**MINISTRY OF EDUCATION AND TRAINING**

**FPT UNIVERSITY**

Anomaly Detection for Vietnam Railway using Unsupervised Learning based on IoT Device Data Monitoring Railway Level Crossings

by

Ngo Anh Tuan

A thesis submitted in conformity with the requirements  
for the degree of Master of Software Engineering

© Copyright by Ngo Anh Tuan 2025

**MINISTRY OF EDUCATION AND TRAINING**

**FPT UNIVERSITY**

Anomaly Detection for Vietnam Railway using Unsupervised Learning based on IoT Device Data Monitoring Railway Level Crossings

by

Ngo Anh Tuan

A thesis submitted in conformity with the requirements  
for the degree of Master of Software Engineering

Supervisor:

Assoc. Prof. Phan Duy Hung

© Copyright by Ngo Anh Tuan 2025

Anomaly Detection for Vietnam Railway using Unsupervised Learning based on IoT Device Data Monitoring Railway Level Crossings

Ngo Anh Tuan

Degree Master of Software Engineering

FPT University

2025

Abstract

In Vietnam, the railway industry has always been of interest to society, with its scale constantly expanding and upgrading. According to statistics, there are approximately 10,000 railway level crossings across the nation. At railway level crossings, system of protective equipment is deployed to ensure traffic safety. The operational data of these protective devices is collected and transmitted to the data center for monitoring and management. We developed and evaluated machine learning (ML) models to detect potential anomalies in crossing operations and identify potential safety risks. The system analyzes real-time sensor data including magnetic sensors, barrier status monitors, power supply indicators, and environmental sensors from multiple crossing locations. By using one class support vector machine (OCSVM) models to exploit the collected data, with the dual objectives of early detection of potential risks as well as the stable operation of the equipment we achieved 96% accuracy. This research proposes a methodology for improving the quality of management as well as the application of science and technology in the digital transformation and modernization of the railway industry in Vietnam.

Keywords: Railway Safety, Unsupervised Learning, SVM One Class, Anomaly Detection.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Assoc. Prof. Phan Duy Hung, for his invaluable assistance, guidance, and insightful feedback throughout the development and writing of this thesis.

I am particularly grateful to all members of staff at the FTP School of Business & Technology for their kind support during my master study.

My sincere appreciation extends to Hanoi Railway Signal and Telecom, JSC for the datasets and support.

# List of Figures

[**Figure 1.** Infrastructure Architecture 7](#_Toc197327809)

[**Figure 2**. Overall architecture 19](#_Toc197327810)

[**Figure 3**. Train event data by time 21](#_Toc197327811)

[**Figure 4.** Magnetic sensors value by time 23](#_Toc197327812)

[**Figure** **5.** Explain of measure AD3 in data set with PCA 27](#_Toc197327813)

[**Figure 6**. PCA and t-SNE distribute 27](#_Toc197327814)

[**Figure 7.** Magnetic sensor values for a normal train event and a simulated faulty train event 31](#_Toc197327815)

# List of Table

[**Table 1.** The raw data example from railway level crossing 20](#_Toc195280049)

[**Table 2.** Key IoT Data Fields for Railway Crossing Monitoring 22](#_Toc195280050)

[**Table 3.** Features Vector detail 26](#_Toc195280051)

[**Table 4.** Summary of dataset 30](#_Toc195280052)

[**Table 5.** Anomaly types and implementation methods 30](#_Toc195280053)

[**Table 6.** OCSVM default hyper params metrics 33](#_Toc195280054)

[**Table 7.** Tuning OCSVM hyper params 33](#_Toc195280055)

[**Table 8.** Best metrics of OCSVM model 33](#_Toc195280056)

Table of Contents

[Acknowledgments 2](#_Toc195283920)

[List of Figures 3](#_Toc195283921)

[List of Table 4](#_Toc195283922)

[Table of Contents 5](#_Toc195283923)

[Chapter 1 Introduction 7](#_Toc195283924)

[1.1. Problem And Motivation 7](#_Toc195283925)

[1.2. Related Works 9](#_Toc195283926)

[1.3. Contribution 10](#_Toc195283927)

[1.4. Thesis structure 11](#_Toc195283928)

[Chapter 2 Background Study 12](#_Toc195283929)

[2.1. IOT 12](#_Toc195283930)

[2.1.1. Sensor Data Acquisition 12](#_Toc195283931)

[2.1.2. Data Transmission via network 12](#_Toc195283932)

[2.2. Data mining 13](#_Toc195283933)

[2.2.1. Visualization Data 13](#_Toc195283934)

[2.2.2. Data Preprocessing 13](#_Toc195283935)

[2.2.3. Feature Extraction and Engineering 14](#_Toc195283936)

[2.2.4. Temporal Pattern Analysis 14](#_Toc195283937)

[2.3. Unsupervised Learning Techniques 14](#_Toc195283938)

[2.3.1. Overview of Unsupervised Learning Methods 14](#_Toc195283939)

[2.3.2. One-class SVM modeling 15](#_Toc195283940)

[2.3.3. Hyperparameter Tuning 15](#_Toc195283941)

[2.3.4. Evaluation Metrics 16](#_Toc195283942)

[Chapter 3 System Design 19](#_Toc195283943)

[3.1. System Architecture 19](#_Toc195283944)

[3.2. Data Preprocessing 20](#_Toc195283945)

[3.2.1. Filter and clean data 20](#_Toc195283946)

[3.2.2. Feature Extraction 22](#_Toc195283947)

[3.3. Modeling Module 26](#_Toc195283948)

[3.3.1. Dimensionality Reduction 26](#_Toc195283949)

[3.3.2. Tuning model 28](#_Toc195283950)

[3.4. Realtime processing 28](#_Toc195283951)

[3.4.1. SlidingWindows 28](#_Toc195283952)

[3.4.2. DynamicTemplate 29](#_Toc195283953)

[Chapter 4 Experiments And Results 30](#_Toc195283954)

[4.1. Data collection 30](#_Toc195283955)

[4.2. Experiments 30](#_Toc195283956)

[4.3. Results 32](#_Toc195283957)

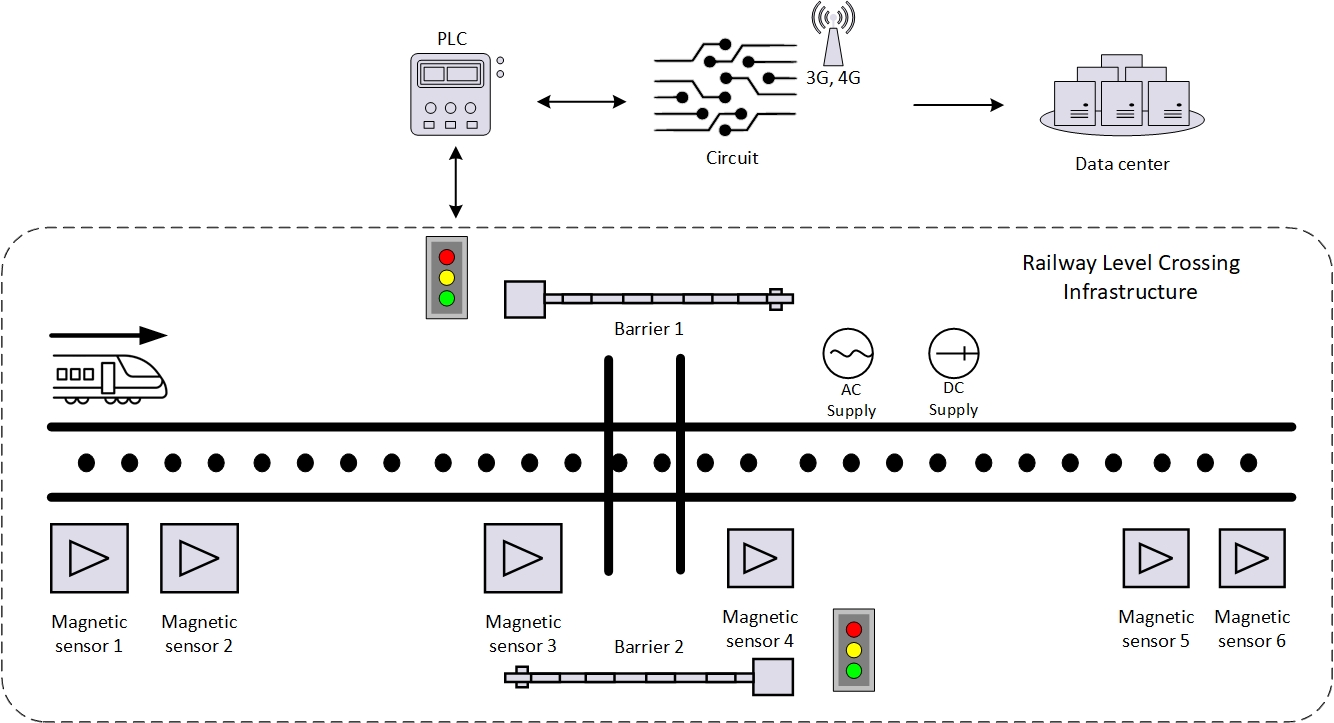
[4.4. Conclusion And Future Work 33](#_Toc195283958)

[References 34](#_Toc195283959)

# Introduction

## Problem And Motivation

At railway level crossings in Viet Nam, a comprehensive system of devices and sensors is deployed to ensure safety operations. The core infrastructure includes six magnetic sensors strategically positioned along the tracks, these magnetic sensors function by counting the pulses generated each time a train wheel axle passes over the sensor location, with two sensors placed one to five kilometers before the crossing point for early train detection, and two sensors at the intersection to confirm train presence. The system also incorporates barrier gates equipped with traffic light signals, powered by both AC and DC supplies to ensure operational reliability. The IoT infrastructure continuously collects various parameters from these devices.



