineering, India.
- [5] Ayan Chatterjee, Bestoun S. Ahmed, —IoT Anomaly Detection Methods and Applications: A Survey,|| Karlstad University, Sweden.
- [6] Kyle DeMedeiros, Abdeltawab Hendawi, Marco Alvarez, —A Survey of AI-Based Anomaly Detection in IoT and Sensor Networks,|| University of Rhode Island, USA.