Clustering-Based Anomaly Detection for Connected Trucks

Abdulkadir Bilge[1], Erdem Ergen[1], Onur Ciliz[1], Aslihan Reyhanoglu[1], Feyzi Ege Kumec[1],
Hira Bakirhan[2], Bugra Turan[3], and Sinem Coleri [3]
[1]Koc University Ford Otosan Automotive Technologies Laboratory (KUFOTAL), Sariyer, Istanbul, Turkey, 34450
[2]Ford Otosan, Sancaktepe, Istanbul, Turkey, 34885
[3]Department of Electrical and Electronics Engineering, Koc University, Sariyer, Istanbul, Turkey, 34450
E-mail: [abilge20, eergen20, ociliz21, areyhanoglu, fkumec, buturan, scoleri] @ku.edu.tr and hertugru@ford.com.tr

***Abstract*—Proactive fault detection and anomaly detection are among key enablers of autonomous driving technologies provisioned to reduce operational downtime and increase reliability. Unsupervised learning methods are utilized to detect patterns in large datasets for clustering. In this study, we demonstrate clustering-based anomaly detection using engine-related onboard telemetry from connected trucks. First, we prepare the data by missing value removal, relevant feature selection, and data standardization in the preprocessing step. Then, we apply five clustering methods: K-Means, Isolation Forest, Z-Score Analysis, Gaussian Mixture Model (GMM) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Finally, the silhouette score, Calinski-Harabasz index, and domain-specific thresholds are employed to validate the models' performance in anomaly detection. The results demonstrate that the best performing clustering method is Z-Score.**

Keywords*— Connected trucks**; ***anomaly detection; ***unsupervised learning***; ***clustering***; ***vehicle telemetry***.

I. Introduction

Anomaly detection in connected trucks is essential for predictive maintenance and enhanced fleet efficiency, leveraging vehicle diagnostics, geolocation, driving behavior, and weather data to extract actionable insights despite large, complex datasets. Machine learning models predict maintenance needs, detect anomalies, and optimize operations for safer, more cost-effective fleet management [1]–[3].

In the automotive industry, corrective maintenance is employed to fix component failure [4]. On the other hand, Predictive Maintenance (PdM) is a novel concept that focuses on predicting potential vehicle failures before they occur by analyzing vehicle data [5]. PdM reduces unplanned downtime, leading to fewer unexpected repairs, increasing efficiency and user availability. Unsupervised learning techniques offer robust and effective anomaly detection algorithms to handle complex and large-scale datasets in the context of connected truck data analysis. They provide an optimal framework for the design of anomaly detection algorithms to identify anomalies in vehicle data, improve predictive maintenance, and optimize fleet efficiency [6]. Unsupervised methods are commonly preferred for pattern recognition and outlier identification [7].

Building on these methodological insights, recent studies have applied unsupervised learning and anomaly detection techniques across various domains within the automotive industry, including driving behavior anomaly detection [8], Electric Vehicle (EV) battery anomaly detection [9], sensor data abnormality monitoring [10], intravehicular communication abnormality [11] and cybersecurity-intrusion protection [12] [13]. The authors in [8] compare the performance of supervised (Support Vector Machine (SVM), k-Nearest Neighbor (KNN)) and unsupervised (isolation forest [14], Local Outlier Factor (LOF), Z-score [15]) approaches for detecting behavioral abnormalities. Their study utilizes Principal Component Analysis (PCA) and the Minimum Covariance Determinant (MCD) methods for dataset analysis. They conclude that the MCD algorithm demonstrated remarkable results regarding the metrics of accuracy, F1-score, and Mean Absolute Error (MAE). [9] introduces an anomaly detection framework for EV batteries to avoid problems such as thermal runaway and overheat, based on time series analysis. They report that their framework outperforms traditional anomaly detection methods, including PCA, KNN and autoencoders. [10] develops a univariate time series data driven anomaly detection model to detect anomalies in time series vehicle sensor data. They conclude that isolation forest executes the best performance for unsupervised anomaly detection on time series vehicle sensor data. [16] proposes a two-stage approach for radio resource management in Cellular Vehicle-to-Everything (C-V2X) networks. In the first stage, DBSCAN is utilized to cluster vehicles based on their geographical locations, predicted positions, and speeds. The study demonstrates that using DBSCAN for vehicle clustering enhances spectral efficiency and optimizes resource allocation in vehicular networks. In [17], the authors propose an innovative anomaly detection algorithm, namely the Long Short Term Memory (LSTM) Autoencoder with GMM, to detect anomalous behavior in Connected and Autonomous Vehicle (CAV) trajectories. Previously mentioned studies lack a comprehensive, multi-algorithm approach that integrates diverse vehicle subsystems, domain-specific thresholds, and robust preprocessing for heterogeneous vehicle sensor data, limiting their effectiveness in general-purpose, real-world predictive maintenance.