**Figure 1.** Infrastructure Architecture

The crossing protection devices are connected to a PLC (Programmable Logic Controller), which controls the operation of the entire infrastructure. Sensor data is sent to the PLC, and circuits collect it through indirect interfaces with the PLC. The processed data is then transmitted to the server through a Wi-Fi module, which connects to 3G/4G access points for reliable data transmission to centralized data centers.

Despite the centralized collection of sensor data from railway level crossings, its current application is limited to displaying information at monitoring centers, with incident response decisions still heavily dependent on human operators. This manual approach faces significant challenges as monitoring centers are responsible for supervising numerous level crossings simultaneously. The high volume of real-time data from thousands of crossing points creates a substantial cognitive load for operators, potentially leading to delayed response times and increased risk of oversight in critical situations.

The monitoring of level crossings faces numerous challenges due to the inability to comprehensively define all safety-affecting scenarios. The railway system extends throughout Viet Nam with diverse operating conditions, and the message structure reveals over 40 different sensor parameters that must be monitored simultaneously. The combinations of these parameters create an extensive anomaly space, and many potentially dangerous situations have never occurred, making it impossible to pre-label them for traditional analysis methods.

The monitoring of level crossings faces numerous challenges due to the inability to comprehensively define all safety-affecting scenarios. The railway system extends throughout Viet Nam with diverse operating conditions, and the message structure reveals over 40 different sensor parameters that must be monitored simultaneously. The combinations of these parameters create an extensive anomaly space, and many potentially dangerous situations have never occurred, making it impossible to pre-label them for traditional analysis methods.

Furthermore, the data exhibits complex temporal relationships. Train detection follows a specific sequence (approaching, passing through, and departing the crossing) that must trigger synchronized responses from barrier movements, light signals, and sound alerts. Any desynchronization in this sequence could indicate a potential safety risk that might be overlooked in manual monitoring.

We have recognized the potential to develop unsupervised learning models to detect these hidden risks. Unlike supervised learning methods that require extensively labeled training data, unsupervised learning can identify patterns and detect anomalies by leveraging the inherent structure of the data itself without explicit programming. This approach is particularly valuable for railway crossing safety where normal operational patterns can be learned, allowing deviations that may indicate potential safety issues to be automatically flagged for operator attention.

Therefore, there is a pressing need for automated systems that can not only collect data but also intelligently process and identify potential anomalies, supporting operators in making more timely and informed decisions.

## Related Works

Statistics from 1990 to 2020 show that even countries with developed railway infra-structure such as China and Japan face significant challenges [1]. Both countries, along with many others [1-2], continue to experience railway accidents due to environmental factors, poor management, and other challenges.

The complexity of railway safety management varies significantly across different economic contexts. Research on transport policy and management across low, middle, and high-income Asian countries demonstrated that while technological solutions are important, the effectiveness of safety measures heavily depends on the local context and management capabilities [3]. This is particularly relevant for Vietnam's railway system, where the railway infrastructure is outdated compared to global standards.

In the context of railway management, proactive safety is always the first priority [4]. Traditional monitoring methods often fail to detect potential risks before they develop into accidents. Railway safety systems must integrate risk management with real-time monitoring capabilities, indicating that advanced technologies such as Machine Learning [5] can play a crucial role in enhancing safety measures. The integration of ML technologies offers promising solutions for enhancing railway safety systems through improved risk detection and management capabilities [6-8]. The deployment of IoT devices at these crossings generates an exceptionally large volume of data [9]. Machine learning excels at identifying subtle patterns in large, complex datasets that may not be apparent through traditional statistical analysis [10].

To build a machine learning model with high accuracy, data preprocessing and feature extraction play a crucial role, as sensors at railway level crossings typically have different measurement scales [11-12]. Normalization methods such as min-max scaling or Z-score are particularly important for magnetic sensors used in train detection, where environmental conditions can significantly affect raw measurements, requiring normalization to maintain consistent detection thresholds across different locations and weather conditions [13]. Additionally, dimensionality reduction helps significantly improve system performance. Methods such as Principal Component Analysis (PCA) or t-SNE are effective in reducing data complexity for railway systems with extensive data [14].

The One-Class Support Vector Machine (OCSVM) algorithm is particularly well suited to the problem of anomaly detection at railway crossings, as data in this area is often unlabeled. OCSVM is highly sensitive to rare and unusual data patterns [15], such as incidents with very low occurrence frequencies. This makes OCSVM particularly useful for safety monitoring at crossings, where severe anomalies may not have been present in historical data. However, the ability to detect these anomalies early is key to preventing accidents and ensuring safe operations.

## Contribution

Our study proposes data filtering and normalization methods, thereby building a highly reliable training dataset that is suitable for the characteristics of the railway industry [32].

This research serves as a foundation for future approaches to processing data collected from devices at railway crossings via the Internet of Things (IoT) infrastructure. The proposed method demonstrates the ability to efficiently integrate and analyze multiple data streams, thereby supporting the early detection of potential safety-related risks and system operation anomalies.

This research offers valuable insights into the challenges and solutions associated with implementing smart monitoring systems in railway networks in developing countries where infrastructure and management capacity are limited. These insights are particularly important for countries facing similar issues in infrastructure development and modernization, providing a practical application framework that can be flexibly adapted to different operational contexts considering local technological and managerial constraints.

## Thesis structure

The remainder of this paper is organized as follows. Section 2 describes the background study. Section 3 provides system design. Finally, Section 4 discusses experimental results.

# Background Study

## IOT

### Sensor Data Acquisition

The level crossing monitoring system collects data from various sensors and measuring devices. Understanding the measurement methodology as well as the raw data plays a crucial role as input for subsequent anomaly detection analysis processes.

### Data Transmission via network

Data transmission is an essential component in the IoT system for monitoring railway crossings, serving as the connection between field sensors and the analysis center. The infrastructure uses multiple communication protocols to ensure flexibility and reliability:

* Local protocols: Communication between sensors and controllers typically uses industrial protocols such as Modbus, RS-485, or I2C for data transmission within a limited range.
* Wide-area network protocols: Data from sensors at railway crossings is transmitted to the central analysis system using secure 3G/4G cellular networks. This transmission channel is isolated and protected by TLS encryption to ensure both security and efficient bandwidth utilization. Communication between users and field devices is facilitated through an intermediary server via WebSocket protocols, enabling real-time interaction and monitoring
* Message structure:Messages are packaged according to JSON or XML standards

At crossing points, 3G/4G transmission technology is used to transmit data, with the use of 3G/4G technology leveraging Vietnam's leading telecommunications infrastructure.

## Data mining

Data mining plays a crucial role in transforming raw IoT sensor data collected from railway level crossings into valuable, practically applicable insights. The techniques presented in this section provide the theoretical foundation for extracting meaningful features from sensor data [16] at railway crossings, thereby supporting the development of high-performance unsupervised learning models.

### Visualization Data

Data visualization in data mining is the process of using visual tools such as charts, graphs, and images to clearly and intuitively present data and analytical results [17]. Instead of delving deeply into complex data tables filled with numbers and technical details, visualization enables users to quickly grasp essential information, identify trends, patterns, and hidden anomalies within large datasets.

In the context of operations, maintenance, and research in railway systems, visualization offers significant benefits. It provides managers, technical teams, and researchers with clear, direct insights into system behavior, facilitating faster, more accurate, and efficient decision-making processes. Additionally, combining data mining with visualization plays a crucial role in early detection of operational anomalies or potential risks arising from system degradation, thus enabling proactive maintenance and enhancing overall railway safety and reliability.

### Data Preprocessing

Effective data preprocessing plays a critical role, directly influencing the performance of subsequent anomaly detection algorithms. If input data is not properly cleaned, standardized, and processed, anomaly detection systems may perform inaccurately [18], generating excessive false alarms that waste resources or, more seriously, missing important safety-related events. This negatively impacts the reliability and overall safety of the entire system. Therefore, a careful and well-designed data preprocessing strategy enhances accuracy, reduces error rates, and ensures the system can promptly identify and respond to genuinely critical anomalies.

