# Anomaly Detection in Industrial IoT Sensor Data in Matlab

**Kavita Mahendra Somoshi, Prof.Bansode Rahul S.**

**Abstract- In modern industrial environments, the Internet of Things (IoT) has become essential for real-time monitoring of machines and processes through intelligent sensors. These sensors continuously collect operational data such as temperature, pressure, and vibration, help us to ensure smooth, safe, and efficient system performance. However, the massive volume of generated data makes it challenging to manually identify unusual patterns or system faults. To address this issue, this project focuses on detecting anomalies in Industrial IoT sensor data using two machine learning approaches: Support Vector Classifier (SVC) and Random Forest, implemented in MATLAB. The collected data is first preprocessed by cleaning, normalizing, and splitting it into training and testing datasets. The SVC model with a Radial Basis Function (RBF) kernel is used to capture complex nonlinear relationships and classify data points as either normal or abnormal. In parallel, the Random Forest algorithm, which is based on an ensemble of multiple decision trees, is employed to enhance prediction accuracy and effectively manage noisy or inconsistent data. Both models are evaluated using performance metrics such as accuracy and confusion matrix analysis. The results indicate that both techniques are effective in identifying anomalies in industrial sensor data, with Random Forest showing slightly superior performance due to its ensemble-based learning strategy and strong generalization capability. Overall, the study highlights that applying machine learning techniques in Industrial IoT environments can greatly improve predictive maintenance, minimize downtime, and enhance the safety and reliability of industrial operations.**

**Index Terms— anomaly ,SVC, random forest.**

## I. INTRODUCTION

The rapid evolution of digital technologies has transformed modern industries into highly interconnected, intelligent, and automated ecosystems. Among the most significant technological advancements driving this transformation is the Industrial Internet of Things (IIoT), a specialized branch of the Internet of Things (IoT) tailored specifically for
industrial environments. IIoT integrates industrial machines, sensors, communication networks, cloud computing, and intelligent analytics into a unified framework capable of

monitoring as well as optimizing industrial operations in real time.

In modern industrial systems, thousands of interconnected sensors are deployed across manufacturing plants, power systems, oil refineries, transportation systems, chemical

processing units, and smart factories. These sensor continuously measure and transmit operational parameters such as temperature, pressure, humidity, vibration, rotational speed, current, voltage, torque, flow rate, gas concentration, and machine utilization metrics. The collected sensor readings generate enormous volumes of time-series data every second, creating a rich source of information that reflects the real-time health, performance, and operational state of industrial assets. The ability to collect massive amounts of operational data has opened new opportunities for

industries to improve productivity, efficiency, reliability, and safety. However, the challenge lies not in collecting data but in effectively analyzing it to extract meaningful insights.

Industrial environments are dynamic and highly complex, involving numerous interconnected subsystems where a small deviation in one component may affect the performance of the entire system. As industries become increasingly automated and interconnected, ensures that the continuous and reliable operation of machinery becomes important. One of the most important tasks in industrial monitoring systems is anomaly detection. An anomaly refers to unusual pattern irregular behavior of system, or deviation from normal operational conditions that may indicate faults, equipment degradation, cyber attacks, sensor failures, process instability, or unsafe working conditions. In industrial systems, anomalies can occur due to multiple reasons, including mechanical wear and tear, overheating,
lubrication failure, electrical faults, environmental disturbances, process drift, communication errors, calibration issues, or human mistakes. Early detection of such anomalies is extremely important because even minor abnormalities

can gradually evolve into severe failures if left unnoticed. Unexpected equipment breakdowns can result in production downtime, financial losses, reduced product quality, environmental hazards, and safety risks for workers. In critical industries such as energy generation, healthcare, system failures can have catastrophic consequences. Therefore, developing intelligent systems capable of identifying anomalies accurately and at an early stage has become a major research and industrial priority. Traditionally, industrial anomaly detection relied on manual inspection methods, threshold based monitoring systems, and rule-based approaches. Threshold-based systems operate by defining upper and lower limits for sensor parameters. If a sensor reading exceeds these predefined boundaries, an alarm is triggered. Rule-based systems similarly depend on manually programmed conditions created by domain experts. While such approaches are simple and easy to implement, they suffer from several limitations. Industrial data is highly nonlinear, multidimensional, and time-dependent. Different