This paper benchmarks multiple unsupervised anomaly detection methods on heterogeneous, real-world engine-related connected-truck telemetry within a unified evaluation pipeline. The approach leverages unsupervised learning techniques, specifically K-Means clustering [18], Isolation Forest, Z-score analysis, GMM and DBSCAN to identify abnormal patterns and potential system faults.

The paper is organized as follows: Section II describes the

 ISBN: 978-1-68558-348-4

dataset and preprocessing. Section III presents the anomaly detection algorithms. Section IV provides the performance evaluation and results. Section V concludes the paper with remarks and possible future work.

II. System Model

This section outlines the system model for detecting anomalies in engine-related connected-truck telemetry data.

The dataset used in this study is collected over a period of 14 consecutive days from two identical connected trucks. The entire dataset includes 90 features and 952815 data points. The data is off-boarded from the truck's onboard telemetry system, which captures a comprehensive set of features that reflect its operational behavior and performance. Each feature is recorded with its own specific frequency, creating a heterogeneous dataset requiring careful alignment and synchronization during preprocessing. Data challenges include missing values due to sensor failures and irregular sampling rates stem from features being logged at different intervals. These issues are meticulously addressed during the preprocessing stage to ensure the reliability and suitability of the data set for anomaly detection.

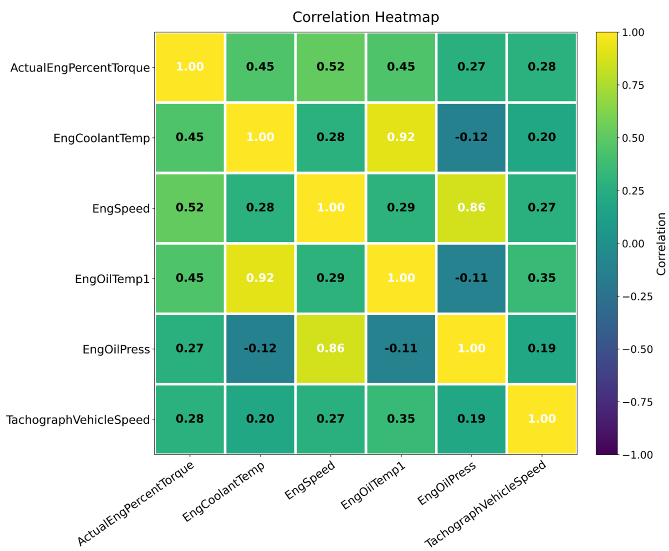

Figure 1. Correlation Heatmap of Vehicle Health Features.

Data preprocessing is a critical step in preparing the collected dataset for analysis. Preprocessing techniques such as standardization, null removal and data type conversion are applied to ensure usability for anomaly detection. Feature correlation is extracted by using a heatmap as in Figure 1. For instance, *EngCoolantTemp* and *EngOilTemp1* exhibit a strong positive correlation (0.92), indicating that engine load and thermal dynamics influence both parameters simultaneously. Similarly, *EngSpeed* and *EngOilPress* display a high correlation (0.86), reflecting the mechanical relationship between engine speed and oil pressure. Features with a correlation higher than 0.3 with any other feature is selected. The selected features are; actual engine torque percent (*ActualEngPercentTorque*), engine coolant temperature (*EngCoolantTemp*), engine speed (*EngSpeed*), engine oil temperature (*EngOilTemp1*), engine oil pressure (*EngOilPress*), tachograph vehicle speed (*TachographVehicleSpeed*). 43858 samples were obtained from 6 features in total by data preprocessing. The data is normalized through a standard scaler by setting $\sigma = 1$ and $\mu = 0$, for each feature. The preprocessed data is fed into the K-Means, Isolation Forest, Z-Score, GMM and DBSCAN algorithms to perform anomaly detection.