### Feature Extraction and Engineering

Sensors in railway systems typically exhibit direct interactions and interdependencies rather than merely representing independent statistical properties. To effectively exploit these critical characteristics, it is essential to apply appropriate data transformation techniques [19]. These techniques enable the extraction of meaningful features that accurately reflect the complex interactive nature among sensors, while simultaneously reducing noise and eliminating redundant factors. Selecting and employing suitable data processing methods directly influence the model's ability to detect anomalies, thereby significantly enhancing the accuracy and reliability of anomaly detection models.

### Temporal Pattern Analysis

Understanding time dependencies is particularly important in managing railway crossings, where the sequence and timing of each event (such as detecting an approaching train, operating the barrier system, activating warning signals) must strictly adhere to safety protocols. Any deviation from the standard time model can become an early warning signal of potential safety issues [20]. Timely identification of these anomalies forms the foundation for building an effective anomaly detection system.

## Unsupervised Learning Techniques

### Overview of Unsupervised Learning Methods

Unsupervised learning is an important branch of machine learning, where algorithms automatically detect hidden patterns and structures within unlabeled data. For anomaly detection problems, this method has significant advantages due to its ability to identify rare or abnormal data patterns without depending on previously labeled datasets [21].

In the context of safety management at railway crossings, where collected sensor data often lacks or has very limited anomaly labels, unsupervised learning becomes an ideal solution. These algorithms can actively identify early warning signs from raw data based on statistical properties and data structures, thereby timely detecting potential anomalies, such as malfunctions in barrier operations, warning errors, or time deviations of events. This significantly improves monitoring efficiency and ensures the safe operation of the level crossing system, even when lacking anomalous data samples to train models according to traditional supervised learning methods

### One-class SVM modeling

Support Vector Machines (SVM) are a popular supervised learning algorithm for classification and regression tasks. SVM work by finding the optimal hyperplane that separates different classes in feature space while maximizing the margin between them. This hyperplane is based on a subset of training data points called support vectors. Unlike traditional SVM, which handle binary classification tasks, One-class SVM exclusively trains on data points from a single class, known as the target class. One-class SVM aims to learn a boundary or decision function that encapsulates the target class in feature space, effectively modeling the normal behavior of the data.

One-class SVM aims to discover a hyperplane with maximum margin within the feature space by separating the mapped data from the origin. On a dataset Dn = {x1 ...xn} with xi ∈ X are data points, and X is the n-dimensional feature space. Equation represents the primal problem formulation for OC-SVM:

: is the separating hyperplane

: is the offset from the origin

: are slack variables of data i

: a hyperparameter controls the effect of the slack variable

### Hyperparameter Tuning

The ν parameter, introduced in the primal problem formulation, serves a dual role in the One-Class SVM. Firstly, it represents an upper bound on the fraction of anomalies or outliers permitted in the training data. Secondly, it serves as a lower bound on the fraction of support vectors relative to the total number of training samples. Mathematically, ν ∈ (0, 1] controls the trade-off between maximizing the margin (distance from the origin) and encompassing most data points within the decision boundary. A smaller ν value results in a tighter boundary around normal data points, potentially increasing the false negative rate. Conversely, a larger ν value creates a looser boundary, potentially increasing false positives.

In addition to ν, the kernel selection and its parameters significantly influence model performance. This study applies two kernel functions for hyperparameter tuning: the Radial Basis Function (RBF) and the Linear kernel.

When using the RBF kernel, widely employed in anomaly detection tasks due to its flexibility, the parameter determines the complexity and smoothness of the decision boundary. The kernel function is defined as:

A smaller γ value produces a smoother, more generalized decision boundary that might miss subtle anomalies, while a larger γ value creates a more complex boundary that closely fits training data, increasing the risk of overfitting.

Alternatively, the Linear kernel, defined mathematically as:

Produces a simpler and less flexible boundary. Although computationally efficient and effective when data is linearly separable, it may perform poorly if the anomaly patterns exhibit complex, nonlinear structures.

Therefore, tuning ν and selecting appropriate kernel parameters for both RBF and Linear kernels is crucial for optimizing anomaly detection performance, balancing generalization, complexity, and accuracy effectively.

### Evaluation Metrics

The following standard metrics are essential:

**Accuracy** measures the ratio of correctly classified instances (both normal and anomalous) to the total number of instances.

**TP (True Positive):** Cases where the model correctly identifies an anomaly.

**TN (True Negative):** Cases where the model correctly identifies normal operation.

**FP (False Positive):** Cases where the model incorrectly flags normal operation as anomalous.

**FN (False Negative):** Cases where the model fails to detect an actual anomaly.

In railway crossing monitoring, while accuracy provides an overall performance assessment, it can be misleading due to the class imbalance inherent in anomaly detection. A model that simply classifies everything as normal could achieve high accuracy but completely fail to detect critical safety anomalies.

**Precision** quantifies the proportion of correctly identified anomalies among all instances predicted as anomalous.

High precision indicates that when the system flags an anomaly, it is likely to be genuine. In railway operations, precision is crucial for maintaining trust in the alert system and preventing unnecessary maintenance interventions or operational disruptions due to false alarms.

**Recall** (also known as sensitivity) measures the proportion of actual anomalies that are correctly identified by the model.

In safety-critical railway applications, recall is particularly important as it represents the system's ability to detect all potential safety issues. A high recall ensures that few genuine anomalies go undetected, which is paramount for preventing accidents and maintaining system integrity.

**F1-Score** is the harmonic mean of precision and recall, providing a balanced measure that accommodates both metrics.

For railway crossing monitoring, the F1-Score helps balance the competing objectives of minimizing false alarms while maximizing anomaly detection. It is especially useful when seeking a single metric to optimize during model development.

**ROC AUC** is Receiver Operating Characteristic Area Under Curve (ROC AUC) quantifies the model's ability to distinguish between normal and anomalous states across various threshold settings.

**N** is the total number of threshold points used to plot the ROC curve.

**i** is the index representing the current threshold point within this sequence of thresholds.

**FPR** (False Positive Rate) is:

**TPR** (True Positive Rate) is:

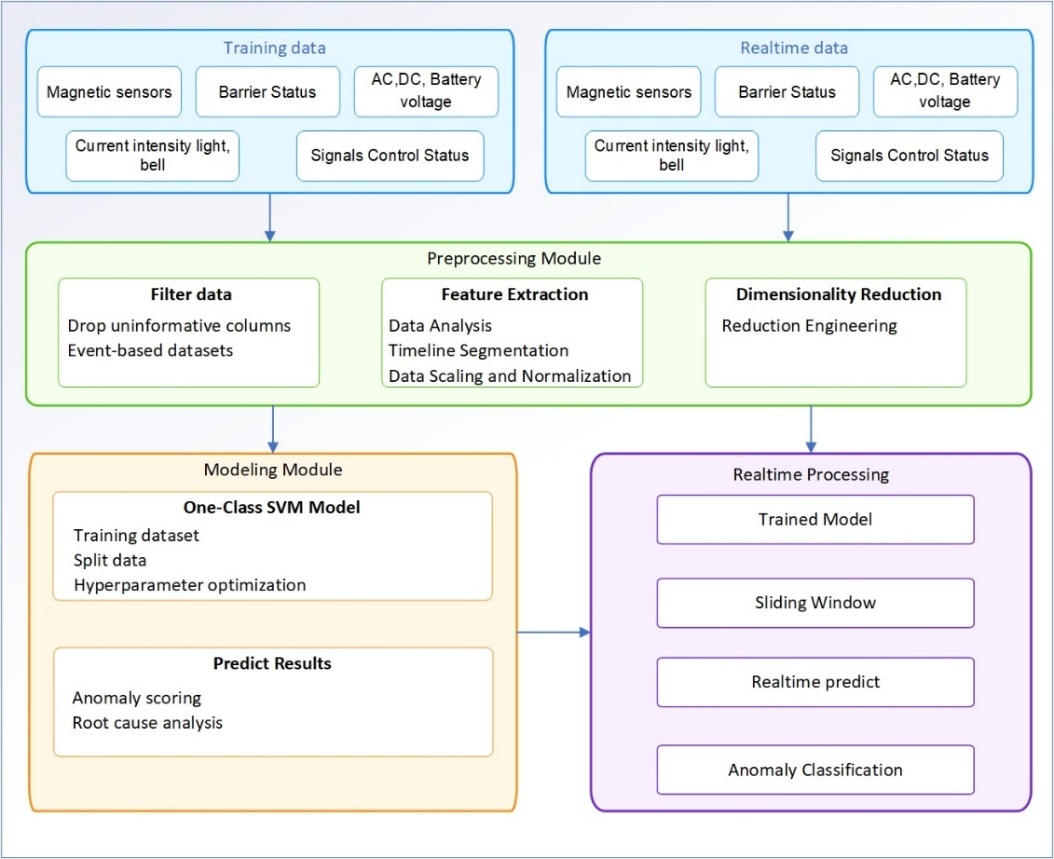
ROC AUC ranges from 0 to 1, where: 0.5 indicates performance no better than random guessing. 1.0 represents perfect classification

ROC AUC is a comprehensive metric when evaluating classification models, especially in difficult problems such as anomaly detection or imbalanced data.

