|  |  |
| --- | --- |
| cetlogo ***CHEMICAL ENGINEERING TRANSACTIONS***  ***VOL. xxx, 2025*** | A publication of  aidiclogo_grande |
| The Italian Association  of Chemical Engineering  Online at www.cetjournal.it |
| Guest Editors: David Bogle, Flavio Manenti, Piero Salatino  Copyright © 2025, AIDIC Servizi S.r.l. **ISBN** 979-12-81206-21-2; **ISSN** 2283-9216 | |

From Reactive to Proactive: a data-Driven Approach to Vehicle Maintenance

G. Sciontia,\*, P. Di Pietrob, M. Midurib, I. Blandinab, M.F. Milazzoa

aDepartment of Engineering, University of Messina, Contrada di Dio, 98166 Messina, Italy

bComet S.r.l., Via Garibaldi, 98121 Messina, Italy

giuseppe.scionti@studenti.unime.it

The digital transition and big data analysis provide a novel approach to predictive maintenance, enhancing efficiency and safety in the automotive industry.

This research aims to develop a groundbreaking methodology capable of forecasting malfunctions in heavy vehicles before breakdowns occur, thereby optimizing maintenance interventions and minimizing downtime. By implementing an IoT architecture, real-time telemetric data from the vehicle's electronic control unit is collected. This data, encompassing engine operating conditions and performance, is subsequently processed using machine learning algorithms. Specifically, various machine learning models are explored to identify correlations between historical data and the emergence of anomalies, which may signal an impending failure.

The system has been successfully applied to a real-world case study at Comet S.r.l., a logistics and goods handling company in Messina. The results of the tests demonstrate the effectiveness of the system, which can be adopted in more complex sites to manage the risk in the intermodal transport of hazardous materials, in particular, to support the prevention of highly dangerous releases of substances.

* 1. Introduction

The safe and efficient operation of industrial vehicles, particularly those handling hazardous materials, is fundamental. Unplanned downtime due to mechanical failures can lead to significant production losses, increased costs, and, most critically, pose severe risks to personnel and the environment, especially when dealing with potentially dangerous substance releases. Traditional maintenance strategies, often reactive or based on fixed schedules, are proving inadequate in today's complex industrial landscape. This necessitates a shift towards predictive maintenance, leveraging the power of machine learning and the Internet of Things (IoT). By integrating IoT architectures for real-time data acquisition from vehicle sensors and employing machine learning algorithms to analyse this data, we can move beyond reactive repairs and proactively predict potential failures. This approach allows for timely interventions, preventing catastrophic events, minimizing downtime, optimizing maintenance schedules, and ultimately enhancing the safety and reliability of industrial vehicle operations, particularly in scenarios where the consequences of failure could be disastrous.

While the field of predictive maintenance (PdM) has seen extensive review, our specific focus on leveraging IoT and machine learning for the real-time monitoring of industrial vehicles, particularly those handling hazardous materials, necessitates a refined approach. Existing literature provides a strong foundation, but often lacks the required integrated perspective. Recently Elkateb et al. (2024) demonstrated the potential of the combination of machine learning algorithms and IoT obtaining appreciable results in the field of PdM of the textile industry. Wu et al. (2011) and Werbińska-Wojciechowska (2019) provided overviews of preventive maintenance models, with the latter specifically addressing technical systems, establishing the groundwork for the exploration into industrial vehicle maintenance. Sutharssan et al. (2015) and Peng et al. (2010) categorized prognostics and health monitoring (PHM) methods, highlighting the relevance of data-driven methods, including machine learning (ML), which is central to the proposed system of this work. The automotive domain presents a distinct set of challenges, Li et al. (2018) surveyed Artificial Intelligence (AI) applications, primarily in autonomous driving, and Nowakowski et al. (2018) briefly surveyed technical system maintenance, with only one automotive case study. Falcini et al. (2017) and Singh and Arat (2019) discussed Deep Learning (DL) for automotive systems, but not specifically for predictive maintenance, highlighting a gap in the literature regarding the application of these techniques for proactive maintenance in this sector. Ahsan et al. (2016) and Sankavaram et al. (2009), combined PdM, data-driven methods, and automotive applications. Byttner et al. (2011) proposed a novel approach that progressively learns system behavior through exploratory analysis of signals available on the vehicle's internal communication network, by using consensus self-organized models’ approach (COSMO). This approach, further refined by Killeen et al. (2019) with improved sensor selection, aligns closely with our objective of leveraging real-time data from the vehicle's Electronic Control Unit (ECU). More recently, Kuftinova et al. (2024) have developed a methodology based on clustering with K-means algorithm in order to optimize transportation network performance and indirectly reducing operational costs through enhanced predictive maintenance capabilities.