III. Anomaly Detection Algorithms

We implement five different algorithms to analyze the preprocessed dataset for anomaly detection: K-Means Clustering, Isolation Forest, Z-Score Analysis, DBSCAN and GMM Clustering.

K-Means clustering is initially utilized to obtain a basic set of clusters. The clustering results are analyzed using PCA, which reduces the high-dimensional data to two components for easy interpretation. Data points deviating farther than 2 standard deviations from the mean distance of each cluster are flagged as anomalies. This method's suitability lies in its ability to capture distinct operational patterns, such as idling or high-speed driving, and effectively highlight deviations from these norms.

Isolation Forest, a machine learning algorithm specifically designed for anomaly detection, is applied to the same dataset. It isolates anomalies by constructing random decision trees and measuring path lengths. Its ability to handle high-dimensional and diverse feature distributions makes it particularly effective for identifying unusual engine states or sensor malfunctions. The algorithm's parameters, such as contamination level, number of estimators, max samples and max features are optimized through grid search. The grid for the number of estimators consists of the values 100, 200, 500. The grid for the maximum samples consists of automatic selection, 50%, 75% of the dataset. The grid for the contamination factor consists of 1%, 5%, 10% of the data. The grid for the maximum feature at each tree consists of 100%, 50% of the features. Isolation Forest is utilized, along K-Means to detect anomalies in the vehicle sensor data by isolating outlier instances based on randomly constructed decision trees, making it suitable for identifying rare and irregular driving patterns or sensor behaviors.

Z-Score Analysis provides a statistical perspective by calculating the number of standard deviations each data point deviates from the mean. Z-Score anomalies are primarily observed in features with normal distributions, such as engine coolant temperature and oil pressure, to identify values outside typical operating ranges. A threshold of three standard deviations is used to flag extreme outliers. Such a threshold is chosen because approximately 99.7% of data points lie within 3 standard deviations from the mean.

GMM, a probabilistic clustering technique that represents each observation as arising from a mixture of multivariate Gaussian distributions with their own means, covariances and standard deviations, is applied to the vehicle dataset.

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-348-4

This flexibility in modeling both the central tendencies and the spread of correlated features (e.g., engine temperature, oil pressure, speed) makes GMM particularly effective for capturing overlapping driving modes and for identifying subtle deviations indicative of early-stage faults. The parameters for GMM is selected through grid search. The grid for the number of components consists of 2, 3, 4, 5, 6. The grid for the covariance matrix type consists of full, tied, diagonal and spherical. The best results are obtained with 2 components and a tied covariance matrix.

DBSCAN is a density-based clustering algorithm that groups together points in high-density regions while marking points in low-density regions as noise. In the context of vehicular data clustering, DBSCAN excels at discovering clusters of typical operational states (cruising, idling, or acceleration patterns) without requiring the number of clusters a priori. Through grid search, the best DBSCAN results are obtained with an epsilon value of 1.120833 and minimum sample value of 15. The grid for epsilon consists of 50 linearly spaced values running from 0.1 to 5.0. The grid for the minimum samples consists of the values 3, 5, 10, 15.

K-Means and GMM are utilized to capture the global structure of the dataset. DBSCAN and Isolation Forest spot local and high-dimensional irregularities; and Z-Score provides a transparent, per-signal sanity check. Their ensemble leverages diverse detection philosophies such as centroid distance, density isolation, distribution probability, tree-based partitioning, and statistical thresholding. Combining the strengths of each algorithm proves to be far more robust than using any single method.