# System Design

## System Architecture

Our proposed method provides a comprehensive approach for anomaly detection at railway crossings, structured into three main modules: Preprocessing, Modeling, and Real-time Processing. Raw data collected from IoT devices installed at railway crossings is stored in our data center. The overall architecture of the proposed system is illustrated in Figure 2.



**Figure 2**. Overall architecture

The data is then processed through two parallel preprocessing threads:

* **The first preprocessing thread** prepares data specifically for the Modeling Module, where we employ a One-Class SVM machine learning model to establish reliable anomaly detection capabilities.
* **The second preprocessing thread** focuses on preparing data for real-time analysis, directly feeding into the Real-time Processing Module. This module leverages a sliding window technique to continuously sample incoming data, enabling immediate anomaly detection and prediction during railway crossing operations.

## Data Preprocessing

### Filter and clean data

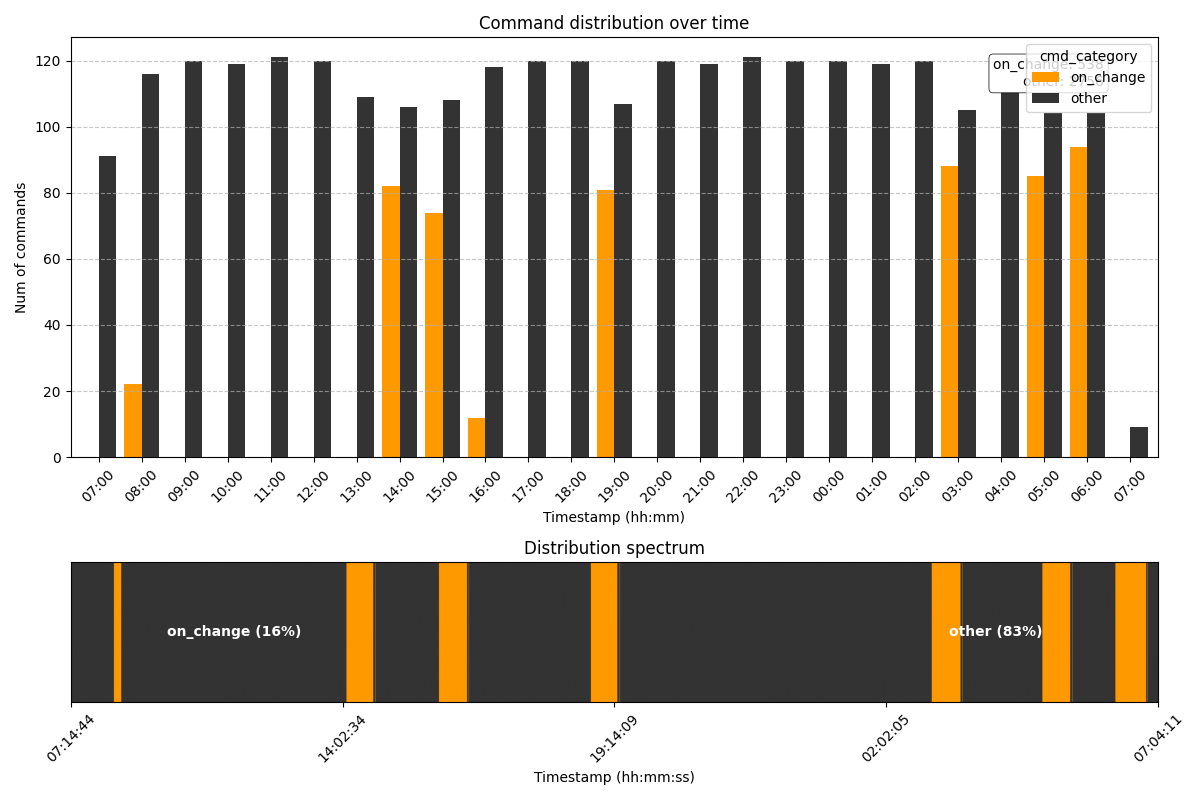
Raw data collected from IoT systems at railway level crossings in Vietnam contains various types of information, primarily distinguished by the "cmd" field. As illustrated in *Table 1*, when a train passes through a level crossing, messages with "cmd=on\_change" are triggered, indicating critical state changes within the monitoring system. Each "on\_change" message is timestamped and sampled at a frequency of one message every 5 seconds, thus forming a continuous sequence of events over time.

Figure 3 illustrates the distribution of collected data by hour over a 24-hour period at a railway crossing. This visualization clearly highlights the specific time intervals when trains are active during the day. Additionally, by observing the width of data segments associated with the on\_change command, we can distinguish the different types of trains passing through this crossing.

The process of filtering and cleaning data will remove noise and standardize information. Data containing a lot of useful information plays a very important role in the next analysis steps so that the model can achieve high accuracy.

**Table 1.** The raw data example from railway level crossing

|  |
| --- |
| 04032025\_081713: "type":"request","cmd":"get\_time"  04032025\_081744:"type":"event","cmd":"send\_data","body":"AD1":25.0,"AD2":12.5,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":222,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":0,"S2":0,"S3":0,"S4":0,"S5":1,"S6":0,"S7":1,"S8":0,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":0,"I2":0,"I3":0,"I4":0,"I5":0,"I6":0,"I7":0,"I8":0,"I9":0  04032025\_081815:"type":"event","cmd":"send\_data","body":"AD1":25.0,"AD2":12.5,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":222,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":0,"S2":0,"S3":0,"S4":0,"S5":1,"S6":0,"S7":1,"S8":0,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":0,"I2":0,"I3":0,"I4":0,"I5":0,"I6":0,"I7":0,"I8":0,"I9":0  04032025\_081844:"type":"event","cmd":"send\_data","body":"AD1":25.0,"AD2":12.5,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":223,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":0,"S2":0,"S3":0,"S4":0,"S5":1,"S6":0,"S7":1,"S8":0,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":0,"I2":0,"I3":0,"I4":0,"I5":0,"I6":0,"I7":0,"I8":0,"I9":0  04032025\_081915:"type":"event","cmd":"send\_data","body":"AD1":25.0,"AD2":12.5,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":222,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":0,"S2":0,"S3":0,"S4":0,"S5":1,"S6":0,"S7":1,"S8":0,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":0,"I2":0,"I3":0,"I4":0,"I5":0,"I6":0,"I7":0,"I8":0,"I9":0  04032025\_081930:"type":"event","cmd":"on\_change","body":"AD1":23.8,"AD2":12.3,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":222,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":0,"S2":0,"S3":0,"S4":0,"S5":1,"S6":0,"S7":1,"S8":0,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":16,"I2":0,"I3":0,"I4":0,"I5":15,"I6":18,"I7":26,"I8":26,"I9":0I1  04032025\_081935:"type":"event","cmd":"on\_change","body":"AD1":23.7,"AD2":12.3,"AD3":0.0,"AD4":0.0,"AS1":0,"AS2":0,"AS3":0,"AS4":0,"AI1":223,"AI2":0,"AI3":0,"DI1":1,"DI2":0,"DI3":0,"DI4":0,"DI5":0,"RL1":0,"RL2":0,"RL3":0,"RL4":0,"RL5":0,"RL6":0,"RL7":0,"RL8":0,"RL9":0,"HN":0,"TN":29,"GPRS":31,"S1":2,"S2":2,"S3":0,"S4":1,"S5":1,"S6":0,"S7":1,"S8":2,"S9":0,"S10":1,"C1":1,"C2":1,"C3":1,"C4":1,"C5":1,"C6":1,"C7":1,"C8":0,"C9":0,"C10":0,"C11":0,"C12":0,"C13":1,"C14":0,"X1":0,"X2":0,"X3":0,"X4":0,"X5":0,"X6":0,"I1":17,"I2":0,"I3":0,"I4":0,"I5":0,"I6":0,"I7":49,"I8":48,"I9":2 |



**Figure 3**. Train event data by time

Raw data will be filtered to remove unnecessary records and fields. Each train passing through a level crossing will be represented as a dataset containing valuable information collected from IoT devices over time, identified by a unique random UUID. The output will be a CSV file, we have filtered the most important information fields such as: magnetic sensor, control signal, barriers status, power supply voltage, current intensity. The information of these fields is described in Table 2.