This document will explore the crucial role of predictive maintenance, supported by machine learning and IoT, in mitigating the risks associated with hazardous substance releases from industrial vehicles.

In this work Section 2 introduces the methodology used for the study; Section 3 illustrates the case study and Section 4 presents and discusses the results; finally, Section 5 reports the main conclusions.

* 1. Methodology
     1. Monitoring System

The Monitoring System (MS) is a low-cost hardware solution designed to extract telemetric data from the Electronic Control Units (ECUs) of industrial vehicles. Its modular architecture ensures ease of maintenance and high functional scalability, enabling integration with compatible ECUs with various communication protocols. This approach significantly reduces costs and implementation time.

Immagine che contiene testo, schermata, diagramma, cartone animato

Il contenuto generato dall'IA potrebbe non essere corretto.

Figure 1: Monitoring system architecture

As shown in Figure 1, the MS comprises three main hardware modules: the CanBus Adapter Module (CBA-M), the GPS Module (GPS-M), and the NFC Module (NFC-M). Each module operates independently, performing specific functions. The CBA-M interfaces directly with the industrial vehicle, acquiring engine status data. The GPS-M collects geolocation data (latitude and longitude) and calculates dynamic metrics such as average speed and vehicle heading. Finally, the NFC-M provides the hardware for identifying resources by reading NFC tags, enabling unique driver identification and asset association, such as the type of cargo being transported.

The MS modules adopt a master-slave architecture, as depicted in Figure 2a. This approach separates data acquisition and pre-processing operations from transmission operations, optimizing system efficiency.

In particular, as highlighted in Figure 2b, the slave node of the CBA\_M consists of an Arduino Uno microcontroller and an MCP2515 board, which enables low-level communication with the vehicle's ECU. The node's firmware is designed to support the SAE J1939 communication standard and includes filtering functions

to select and process only specific groups of CANBus messages, reducing the computational load and optimizing information processing. The master node is based on an ESP32 microcontroller with integrated WiFi capabilities. Its primary functions include real-time buffering of CANBus messages and subsequent transmission of this data. The master and slave nodes within each hardware module communicate via an I2C bus, as depicted in Figure 2a. This design choice ensures low-latency information exchange and reliable synchronization of operations.

Immagine che contiene testo, schermata, Ingegneria elettronica, elettronica

Il contenuto generato dall'IA potrebbe non essere corretto.

Figure 2: CANBus Adapter master-slave architecture (a) architecture scheme (b) hardware solution

Figure 3 illustrates IoT architecture integrating the MS. This architecture is structured across three abstraction layers: Edge, Cloud, and Application Layer, enabling a direct virtual connection between the analysis station and the vehicle being monitored. The Edge Layer comprises hardware in direct contact with the vehicle. In addition to the previously described MS, it includes a Control Unit (CU), implemented on a Raspberry Pi, which manages data acquisition and transmission. The CU is equipped with a 4G-LTE card for remote communication and a local storage unit for data archiving in case of connectivity loss. The CU implements two primary services: an MQTT server for managing communication with the MS modules, and a REST API server for interfacing with the analysis station. The Cloud Layer encompasses devices dedicated to networking, data storage, and processing. Finally, the Application Layer provides a set of machine learning strategies for telemetric data analysis, with the goal of implementing a monitoring service for industrial vehicles.