machine parameters often interact with one another in complex ways that cannot be fully captured using fixed thresholds or handcrafted rules. For example, a rise in temperature may be normal under certain operating loads but abnormal under others. Similarly, vibration patterns may vary depending on machine speed, environmental conditions, and production cycles. Traditional systems struggle to adapt to these dynamic relationships and often generate excessive false alarms or fail to detect subtle faults. Machine learning-based anomaly detection offers several significant advantages over traditional methods. These techniques can handle high-dimensional datasets, capture nonlinear relationships, process multivariate sensor inputs, and identify hidden anomalies that may not be detectable using simple threshold rules. ML algorithms also enable predictive maintenance strategies by identifying early warning signs of equipment degradation before catastrophic failures occur.

Predictive maintenance has become one of the most valuable applications of IIoT and machine learning in Industry 4.0. Unlike reactive maintenance, where repairs occur only after failure, or preventive maintenance, where servicing is scheduled at fixed intervals, predictive maintenance uses real-time and historical data analysis to estimate the actual health condition of equipment. This approach reduces maintenance costs, minimizes downtime, extends machine lifespan, and improves operational efficiency.

## II. HISTORY AND BACKGROUND

The development of anomaly detection can be traced to the early phase of industrial automation and fault diagnosis in the late 20th century. In its initial stage, industries depended heavily on manual observation, where engineers visually inspected machine behavior and instrument readings or relied on basic alarm systems. Gradually, rule-based logic and fixed threshold monitoring became common in condition monitoring systems. For instance, vibration sensors would generate alerts when signal levels crossed predefined limits, often indicating issues such as bearing failures.

As industrial environments grew more complex and sensor data volumes increased, these simple methods became inadequate. They were unable to capture relationships among multiple variables, environmental variations, and temporal dependencies. During the 1990s, more advanced statistical approaches such as Statistical Process Control (SPC), Principal Component Analysis (PCA), and Auto-Regressive models were introduced to improve monitoring accuracy and fault detection capability.

With the rise of Internet of Things (IoT) technologies in the early 21st century, a major transformation occurred. Improvements in communication networks, cloud infrastructure, and sensor miniaturization enabled the deployment of large-scale interconnected sensor systems. This led to the emergence of the Industrial Internet of Things (IIoT), where machines, sensors, and analytics platforms work together to enable continuous monitoring and intelligent industrial operations.

Despite these advancements, IIoT systems generated massive and complex datasets, creating challenges in extracting meaningful patterns from high-dimensional and often unstructured data. To address this, machine learning techniques became increasingly important. Methods such as Support Vector Machines (SVM), Decision Trees, Neural Networks, and Random Forests began to play a key role in fault detection, predictive maintenance, and quality control tasks.

In particular, the Support Vector Classifier (SVC), a supervised learning technique grounded in statistical learning theory, became widely used for binary classification problems such as distinguishing between normal and faulty operating conditions. Similarly, Random Forest, an ensemble-based approach, improved model stability and performance by combining multiple decision trees, making it more effective in handling noisy and incomplete datasets.

Over the past decade, the integration of IIoT data with machine learning analytics has significantly advanced industrial systems. Anomaly detection has shifted from a reactive approach—identifying faults after they occur—to a predictive strategy focused on preventing failures in advance. This evolution forms the basis of modern intelligent industrial systems and supports the present project, which combines traditional and modern machine learning methods to improve reliability, efficiency, and operational intelligence.

## III. THEME

The core focus of this project is to design a data-driven anomaly detection system for industrial settings using machine learning techniques. It investigates how historical Industrial IoT (IIoT) sensor data can be processed using models such as Support Vector Classifier (SVC) and Random Forest to automatically detect unusual or faulty conditions without relying on manual monitoring or fixed threshold values.

The main idea is to move away from traditional rule-based monitoring systems toward intelligent systems that can learn from data and understand complex relationships between multiple sensor readings. By learning patterns from past “normal” operating conditions, these models are able to identify deviations that may signal potential equipment failures or abnormalities.