**Table 2.** Key IoT Data Fields for Railway Crossing Monitoring

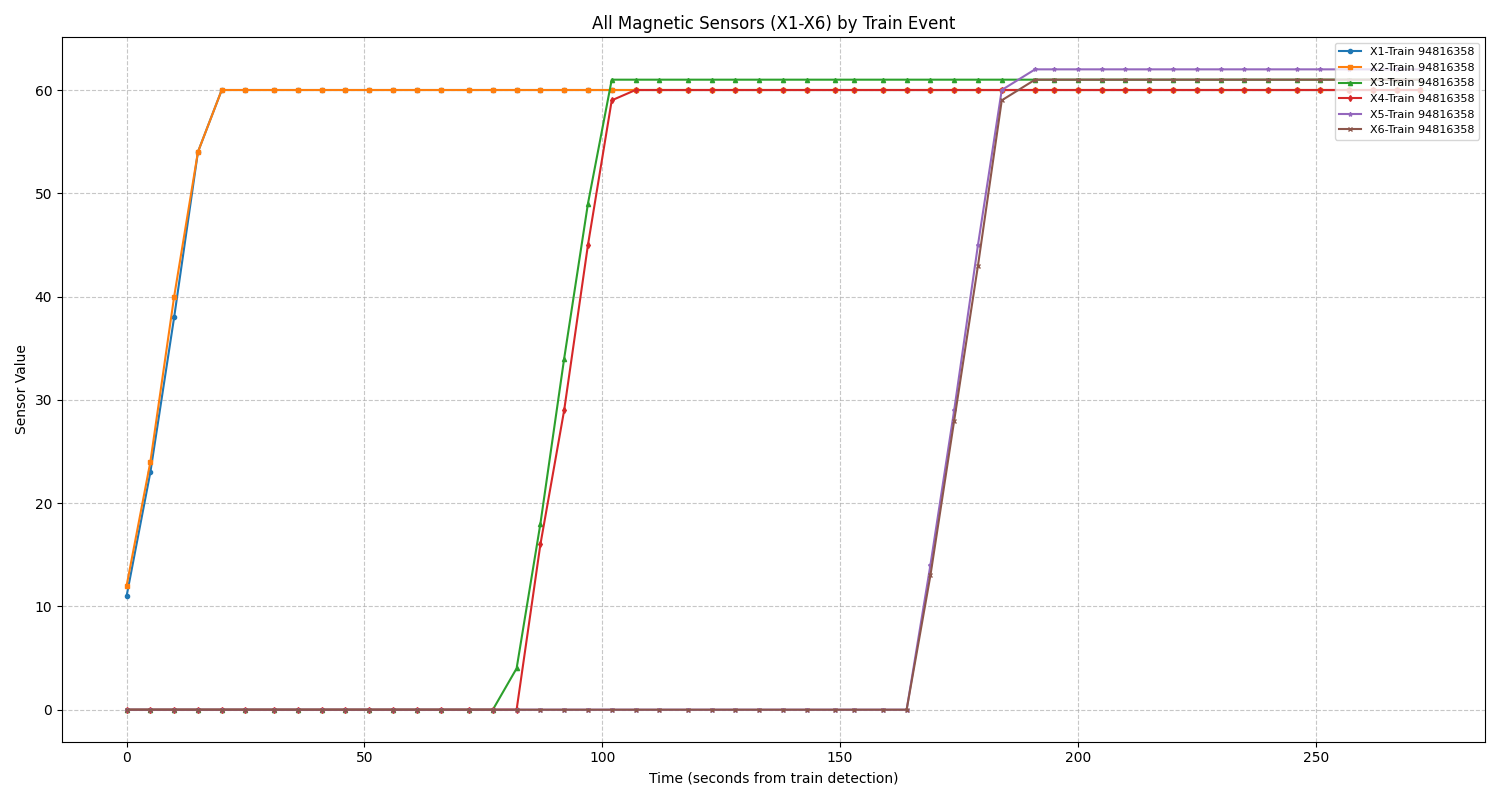
|  |  |  |  |
| --- | --- | --- | --- |
| No | Field Name | Description | Value Range/Unit |
| 1 | UUID | Train event id | Random 6-digit integer |
| 2 | time\_stamp | Timestamp | ddMMyyyy\_HHmmss |
| 3 | X1, X2 … X6 | Magnetic sensor values | Integer values |
| 4 | S5 | PLC control signal | 0: Not operational 1: Ready 2: Controlling |
| 5 | S6, S7 | Barrier 1 control signal | Same as above |
| 6 | S9, S10 | Barrier 2 control signal | Same as above |
| 7 | C7, C8 | Open/closed status of barrier 1 | 0: Undetermined  1: Opened/Closed |
| 8 | C13, C14 | Open/closed status of barrier 2 | Same as above |
| 9 | AI1 | AC power voltage | 220 Volt |
| 10 | AD2 | DC power voltage | 24 Volt |
| 11 | AD3 | Battery voltage | 12 Volt |
| 12 | I1 | Warning light current intensity | Blink. 0 to 2000mA |
| 13 | I3, I4 | Yellow lights current intensity | Same as above |
| 14 | I5, I6, I7, I8 | Red lights current intensity | Same as above |
| 15 | I9 | Bell current intensity | Always on |

### Feature Extraction

As described in the previous section, after filtering and cleaning, the dataset will consist of multiple subsets. Each subset represents a single instance of a train passing through a level crossing, uniquely identified by a UUID. This section outlines the techniques applied to extract relevant features from the dataset. Based on the information described in ***Table 2***, the collected data related to infrastructure at railway crossing points can be divided into three main groups as follows:

* Magnetic sensor group (X1–X6)
* Control signal group ( , measuring sensors)
* Group of values related to protective equipment (bells, lights, power supply)

The magnetic sensor group (X) is responsible for counting pulses generated as train wheel axles pass through the sensors, as clearly illustrated in ***Figure 4***.



**Figure 4.** Magnetic sensors value by time

The data will be normalized and feature vectors will be extracted from the dataset of each train crossing event. The methods implement:

**Min-max Scale.** The maximum value of the magnetic sensor signal, denoted as X\_max​, directly reflects the number of wheel axles in a train. An important characteristic of the data from group X is that the values gradually increase over time and reach their peak as the train passes, with this maximum value being relatively stable. Therefore, applying analytical methods based on the maximum and minimum values of magnetic sensor data significantly improves the performance of prediction models and anomaly detection [22-23]. With the threshold for train detection, the conversion formula is defined as follows:

**Parameter Analysis.** Based on the data density chart in Figure 3 and the magnetic sensor value variation chart in Figure 4, it is evident that the total amount of data collected for each train journey varies. This variation depends on the train length (number of connected carriages), average velocity, and the weight of the train's current load.

The main features are calculated as follows:

* *Train velocity feature*:This index reflects the movement speed of the train when passing the sensor.

*data\_size* is the total number of data messages collected for each train journey

is the time interval between the start and end of the train journey

* *Data Density Feature***:** This index represents the frequency of data collection or the level of detail in the recording process, thereby helping to distinguish between different types of trains.

* *Relative Weight Feature***:** With the same X\_max value (same number of carriages), heavily loaded trains typically move more slowly**.**

***Peak* Difference Analysis (PDA)** for sensor pairs. PDA is typically used when we have two or more variables/series that are closely correlated. In railway level crossing system, sensor pairs are deployed consecutively to backup each other, so monitoring the at peak values is critically important. By extracting peak features and calculating the differences, we can highlight sudden deviations that models based on raw values might overlook:

and are the timestamps when peaks occur at sensors i and i-1.

This is especially useful for detecting anomalies [24] related to irregular train speeds.

***Temporal* Feature Extraction**.Identifying time markers to include in the feature vector is very significant [25]. For this system, there is a strong relationship between sensor values and timestamps.

For example, when a train is detected entering the level crossing (X1 > Threshold), the time interval from when the barrier closure control signal is given until the barrier is completely closed has a major impact on safety. We will analyze and incorporate these time intervals ∆t into the feature vector:

**Z-score.** Is a powerful tool that can significantly improve the performance of models in anomaly detection tasks [26], especially when combined with min-max scaling, peak difference analysis, and Temporal Feature Extraction [27].

Apply Z-Score for each phase: train entering, train passing through, and train leaving the level crossing to enhance the sensitivity of anomaly detection when trains are running:

value of sensor at time t in the train event

are mean value and standard deviation of sensor at time t across multiple train events.

For example, to detect early mechanical degradation of the barrier system before failures occur, we can apply the formula.

and are the mean value and standard deviation of the barrier closing time.

## Modeling Module

At this phase, data is classified and anomaly levels are evaluated through unsupervised learning methods, where the One-Class SVM model is selected to build the system. The information of the data fields after preprocessing is described as Table 3.