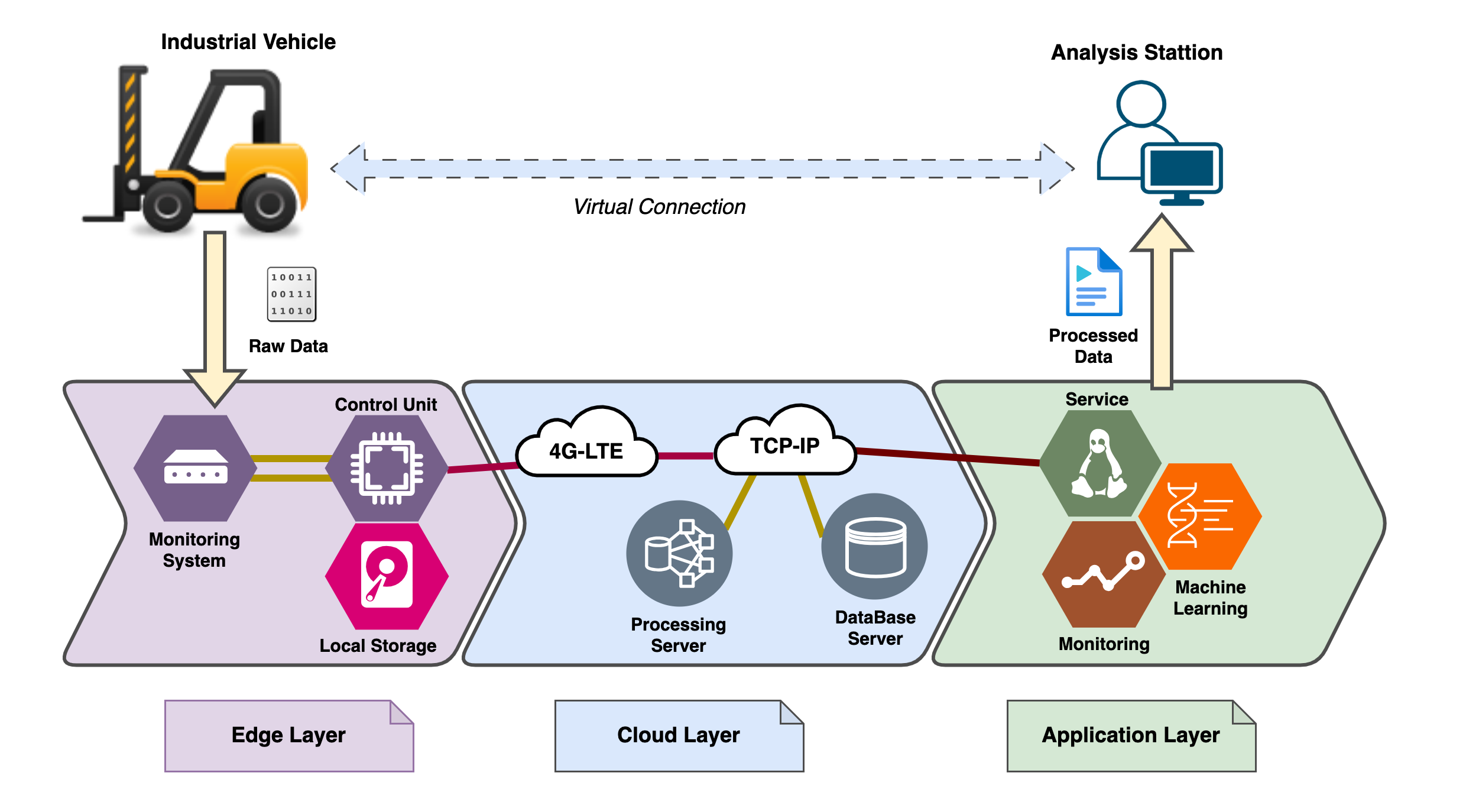


Figure 3: Monitoring system IoT architecture

* + 1. Data Mining

This section outlines a data mining methodology for identifying anomalies within a telemetry signal dataset, leveraging machine learning techniques. Anomaly detection involves discerning anomalous conditions or instances within the dataset. "Anomaly" refers to any event or value that deviates significantly from the expected behavior or established norm. Figure 4 illustrates the methodology proposed in this work. Three main phases can be identified: signal pre-processing, model building, and anomaly detection. Data pre-processing is essential to enhance the quality of the input signals, reduce noise, and convert them into a suitable format for analysis. As shown in Figure 4, pre-processing starts with a filtering operation, implemented using a smoothing window.

Immagine che contiene testo, diagramma, linea, Carattere

Il contenuto generato dall'IA potrebbe non essere corretto.

Figure 4: Data mining methodology

This involves replacing each signal sample with the average value of the points within a defined neighbourhood. The next step is parameterization, which involves extracting representative statistical features from the signal. These parameters enable the signal behaviour to be described quantitatively and its dimensionality to be reduced. Finally, feature selection identifies a subset of relevant parameters for the clustering phase. Model building is based on a training dataset of anomaly-free telemetry signals, representing the average behaviour to be modelled. The dataset is partitioned into clusters using the K-Means algorithm, with the optimal number of clusters determined through an iterative process. In each iteration, the silhouette score is computed, a metric that evaluates both the internal cohesion and separation of the clusters.