IV. Performance Evaluation

The data is preprocessed and analyzed using Python and libraries such as Scikit-Learn, NumPy, SciPy, Matplotlib, Optuna, Seaborn and Pandas. The experiments are run in an NVIDIA Drive AGX Orin device. The analysis employed K-Means Clustering, Isolation Forest, Z-Score, DBSCAN and GMM methods to identify anomalies in connected truck telemetry data. Anomalies are defined as indicators of extreme aggression in vehicle driving behavior or internal vehicle mechanics that could lead to future vehicle failures. For instance, a data point with an *ActualEngPercentTorque* value of 94.0, a *EngCoolantTemp* value of 100.0, a *EngSpeed* value of 1386.0, a *EngOilTemp1* value of 119.65625, a *EngOilPress* value of 300.0, a *TachographVehicleSpeed* value of 38.210938 is flagged as an anomaly, due to having *ActualEngPercentTorque* value higher than the threshold of 90 and a *EngOilTemp1* value of 119.65625 higher than the threshold of 110.

A. Evaluation Metrics

To assess the effectiveness of the anomaly detection algorithms, several evaluation metrics are employed. These metrics are selected to provide insights into the clustering performance, agreement between methods, and the ability of each algorithm to identify meaningful anomalies within the dataset.

The performance of Z-Score Analysis and Isolation Forest is further validated by comparing their flagged anomalies against domain-specific thresholds for engine-related features. The threshold for *ActualEngPercentTorque* is set at 90%, similar to [19], which maps a John Deere 4.5 L engine's torque curve and identified that 90 % of rated torque corresponds to a high-load operating point. The threshold for the variable *EngCoolantTemp* is chosen to be 100 degrees Celsius. [20] proves that temperatures approaching or exceeding 100 degrees Celsius precipitate notable performance losses and component damage. *EngSpeed* as a threshold is set to 2500 revolutions per minute. In [21], the upper bound for a Caterpillar 3126B engine is found to be 2300 revolutions per minute, for emissions purposes. Since our work focuses on testing heavy duty vehicle load, we have set the threshold to be slightly higher than the 2300 rpm limit. *EngOilTemp1* is set to have a threshold of 110 degrees Celsius. [22] shows that a heavy-duty single-cylinder diesel engine achieved its lowest specific fuel consumption 83 degrees Celsius and 88 degrees Celsius. Their work argues that beyond 95 degrees Celsius, engine oil consumption begins to increase again. Industry-standard ageing tests show that key anti-wear additives degrade markedly once oil temperatures exceed 110 degrees Celsius [23]. *EngOilPress* is set at 600 kPa as a threshold. The value is set at such a level according to [24]. In the work, the authors have set their maximum measurement interval between 0-700 kPa for the engine oil pressure, where the majority of the points lie within 0-600 kPa. The points above 600 kPa are thus treated as anomalies in our work. *TachographVehicleSpeed* is set at 90 kmh. The threshold was chosen according to [25], where they demonstrate that carbon monoxide and soot spikes occurred in the interval of 100–130 kmh of vehicle speed. Since high carbon monoxide and soot generation of the system means erroneous operation, a speed value of 90 km/h serves as a good barrier for prevention. The silhouette score is used to evaluate the quality of the K-Means clustering. This metric measures the cohesion and separation of clusters by comparing the average distance of points within a cluster to the average distance to points in the nearest cluster [26]. A higher silhouette score indicates well-defined and distinct clusters.

B. Evaluation Setup

The Calinski-Harabasz Index is utilized as an internal validation metric to assess the clustering quality in the K-Means algorithm. This index evaluates the ratio of dispersion between groups to dispersion within groups, effectively measuring how well defined and compact clusters are [27]. A higher Calinski-Harabasz score indicates that the clusters are both tightly packed and well separated from each other. Furthermore, [28] highlights the index's robustness in evaluating clustering performance across varying datasets and its sensitivity to data separability, making it a reliable metric for assessing the clustering model's effectiveness. This metric complements the silhouette score by providing an additional perspective on the effectiveness of the clustering model in distinguishing between normal and anomalous behaviors within the dataset.