**Table 3.** Features Vector detail

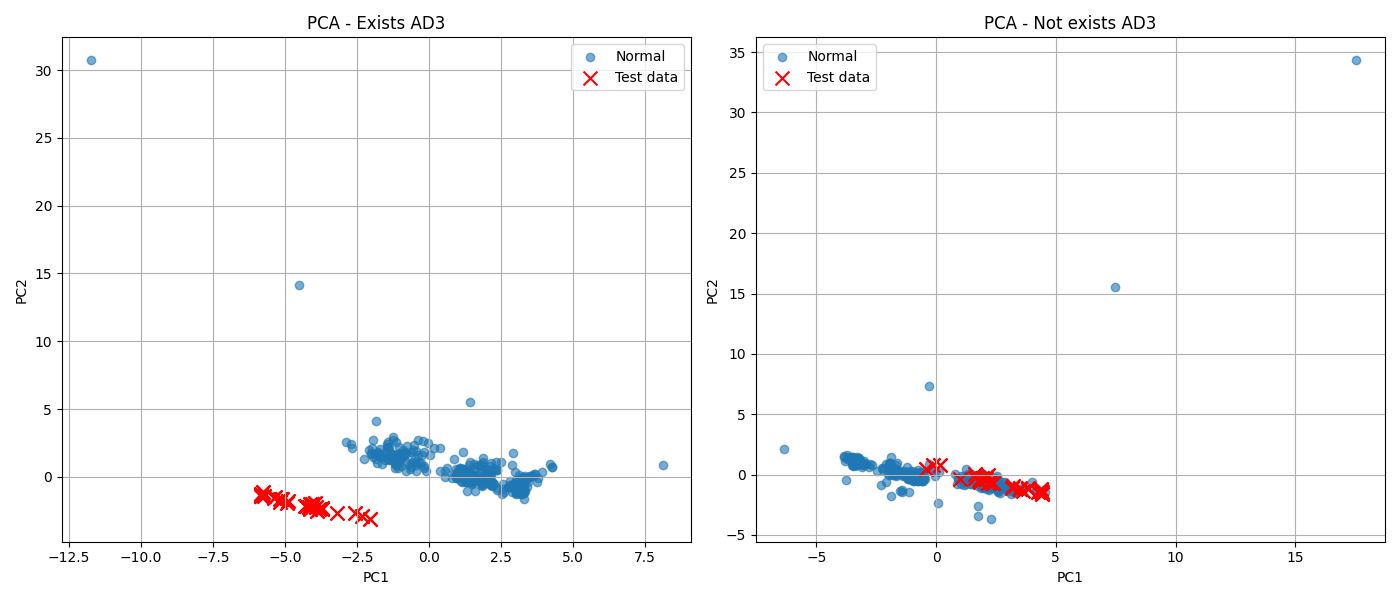
|  |  |  |  |
| --- | --- | --- | --- |
| No | Field Name | Description | Value Range/Unit |
| 1 | max\_X1, … max\_X6 | Pulses of magnetic sensors | Pulse(integer) |
| 2 | delta\_X1X2, delta\_X3X4, delta\_X5X6 | Discrepancy between paired magnetic sensors | Seconds(float) |
| 3 | X1\_ascending ... X6\_ascending | Time until magnetic sensor reaches max value | Normalized(float) |
| 4 | mean\_AI1, mean\_AD2, mean\_AD3 | Mean voltage values of power sources | Volt |
| 5 | std\_AI1, std\_AD2, std\_AD3 | Voltage standard deviation | Volt |
| 6 | min\_AI1, min\_AD2, min\_AD3 | Minimum voltage observed | Volt |
| 7 | max\_AI1, max\_AD2, max\_AD3 | Maximum voltage observed | Volt |
| 8 |  |  |  |
| 9 |  |  |  |
| 10 |  |  |  |
| 11 |  |  |  |
| 12 |  |  |  |
| 13 |  |  |  |
| 14 |  |  |  |
| 15 |  |  |  |

### Dimensionality Reduction

At this phase, data is classified and anomaly levels are evaluated through unsupervised learning methods, where the One-Class SVM(OCSVM) model is selected to build the system. We choose OCSVM for anomaly detection in railway systems because:

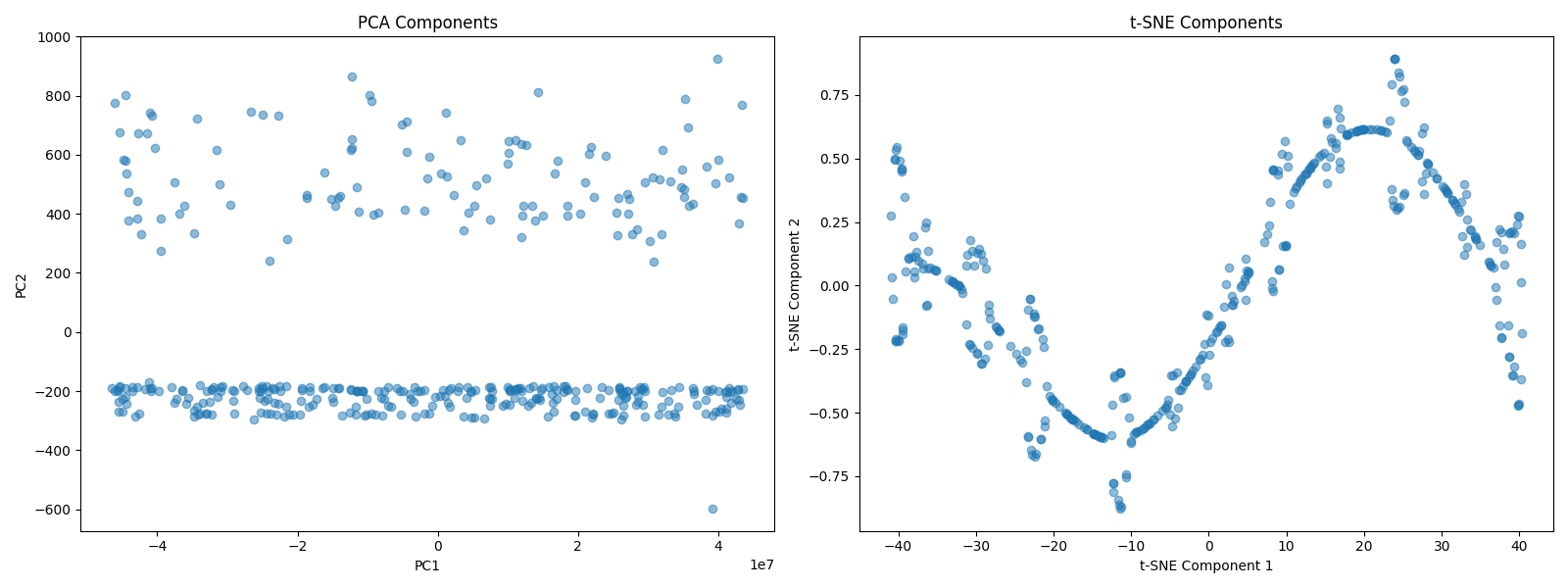
* The data collected from railway systems are mainly normal operating cases, while anomalies are rare. OCSVM is specially designed for this imbalanced data, where normal data samples are the majority.
* One-class SVM can detect nonlinear relationships between features. This is important in this problem because the relationships between parameters such as magnetic sensor, ∆t, and the states of the barriers are often nonlinear.

In some practical scenarios, including the AD3 (Battery Voltage) feature in the model may introduce noise during the training process, since not all horizontal lines are equipped with battery power sources. To assess the contribution of this feature, we apply Principal Component Analysis (PCA) to visualize its effect on the feature space and its influence on data separability. **Figure 5.** illustrates in detail the differences in data distribution when PCA is applied to the dataset with and without the AD3 feature.



**Figure** **5.** Explain of measure AD3 in data set with PCA

After the data has been normalized using the aforementioned methods, we will proceed with dimensionality reduction to make the dataset more concise while still preserving the data characteristics. Common dimensionality reduction methods such as PCA (Principal Component Analysis), t-SNE (t-Distributed Stochastic Neighbor Embedding) will be applied in this step.



**Figure 6**. PCA and t-SNE distribute

### Tuning model

To implement One-Class SVM, the model is trained on a dataset representing normal operations to determine the boundary of standard behavior. The RBF (Radial Basis Function) kernel and Linear kernel techniques is used to better capture non-linear relationships between features. Hyperparameter tuning is applied to find optimal values for the hyperparameter pair (, ), helping to balance between model sensitivity and specificity. Finally, a decision function is implemented to classify new observations, determining whether they belong to normal or anomalous behavior.

To evaluate the model, the first step is to extract normal operational data from various level crossings for training. Next, a synthetic anomaly dataset is generated with varying anomaly occurrence rates to assess the model’s detection performance. Performance is then measured using precision, recall, F1-score, and AUC-ROC. Finally, a hyperparameter sensitivity analysis is conducted to ensure the model’s stability and generalization capability.

## Realtime processing

To deploy and apply trained models in production system, there is a huge challenge: there will be many incidents that are only detected when the train reaches a specific time point in the train running event chain. For each stage, we can detect different abnormalities. To solve this problem, we have implemented a method that combines Sliding Window and Dynamic Template.

### Sliding Windows

Sliding Window technique plays an important role in detecting abnormalities by analyzing multivariate data series from real stages for the railway system, through analyzing multivariate data series from IoT points. This method is based on a number of defined technical characteristics, allowing to track statistical characteristics and correlations between sensors in each specific stage of the train passing event.

We can evaluate the similarity between windows and the standard template [28] using the formula:

is the inner product between and ,

is a weight vector which is based on the eigenvalues of dataset

is the angle between and

### Dynamic Template

Applying the Dynamic Template allows early detection of anomalies without waiting for the event to complete in order to issue an alert [29]. Real-time data of train crossing events are continuously updated, while the standard template automatically adjusts when gradual changes occur. As a result, the system does not need to store the entire historical data, helping to optimize computational resources.