To finalize the process, the anomaly identification threshold, denoted as **Z**, is determined by analysing the maximum inter-cluster distance. Given **K** = **{*K1*, *K2*, …, *Km*}** as the set of *m* clusters resulting from the K-Means algorithm, and **C = {*c1*, *c2*, …, *cm*}** as the set of their respective centroids. The inter-cluster distance is defined as the mean Euclidean distance between the points *ph* of ***Ki***, where *h* ranges from 1 to the dimensionality of ***Ki*** (h=1 … *dim*(***Ki***)) and the centroids of all other clusters. The threshold **Z** is then determined according to the following formula:

|  |  |
| --- | --- |
|  | (1) |

The **M** = {**K**, **C**, **Z**} model, comprising the set of clusters **K**, centroids **C**, and threshold **Z**, is employed during the anomaly detection phase to identify atypical patterns within the testing dataset. In this phase, data points are assigned to the clusters of the **M** model through a classification process leveraging the k-Nearest Neighbors (k-NN) algorithm. Let **A = {*A1, A2, …, An*}** be the set of *n* clusters obtained through the k-NN algorithm, anomalies are identified by comparing the minimum of the average Euclidean distances between the points of ***Ax*** (where *x* ranges from 1 to *n*) and the centroids **C**, with the threshold **Z**. Consequently, the set ***Ax*** is considered anomalous if the following condition is met:

|  |  |
| --- | --- |
|  | (2) |

Subsequently, the set of **A** elements, excluding those identified as anomalous, is employed to update the model **M**.

* 1. Case Study

To validating the data mining methodology outlined in Section 2.2, this study examines the case of engine telemetry monitoring from a Comet S.r.l. terminal truck, by a prototype of MS deployed on the vehicle. During normal vehicle operation, the MS was configured to collect data from the Intake Air Temperature (IAT) sensor. The IAT measures the temperature of the air entering the intake system of an internal combustion engine, a critical parameter for controlling combustion and optimizing engine efficiency. In a turbocharged engine equipped with an intercooler, such as the one in the monitored vehicle, the intake air temperature typically ranges from 20°C to 60°C. The acquired dataset was partitioned into two subsets: one for model training and the other one for validation. For each signal in the dataset, the parameters listed in Table 1 were calculated. Subsequently, a subset of independent features was selected for the clustering phase, using the correlation matrix of all calculated parameters.

*Table 1: Features and formulas for dataset signals*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Feature | Formula | Feature | Formula | Feature | Formula |
| *Mean* |  | *RMS* |  | *Peak to Peak* |  |
| *Variance* |  | *Standard Deviation* |  | *Power* |  |
|  |  |  |  |  |  |
| *Crest Factor* |  | *Form Factor* |  | *Pulse Indicator* |  |
|  |  |  |  |  |  |

The identified subset of features includes: mean value, Crest Factor, Form Factor, and Pulse Indicator. During the training phase, the M model, described in Section 2.2, was generated, consisting of m=4 clusters and with a threshold for anomaly detection set at Z=6.80.

To assess the effectiveness of anomaly detection, three signals indicating malfunctions within the air intake system were incorporated into the testing dataset. These signals were derived from a telemetric dataset of a terminal truck, which was the same model as the one under monitoring. Specifically, the signals designated as Trip7 and Trip2 correspond to anomalies detected by the IAT sensor, while the Trip10 signal pertains to an unidentified mechanical failure.

* 1. Results

Figure 5 shows the anomaly detection results. To visualize the data in a 2D space while preserving as much variance information as possible, we applied Principal Component Analysis (PCA) to the dataset, retaining the first two principal components. Specifically, Figure 5a displays the classification output obtained using the K-NN algorithm and formula 1. It's clear how the points corresponding to the three anomalous signals are isolated from the other clusters, as they deviate significantly from the dataset representing normal IAT operation, as reported in detail in Table 2, according to Eq(2). A similar analysis can be performed by examining Figure 5b, which shows the raw and filtered IAT signals. Notably, the signals labelled Trip7 and Trip2 exhibit anomalous behaviour, falling outside the IAT's nominal operating range of 20 to 60 °C. Additionally, the Trip10 signal, while generally within the expected operating range, displays an atypical trend compared to the signals classified as normal.