This work is also closely connected to the concepts of Industry 4.0 and smart manufacturing, where automation, machine learning, and data analytics are integrated to support real-time monitoring and faster decision-making. Using MATLAB enhances this process by providing strong support for machine learning toolboxes, data visualization, and database handling, enabling a smooth workflow from data preparation to model training, testing, and evaluation.

Overall, the project combines machine learning, industrial automation, and data analytics to improve system reliability,

reduce maintenance costs, and support more efficient and informed industrial operations.

## IV. OBJECTIVES

The primary objectives of this project are defined to balance both technical implementation and real-world industrial applicability within an IIoT environment:

1. To develop a robust anomaly detection system using Support Vector Classifier (SVC) and Random Forest algorithms implemented in MATLAB, ensuring effective model design and deployment.
2. To work with pre-collected historical IIoT sensor datasets stored in centralized databases, using them for model training, testing, and validation instead of relying on real-time sensor data streams.
3. To carry out comprehensive data preprocessing steps, including data cleaning, normalization, and feature extraction, to transform raw sensor readings into a suitable format for machine learning models.
4. To train machine learning models capable of accurately distinguishing between normal operational behavior and abnormal or faulty conditions using labeled historical data.
5. To assess the performance of the developed models using evaluation metrics such as accuracy, precision, recall, and F1-score, ensuring dependable and consistent anomaly detection results.
6. To conduct a comparative analysis between SVC and Random Forest models based on detection accuracy, computational efficiency, and resilience to noisy or imperfect data.
7. To build a conceptual foundation for predictive maintenance systems that can identify potential faults at an early stage and support preventive decision-making to minimize system downtime.
8. To highlight the effectiveness of machine learning–based anomaly detection techniques in real industrial environments, contributing to better process monitoring, control, and operational optimization.

## V. SYSTEM DEVELOPMENT

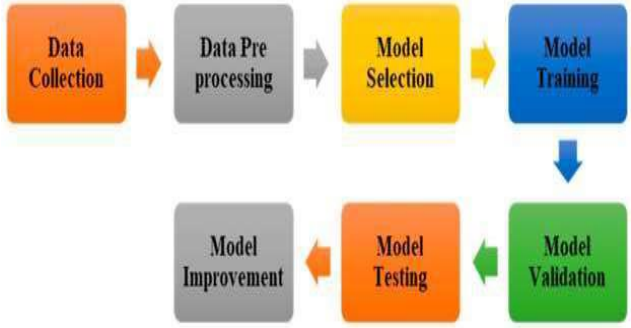



Fig. Proposed Anomaly detection method

**Data Source and Pre-processing**

In this project, Industrial Internet of Things (IIoT) sensor data is utilized as the primary input for developing an anomaly detection system. The dataset is collected from a historical industrial monitoring database consisting of multivariate time-series readings obtained from different machines and equipment over time. Each data record captures the real-time operational condition of the system at a specific timestamp. The dataset includes some sensor parameters such as temperature, pressure, vibration , humidity. which reflect the health status and performance behavior of industrial machinery.

The data is imported into MATLAB using database connectivity features or by loading structured files in CSV or Excel format. After loading, the dataset undergoes systematic pre-processing to improve data quality and make it suitable for machine learning-based analysis.

Pre-processing Steps

Data Cleaning:

The raw dataset often contains missing entries, duplicate records, and inconsistent sensor outputs due to communication delays or hardware disturbances. These issues are handled by removing invalid records and replacing missing values using statistical methods such as mean, median substitution, or interpolation techniques.

Noise Reduction:

Industrial sensor signals are frequently affected by unwanted noise caused by electrical interference and environmental variations. To improve data reliability, smoothing techniques like moving average filters or Savitzky-Golay filtering are applied to reduce fluctuations and preserve meaningful trends.

Feature Scaling (Normalization):

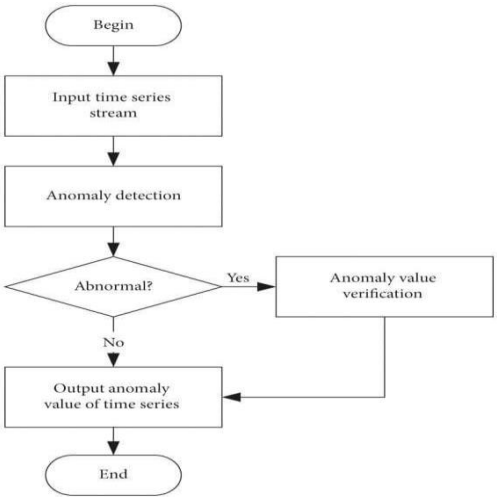
Since sensor values operate on different scales, normalization is applied to convert all features into a uniform range. This prevents any single parameter from dominating the learning process and improves the stability and convergence of machine learning models.