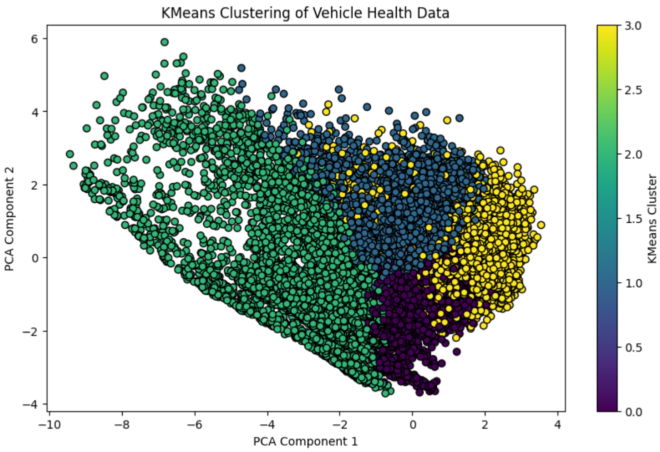

Figure 2. PCA with 2 Components Through K-Means.

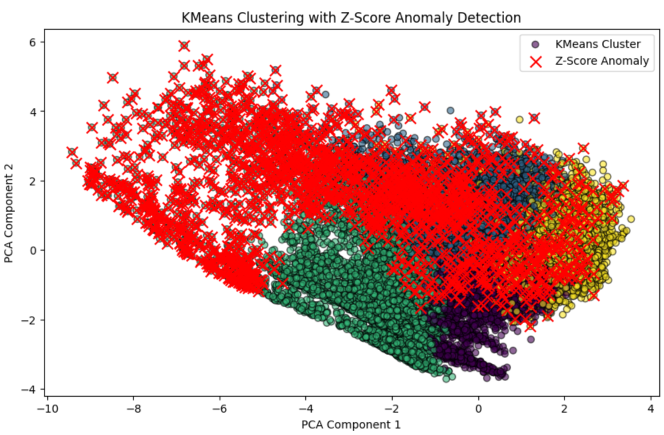

Figure 3. PCA Component Analysis Through K-Means and Z-Score Anomaly Detection.

During K-Means analysis, four clusters are initially utilized to capture major operational modes of the vehicle; idle, cruise, acceleration, deceleration. Each operational mode has its own level of engine load and thus needs its own analysis. Utilizing K-Means clustering and PCA, the clusters in Figure 2 are obtained. Each color in the cluster diagram denotes a unique mode of vehicle operation. Then, obtained clusters are relayed to the Z-Score algorithm, for further processing of anomalies and detection thereof. The Z-Score algorithm is then run with a Z-Score threshold value of 3. The resulting anomaly detection plot is given in Figure 3, where the cross-markers denote the anomalies. The Isolation Forest algorithm is then run with a contamination factor of 0.05 and a hundred estimators, to further detect anomalies or modify the existing anomalies.

Concurrently, the GMM is utilized. The pre-processed data is further processed for utilizing in GMM. Sorting, windowing and sampling operations are done on data before utilizing GMM. Sorting is done as per the timestamp of the data. For windowing, data is first resampled by splitting the data into contiguous, non-overlapping 60-second bins, for each feature. Then, the data is windowed to bins of separate statistic types, namely the mean, the standard deviation, the minimum, the maximum of each 60 second period, for each feature. Then, columns with an 80 percent of missing values or more are dropped. After windowing is completed, the data is imputed and scaled via a standard scaler. The GMM algorithm is run with a tied covariance matrix and 2 components. Similarly, after hyperparameter optimization, the DBSCAN algorithm is run with an epsilon value of 1.121, minimum sample value of 15.

C. Evaluation Results

In method evaluation, the silhouette score confirms the suitability of the chosen number of clusters, reflecting clear separations between operational patterns in the truck's telemetry data. On the other hand, the Calinski-Harabasz Index validates the clustering solution by confirming the distinctiveness of the operational patterns identified in the truck's telemetry data.

TABLE I.
COMPARISON OF CLUSTERING AND ANOMALY-DETECTION METHODS ON VEHICLE TELEMETRY DATA.