# Experiments And Results

## Data collection

Our data was collected from the Hanoi-Hai Phong railway route, one of the busiest railway corridors in northern Vietnam. We selected 8 consecutive level crossings with the most stable operations to collect data over a 5-day period (from March 1 to March 5, 2025). These crossings were chosen because they have reliable infrastructure, creating favorable conditions for the research process.

For each day, the raw data is stored as a text file. After filtering the data, our dataset is transformed into CSV files, with a summary presented in Table 4:

**Table 4.** Summary of dataset

|  |  |
| --- | --- |
| Total CSV files | 40 |
| Total train sessions | 398 |
| Total data rows | 30421 |
| Rows per event | 76.43 |

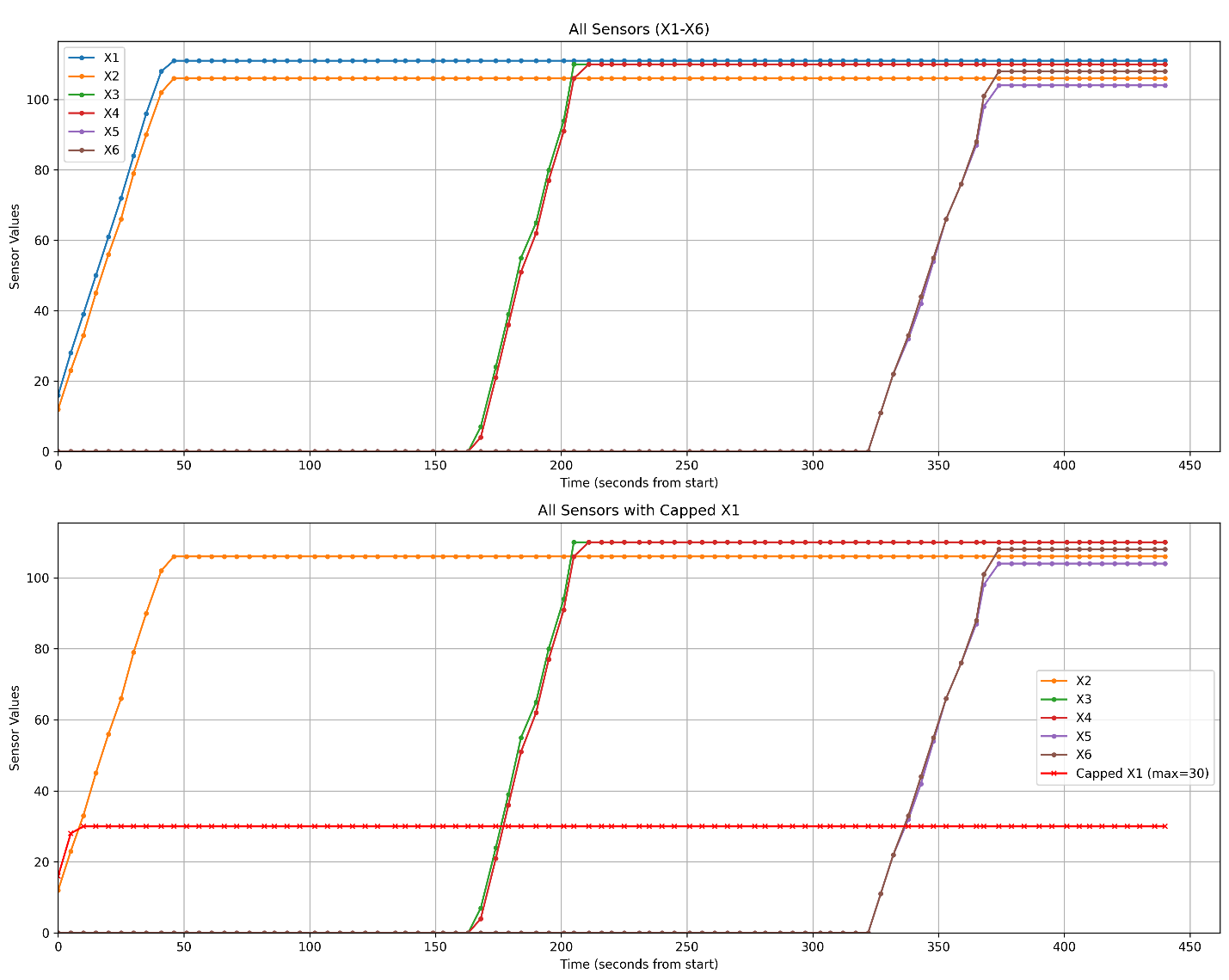
## Experiments

Since performance evaluation in unsupervised learning is crucial, all our collected data consists of train events operating under normal conditions (without incidents). This creates a major challenge known as class imbalance [30] in machine learning. To address the lack of incident data samples, we used historical error data, combined with simulated scenarios developed by railway experts, and integrate these samples into the original dataset. Table 5 summarizes the types of anomalies we simulated and the implementation methods used to generate these anomalies in our dataset.

**Table 5.** Anomaly types and implementation methods

|  |  |
| --- | --- |
| Anomaly Type | Implementation Method |
| Sensor Detection Failure | Set sensor pair value |
| Barrier Malfunction | Extend closing time |
| Signal Control Delay | Increased time intervals |
| Power Instability | Voltage fluctuation |

The magnetic sensors count pulses generated each time a train wheel axle passes, and at peak operation, the variation between sensors typically remains within an expected range of values. To simulate a Sensor Detection Failure scenario for a train event, we will set all X1 values after the pulse count reaches 30 to a fixed value of 30. Figure 7 illustrates the variation in magnetic sensor values for a normal train event and a simulated faulty train event.



**Figure 7.** Magnetic sensor values for a normal train event and a simulated faulty train event

Barrier Malfunction errors are defined by the following process: find the first da-ta point that satisfies the condition C7=0 and C8=1, then mark the next 10 consecutive records with values C7=0 and C8=0. The barrier closure time will be extended by the duration corresponding to the next 10 records, which results in a Barrier Malfunction.

The time from when the number of pulses counted by X1 or X6 exceeds the detection threshold (usually 7 pulses) until S7=1 and S10=1 is defined as the system reaction time. To generate the Signal Control Delay error, we update the value at the first time S7 = 1 and the next two records to 0, because this time interval is usually very short.

AC power typically has a threshold of 220V and DC power with a threshold of 24V usually operates very stably with a 5% error margin. We modify a few random records with a deviation higher than 10% to create the Power Instability error.

## Results

The characteristics of the dataset were collected under normal conditions, including 400 train events as described in Table 2. We divided the data into a training set and a testing set at a 50% ratio, with each set containing approximately 200 train events. From the training set, we proceeded to train the OCSVM model.

To evaluate the model based on metrics such as Accuracy, Precision, Recall, and F1-Score, our approach involves mixing different error ratios with our testing dataset, introducing anomalies at various proportions: 20% (40 samples), 10% (20 samples), 5% (10 samples), and 2% (4 samples) in a dataset of 200 samples. These ratios help evaluate our model's performance under both high anomaly density scenarios and more realistic conditions where anomalies are relatively rare. Testing with such varied ratios helps us gain deeper insights into how the model responds in different scenarios, while also helping identify the sweet spot [31] the point at which the model maintains effectiveness despite very low anomaly rates. This is particularly important in railway safety applications, where the cost of missing an incident (false negative) can be extremely high. Table 6 describes the metrics of the OCSVM model with default parameters v = 0.5, γ = scale, kernel = rbf.

**Table 6.** OCSVM default hyper params metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample error | Accuracy | Precision | Recall | F1-Score | ROC AUC |
| 40 | 0.7865 | 0.8974 | 0.8642 | 0.8805 | 0.6173 |
| 20 | 0.7765 | 0.9429 | 0.8148 | 0.8742 | 0.5185 |
| 10 | 0.7952 | 0.9706 | 0.8148 | 0.8859 | 0.5185 |
| 4 | 0.8293 | 0.9855 | 0.8395 | 0.9067 | 0.5309 |

To increase the model's performance, we performed hyperparameter tuning as shown in **Table 7**. We achieved very good results during experimentation. **Table 8** describes the results after tuning with v = 0.01, γ = scale, kernel = linear.

**Table 7.** Tuning OCSVM hyper params

|  |  |
| --- | --- |
| Parameter | Tuning Range |
| nu | 0.01, 0.05, 0.1, 0.2 |
| gamma | scale, auto, 0.1, 0.01 |
| kernel | rbf, linear |

**Table 8.** Best metrics of OCSVM model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sample error | Accuracy | Precision | Recall | F1-Score | ROC AUC |
| 40 | 0.8989 | 0.9091 | 0.9877 | 0.9467 | 0.8519 |
| 20 | 0.9294 | 0.9518 | 0.9753 | 0.9634 | 0.8519 |
| 10 | 0.9398 | 0.9750 | 0.9630 | 0.9689 | 0.7160 |
| 4 | 0.9634 | 0.9875 | 0.9753 | 0.9814 | 0.8519 |

## Conclusion And Future Work

This research has successfully developed an anomaly detection subsystem for railway crossings in Vietnam, applying unsupervised learning on IoT sensor data. The implementation of unsupervised learning not only helps reduce manual classification efforts but also enhances detection efficiency based on large-scale datasets.