Feature Engineering:

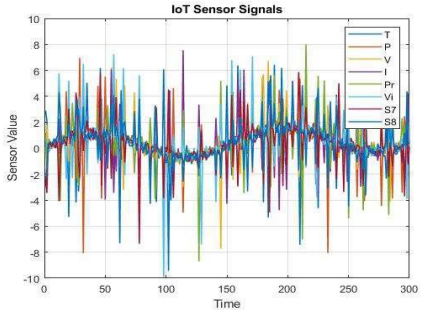
Raw sensor signals are transformed into meaningful statistical representations over time windows. Features such as mean, variance, standard deviation, skewness, kurtosis, and signal energy are extracted to capture hidden patterns in machine behavior. These engineered features form the final dataset used for model training.

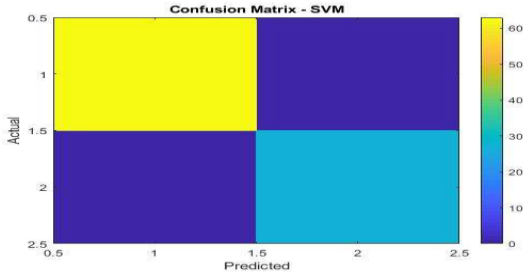
After completing pre-processing, the dataset is split into training and testing subsets to ensure unbiased evaluation. The training set is used to build machine learning models, while the testing set is used to evaluate their performance on unseen data.

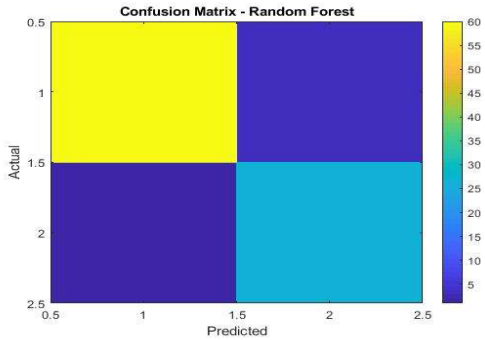
Flowchart:

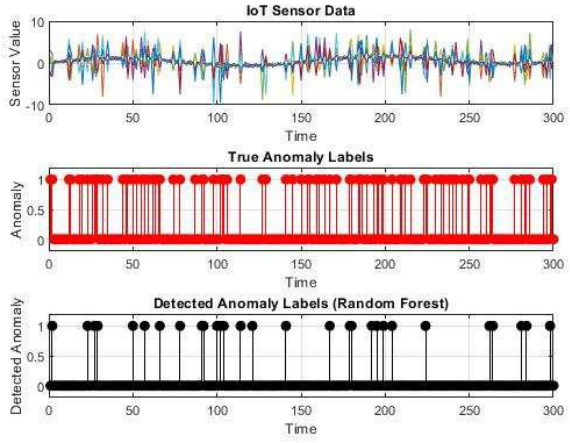

Begin
Input time series stream
Anomaly detection
Abnormal?
Yes
Anomaly value verification
No
Output anomaly value of time series
End

## RESULT


IoT Sensor Signals
Sensor Value
Time
T
P
V
I
Pr
Vi
S7
S8


Confusion Matrix - SVM
Actual
Predicted


Confusion Matrix - Random Forest
Actual
Predicted




## VI. .CONCLUSION

This project successfully demonstrates the use of machine learning techniques for anomaly detection in Industrial IoT sensor data. By applying SVC and Random Forest models on pre-processed historical sensor data, the system effectively identifies abnormal patterns in critical industrial parameters. The results show that combining strong preprocessing with robust classification models improves detection accuracy and reliability. Overall, the approach supports predictive maintenance and helps industries reduce failures, minimize downtime, and improve operational efficiency through data-driven decision-making.

## 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.