METHOD	SILHOUETTE	CALINSKI–HARABASZ
K-Means	0.33	**19281.7**
Isolation Forest	0.473	671
GMM	0.513	6379.11
DBSCAN	0.297	5.916
Z-Score	**0.569**	7777.99

Z-Score analysis resulted in the highest silhouette score of 0.569 and a strong Calinski–Harabasz Index of 7777.99, reflecting its straightforward yet precise anomaly detection. GMM followed with a silhouette score of 0.513 and an index of 6379.11, demonstrating robustness in modeling both central tendencies and spread. Isolation Forest achieved a silhouette of 0.473 but a low Calinski–Harabasz Index of 671, indicating its specialization in outlier isolation over cluster cohesion. K-Means showed moderate performance with a silhouette score of 0.33, Calinski-Harabasz index 19281.7 but was less effective at capturing anomalies. DBSCAN's density-based grouping yielded a low silhouette score of 0.297 and an index of 5.916, highlighting its difficulty in forming cohesive clusters for this dataset.

V. Conclusions and Future Work

This paper presents a unified benchmarking framework for anomaly detection on engine-related connected-truck onboard telemetry using K-Means, Isolation Forest, Z-Score analysis, GMM, and DBSCAN. The results indicate that Z-Score achieved the best overall performance on the studied dataset, providing a transparent and effective per-signal sanity check under domain-informed thresholds. K-Means was useful for separating major operational modes, while Isolation Forest complemented the analysis by isolating rare, high-dimensional outliers. Although DBSCAN can explicitly label low-density samples as noise and discover non-spherical structures, its single-density assumption may lead to fragmented clusters when the data exhibit varying densities. In contrast, GMM provides a probabilistic soft-assignment model that can represent overlapping operating regimes and correlated feature structures, which is valuable for capturing subtle deviations.

Future work will focus on (i) validating the proposed approach on longer time horizons and larger fleets to assess robustness under seasonal and operational variability, (ii) extending the pipeline toward online/streaming anomaly detection to enable real-time edge deployment, and (iii) incorporating weak supervision from maintenance logs or diagnostic trouble codes to calibrate thresholds, reduce false positives, and improve interpretability. In addition, exploring temporal modeling and change-point detection on multivariate telemetry may further enhance early-fault sensitivity beyond point-wise outlier scoring.

REFERENCES

[1] A. Giannoulidis, K. Karras, G. Giannakopoulos, and V. Vassalos, "Exploring unsupervised anomaly detection for vehicle predictive maintenance with partial information," in *Proceedings of the 27th International Conference on Extending Database Technology (EDBT)*, 2024, pp. 753–761, paestum, Italy. [Online]. Available: https://openproceedings.org/2024/conf/edbt/paper-200.pdf

[2] I. J. Chukwudi *et al.*, "An ensemble deep learning model for vehicular engine health prediction," *IEEE Access*, vol. 12, pp. 63 433–63 451, 2024.

[3] O. Gheibi, D. Weyns, and F. Quin, "Applying machine learning in self-adaptive systems: A systematic literature review," *ACM Transactions on Autonomous and Adaptive Systems*, vol. 15, no. 3, pp. 1–37, 2021. [Online]. Available: https://dl.acm.org/doi/10.1145/3469440

[4] J. Selech and K. Przystupa, "Reliability analysis for unrepairable automotive components," *Applied Sciences*, vol. 11, no. 22, p. 10863, 2021.

[5] F. Arena, M. Collotta, L. Luca, M. Ruggieri, and F. G. Termine, "Predictive maintenance in the automotive sector: A literature review," *Mathematical and Computational Applications*, vol. 27, no. 1, p. 2, 2022. [Online]. Available: https://www.mdpi.com/2297-8747/27/1/2

[6] Y. Cherdo, B. Miramond, A. Pegatoquet, and A. Vallauri, "Unsupervised anomaly detection for cars can sensors time series using small recurrent and convolutional neural networks," *Sensors*, vol. 23, no. 11, p. 5013, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/11/5013

[7] R. Domingues, M. Filippone, P. Michiardi, and J. Zouaoui, "A comparative evaluation of outlier detection algorithms: Experiments and analyses," *Pattern Recognition*, vol. 74, pp. 406–421, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320317303916

[8] K. Demestichas, T. Alexakis, N. Peppes, and E. Adamopoulou, "Comparative analysis of machine learning-based approaches for anomaly detection in vehicular data," *Vehicles*, vol. 3, pp. 171–186, 2021.