Immagine che contiene testo, diagramma, linea, Piano

Il contenuto generato dall'IA potrebbe non essere corretto.

Figure 5: Anomaly detection results (a)clustering (b)signals

Table 2: Distances from centroid for all investigated items

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Distance from centroid |  | Thresholding (6.80) |
| Trip1 | 2.29 |  | < |
| Trip2 | 9.32 |  | > |
| Trip6 | 1.41 |  | < |
| Trip7 | 13.65 |  | > |
| Trip10 | 22.78 |  | > |
| Trip15 | 3.78 |  | < |

The model's adaptability to diverse industrial vehicles and operating conditions allows for comprehensive engine diagnosis through multi-sensor signal analysis, with AI-driven signal selection for enhanced efficiency.

* 1. Conclusions

This article demonstrates the effectiveness of a predictive maintenance system for industrial vehicles, particularly those handling hazardous materials. By combining a modular IoT-based Monitoring System (MS) for real-time data acquisition (engine status, location, driver/asset info) with a machine learning-driven anomaly detection methodology, the system successfully identified simulated malfunctions in a case study. Analysis of both clustered and raw/filtered IAT signals clearly demonstrates the system's ability to isolate anomalous signals, including those outside the normal operating range (Trip7 and Trip2) and those exhibiting atypical trends within it (Trip10), showcasing its effectiveness in detecting malfunctions. The results showcase the potential for proactive anomaly detection, enabling timely interventions, minimizing downtime, and improving safety and reliability, thus mitigating the risks of catastrophic failures and environmental damage. This research provides a practical example of how IoT and machine learning can be applied to advance predictive maintenance in complex industrial environments.

References

Ahsan M., Stoyanov S., Bailey C., 2016, Prognostics of automotive electronics with data driven approach: A review. In: 39th international spring seminar on electronics technology, 279–284.

Byttner S., Rognvaldsson T., Svensson M., 2011, Consensus self-organized models for fault detection (COSMO). Engineering Applications of Artificial Intelligence, 24, 833-839.

Elkateb S., Métwalli A., Shendy A., Abu-Elanien A.E.B., 2024, Machine learning and IoT – Based predictive maintenance approach for industrial applications. Alexandria Engineering Journal, 88, 298-309.

Falcini F., Lami G., Costanza A., 2017, Deep learning in automotive software. IEEE Software, 34, 56–63.

Li J., Cheng H., Guo H., Qiu S., 2018, Survey on artificial intelligence for vehicles. Automotive Innovation, 1, 2-14.

Killeen P., Ding B., Kiringa I., Yeap T., 2019, IoT-based predictive maintenance for fleet management, Procedia Computer Science, 151, 607-613.

Kuftinova N.G., Ostroukh A.V., Maksimychev O.I., Pronin C.B., 2024, Efficient Machine Learning Methods for Real-Time Transport System Optimization and Predictive Maintenance. 2024 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex.

Nowakowski T., Tubis A., Werbińska-Wojciechowska S., 2018, Evolution of technical systems maintenance approaches–review and a case study. In: International conference on intelligent systems in production engineering and maintenance. Springer, 161–74.

Peng Y., Dong M., Zuo M., 2010, Current status of machine prognostics in condition-based maintenance: A review. International Journal of Advanced Manufacturing Technology, 50, 297–313.

Sankavaram C., Pattipati B., Kodali A., Pattipati K., Azam M., Kumar S., Pecht M., 2009, Model-based and data-driven prognosis of automotive and electronic systems. In: 5th Annual IEEE international conference on automation science and engineering, 96–101. [h](http://dx.doi.org/10.1109/COASE.2009.5234108)

Singh K. B. and Arat M.A., 2019, Deep learning in the automotive industry: Recent advances and application examples. Engineering, Computer Science, arXiv preprint.

Sutharssan T., Stoyanov S., Bailey C., Yin C., 2015, Prognostic and health management for engineering systems: a review of the data-driven approach and algorithms. The Journal of Engineering, 7, 215–222.

Werbińska-Wojciechowska S., 2019, Preventive maintenance models for technical systems. In: Technical system maintenance: delay-time-based modelling. Cham: Springer International Publishing, 21–100.

Wu S., 2011, Preventive maintenance models: A review. In: Tadj Lotfi, Ouali MSalah, Yacout Soumaya, Ait-Kadi Daoud, editors. Replacement models with minimal repair. London: Springer London, 129–40.