This work establishes a foundation for future intelligent monitoring systems that can be adapted to different operational contexts while maintaining high reliability and performance. Our dataset, extracted in CSV format, can serve as a valuable reference resource for the community, supporting more effective research and development of models.

In the future, there is significant potential for expanding this research. For instance, the system can be upgraded to monitor detailed train journeys as they pass through crossings. Additionally, after the preliminary classification, integrating another machine learning model to specifically identify each type of incident will help obtain accurately labeled data. This data, in turn, can be used to improve and retrain the original model, enhancing the reliability and efficiency of the entire system.

# References

1. Cao, Y., An, Y., Su, S., Xie, G., & Sun, Y. (2022). A statistical study of railway safety in China and Japan 1990–2020. Accident Analysis & Prevention, 175, 106764.

2. Kyriakidis, M., Hirsch, R., & Majumdar, A. (2012). Metro railway safety: An analysis of accident precursors. Safety Science, 50(7), 1535-1548.

3. Shah, S. A. R., Ahmad, N., Shen, Y., Pirdavani, A., Basheer, M. A., & Brijs, T. (2018). Road Safety Risk Assessment: An Analysis of Transport Policy and Management for Low-, Middle-, and High-Income Asian Countries. Sustainability, 10(2), 389.

4. Bulakh, M., Okorokov, A., & Baranovskyi, D. (2021). Risk System and Railway Safety. IOP Conference Series: Earth and Environmental Science, 666, 042074.

5. Alawad, H., Kaewunruen, S., & An, M. (2020). Learning From Accidents: Machine Learning for Safety at Railway Stations. IEEE Access, 8, 633-648.

6. Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., & Hampapur, A. (2014). Improv-ing rail network velocity: A machine learning approach to predictive maintenance. Transportation Research Part C: Emerging Technologies, 45, 17-26.

7. Thaduri, A., Galar, D., & Kumar, U. (2015). Railway Assets: A Potential Domain for Big Data Analytics. Procedia Computer Science, 53, 457-467.

8. Paltrinieri, N., Comfort, L., & Reniers, G. (2019). Learning about risk: Machine learn-ing for risk assessment. Safety Science, 118, 475-486.

9. Singh, P., Elmi, Z., Meriga, V. K., Pasha, J., & Dulebenets, M. A. (2022). Internet of Things for sustainable railway transportation: Past, present, and future. Cleaner Logis-tics and Supply Chain, 4, 100065. https://doi.org/10.1016/j.clscn.2022.100065

10. Siddiqui, H. U. R., Saleem, A. A., Raza, M. A., Zafar, K., Munir, K., & Dudley, S. (2022). IoT Based Railway Track Faults Detection and Localization Using Acoustic Analysis. IEEE Access, 10, 106520-106533.

11. Islam, U., Malik, R. Q., Al-Johani, A. S., Khan, M. R., Daradkeh, Y. I., Ahmad, I., Alissa, K. A., Abdul-Samad, Z., & Tag-Eldin, E. M. (2022). A Novel Anomaly Detec-tion System on the Internet of Railways Using Extended Neural Net-works. Electronics, 11(18), 2813. https://doi.org/10.3390/electronics11182813

12. Inan, M. S. K., Liao, K., Shen, H., Jayaraman, P. P., Georgakopoulos, D., & Tang, M. J. (2023). DeepHeteroIoT: Deep Local and Global Learning over Heterogeneous IoT Sensor Data. Proceedings of the 20th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services.

13. Kumar, N., Krause, L., Wondrak, T., Eckert, S., Eckert, K., & Gumhold, S. (2024). Robust Reconstruction of the Void Fraction from Noisy Magnetic Flux Density Using Invertible Neural Networks. Sensors, 24(4), 1213. https://doi.org/10.3390/s24041213

14. Lasisi, A., & Attoh-Okine, N. (2021). Hybrid rail track quality analysis using nonlinear dimension reduction technique with machine learning. Canadian Journal of Civil Engi-neering, 48(12), 1713-1723. <https://doi.org/10.1139/cjce-2019-0832>

15. Yang, K., Kpotufe, S., & Feamster, N. (2021). An Efficient One-Class SVM for Anom-aly Detection in the Internet of Things. arXiv. <https://arxiv.org/abs/2104.11146>

16. Mangler, J., Seiger, R., Benzin, J.-V., Grüger, et al. (2024). From internet of things data to business processes: Challenges and a framework. arXiv. <https://doi.org/10.48550/arXiv.2405.08528>

17. Hemanth, J., Bhatia, M., & Geman, O. (Eds.). (2020). Data visualization and knowledge engineering: Spotting data points with artificial intelligence. Springer. <https://doi.org/10.1007/978-3-030-25797-2>

18. Bilik, S., & Horak, K. (2022). Feature space reduction as data preprocessing for the anomaly detection. arXiv. <https://arxiv.org/abs/2203.06747>

19. Ma, L., Huang, Z., Peng, B., Zhang, M., He, W., & Wang, Y. (2025). Sensor embedding and variant transformer graph networks for multi-source data anomaly detection. In H. Zhang, X. Li, T. Hao, W. Meng, Z. Wu, & Q. He (Eds.), Neural computing for advanced applications. NCAA 2024. Communications in Computer and Information Science (Vol. 2181). Springer. <https://doi.org/10.1007/978-981-97-7001-4_27>

20. Xu, J., Wu, H., Wang, J., & Long, M. (2022). Anomaly transformer: Time series anomaly detection with association discrepancy. arXiv. <https://arxiv.org/abs/2110.02642>

21. Kadiyala, P., Shanmukhasai, K. V., Budem, S. S., & Maddikunta, P. K. R. (2021). Anomaly detection using unsupervised machine learning algorithms. In A. Makkar & N. Kumar (Eds.), Deep learning for security and privacy preservation in IoT. Signals and Communication Technology. Springer. <https://doi.org/10.1007/978-981-16-6186-0_6>

22. Mejri, N., Lopez-Fuentes, L., Roy, K., Chernakov, P., Ghorbel, E., & Aouada, D. (2024). Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods. Expert Systems with Applications, 256, 124922.

23. Zhang, M., Xu, B., & Wang, D. (2016). An Anomaly Detection Model for Network In-trusions Using One-Class SVM and Scaling Strategy. In S. Guo, X. Liao, F. Liu, & Y. Zhu (Eds.), Collaborative Computing: Networking, Applications, and Worksharing (pp. 280-291). Springer, Cham. <https://doi.org/10.1007/978-3-319-28910-6_24>

24. Lee, B. S., Kaufmann, J. C., Rizzo, D. M., & Haq, I. U. (2023). Peak Anomaly Detec-tion from Environmental Sensor-Generated Watershed Time Series Data. In J. A. Los-sio-Ventura, J. Valverde-Rebaza, E. Díaz, & H. Alatrista-Salas (Eds.), Information Management and Big Data (pp. 156-170). Springer, Cham. <https://doi.org/10.1007/978-3-031-35445-8_11>

25. Zhao, P., Chang, X., & Wang, M. (2021). A Novel Multivariate Time-Series Anomaly Detection Approach Using an Unsupervised Deep Neural Network. IEEE Access, 9, 109025-109041. <https://doi.org/10.1109/ACCESS.2021.3101844>

26. Tutivén, C., Vidal, Y., Insuasty, A., Campoverde-Vilela, L., & Achicanoy, W. (2022). Early Fault Diagnosis Strategy for WT Main Bearings Based on SCADA Data and One-Class SVM. Energies, 15(12), 4381. <https://doi.org/10.3390/en15124381>

27. Salam, A., et al. (2024). Securing Smart Manufacturing by Integrating Anomaly Detec-tion With Zero-Knowledge Proofs. IEEE Access, 12, 36346-36360.

28. Zhang, M., Xu, B., & Gong, J. (2015). An Anomaly Detection Model Based on One-Class SVM to Detect Network Intrusions. In Proceedings of the 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN) (pp. 102-107). IEEE.

29. Lee, S., & Park, D. (2021). A Real-Time Abnormal Beat Detection Method Using a Template Cluster for the ECG Diagnosis of IoT Devices. HCIS, 11(04). <https://doi.org/10.22967/HCIS.2021.11.004>

30. Hosseini, S. M., Shafique, A., Babaie, M., & Tizhoosh, H. R. (2023). Class-imbalanced Unsupervised and Semi-Supervised Domain Adaptation for Histopathology Images. In Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1-7). IEEE.

31. Rufino, V. Q., et al. (2020). Improving Predictability of User-Affecting Metrics to Sup-port Anomaly Detection in Cloud Services. IEEE Access, 8, 198152-198167.

32. Ngo, T. (2025). Vietnamese Railway Crossing IoT Sensor Data [Dataset]. Kaggle. <https://www.kaggle.com/datasets/ngtuan12/vietnamese-railway-crossing-iot-sensor-data>