[9] W. Jia and C. Yang, "A framework for multi-variate time series anomaly detection based on behavioral patterns in electric vehicles," in *2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Dali, China, 2022, pp. 20–24.

[10] C. Derse *et al.*, "An anomaly detection study on automotive sensor data time series for vehicle applications," in *2021 Sixteenth International Conference on Ecological Vehicles and Renewable Energies (EVER)*, Monte-Carlo, Monaco, 2021, pp. 1–5.

[11] C. H. M. Chandralekha, H. M. Chandrashekar, P. S. Nijesh, S. P. S. Pai, and M. K. Ghosh, "Anomaly detection in recorded can log using dbscan and lstm autoencoder," in *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*. Bangalore, India: IEEE, 2022, pp. 1–7.

[12] F. Fenzl, R. Rieke, Y. Chevalier, A. Dominik, and I. Kotenko, "Continuous fields: Enhanced in-vehicle anomaly detection using machine learning models," *Simulation Modelling Practice and Theory*, vol. 105, p. 102143, 2020.

[13] S. Park and J.-Y. Choi, "Hierarchical anomaly detection model for in-vehicle networks using machine learning algorithms," *Sensors*, vol. 20, no. 14, p. 3934, Jul 2020.

[14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422.

[15] Z.-G. Zhou and P. Tang, "Continuous anomaly detection in satellite image time series based on z-scores of season-trend model residuals," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 3410–3413.

[16] H. Davoudi, B. S. Ghahfarokhi, N. Moghim, and S. Shetty, "Dbscan-based vehicle clustering and uav placement for noma-based resource management in cellular v2x communications," *arXiv preprint arXiv:2504.21656*, 2025. [Online]. Available: https://arxiv.org/abs/2504.21656

[17] B. Wang, W. Li, and Z. H. Khattak, "Anomaly detection in connected and autonomous vehicle trajectories using lstm autoencoder and gaussian mixture model," *Electronics*, vol. 13, no. 7, p. 1251, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/7/1251

[18] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/8/1295

[19] R. A. Rohrer, J. D. Luck, S. K. Pitla, and R. Hoy, "Evaluation of the accuracy of machine reported can data for engine torque and speed," *Transactions of the ASABE*, vol. 61, no. 5, pp. 1547–1557, 2018.

[20] J. Adler and T. Bandhauer, "Performance of a diesel engine at high coolant temperatures," *Journal of Energy Resources Technology*, vol. 139, no. 6, p. 062203, 2017.

[21] T. Huai *et al.*, "Analysis of heavy-duty diesel truck activity and emissions data," *Atmospheric Environment*, vol. 40, no. 13, pp. 2333–2344, 2006.

[22] S. Pukalskas, A. Pikūnas, and G. Bureika, "The influence of diesel engine's thermal condition on fuel consumption and smoke of exhaust gases," *Journal of KONES: Powertrain and Transport*, vol. 16, no. 3, pp. 319–324, 2009.

[23] ASTM International, "Astm d7528-09: Standard test method for bench oxidation of engine oils by robo apparatus," ASTM International, West Conshohocken, PA, Standard D7528-09, 2009.

[24] A. Grzesiek, R. Zimroz, P. Śliwiński, N. Gomolla, and A. Wyłomańska, "A method for structure breaking point detection in engine oil pressure data," *Energies*, vol. 14, no. 17, p. 5496, 2021.

[25] J. Gao, H. Chen, K. Dave, J. Chen, and D. Jia, "Fuel economy and exhaust emissions of a diesel vehicle under real traffic conditions," *Energy Science & Engineering*, vol. 8, no. 5, pp. 1781–1792, 2020. [Online]. Available: https://doi.org/10.1002/ese3.632

[26] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748.

[27] S. Łukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Clustering using flower pollination algorithm and calinski-harabasz index," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 2724–2728.

[28] X. Wang and Y. Xu, "An improved index for clustering validation based on silhouette index and calinski-harabasz index," *IOP Conference Series: Materials Science and Engineering*, vol. 569, no. 5, p. 052024, jul 2019.