Efficient Data Fusion Framework for Real-Time Monitoring of Air Pressure Systems in Scania Trucks

Durdona Bakhodirova

Department of International Business Management, Tashkent State University of Economics, Uzbekistan

Email: d.bahodirova.tsue.uz

Abstract

Monitoring air pressure systems in heavy-duty vehicles such as Scania trucks is a key driver for operational safety and efficiency in the automotive industry. However, the complex interaction of sensors and data sources makes it difficult to quickly detect potential system failures. This problem is solved in our paper where we present a special-purpose data fusion framework for real-time monitoring of Scania trucks' air pressure systems. To achieve this, PCA is used to reduce the size of the dataset followed by a voting classifier which combines diverse models such as Decision Trees, Random Forests, Naive Bayes, and Linear Regression using ensemble learning. In particular, our comparative analysis shows that the Voting Classifier outperforms other ML methods in terms of prediction accuracy. These findings suggest that our fusion framework can be utilized for the early detection of air pressure anomalies in heavy-duty vehicles enhancing their safety record.

Keywords: Data Fusion; Real-Time Monitoring; Air Pressure System Analysis; Scania Truck Diagnostics Fault Detection ;Vehicle Health Monitoring.

1. Introduction

The modern automotive industry is continuously pursuing improvements in vehicle safety, reliability, and performance. The air pressure systems for heavy-duty vehicles like the Scania trucks are among the important components that make sure they work safely [1-3]. They help in maintaining tire pressure which is a fundamental part of the roadworthiness and stability of vehicles. Nevertheless, it still remains a big challenge to have these systems working optimally due to the many sensors involved, different sources of data, and operational situations [4-6]. For this reason, there has been an extensive search for more effective ways to monitor and detect failures as they occur immediately within academic circles as well as industry players.

Traditionally, monitoring of vehicles' air pressure systems had relied on the use of only one sensor measurement leading to low accuracy and reliability in fault detection [7]. Recent developments have led to the adoption of data fusion techniques that facilitate the integration and analysis of different sensor information thus providing a more comprehensive picture of how a given system behaves at any given time. In this regard, the effort now shifts towards building an efficient data fusion framework customized specifically for use on Scania trucks with an aim of allowing continuous monitoring in real-time so that any emerging problems can be detected early enough before they become major challenges concerning their air pressure systems [8].

The significance of this research lies in its potential to revolutionize the approach to air pressure system monitoring in Scania trucks. By harnessing advanced data fusion techniques, the proposed framework strives to overcome the limitations of conventional monitoring methods, offering a more robust and proactive solution. Additionally, the integration of real-time monitoring capabilities promises to enhance not only the safety but also the operational efficiency of these heavy-duty vehicles. This research aligns with the industry's aspirations to adopt proactive maintenance strategies, ultimately minimizing downtime and ensuring safer transport operations [9-11].

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

As the automotive landscape evolves, the proposed efficient data fusion framework signifies a crucial step toward predictive maintenance and proactive fault detection in Scania trucks. This research contributes to the broader field of vehicle health monitoring and paves the way for future innovations in real-time monitoring systems.

2. Related Works

This section begins by exploring a diverse array of scholarly works and industry literature encompassing various aspects of vehicle monitoring, data fusion techniques, and fault detection mechanisms. In the realm of heavy-duty vehicle technologies, Alam et al. [10] delved into the domain of fuel-efficient platooning systems, emphasizing their potential impact on enhancing vehicle efficiency. Their work focused on distributed control mechanisms tailored for heavy-duty vehicle platooning, laying the groundwork for advancements in collaborative driving systems. Furthermore, Alam and colleagues [12] extended their research trajectory in fuel-efficient distributed control for platooning, contributing further insights into the realm of cooperative methods to bolster safety and efficiency in freight transportation. Hou et al. [11] conducted a systematic review pertaining to fault detection and diagnosis within air brake systems. Their comprehensive analysis, published in the Journal of Manufacturing Systems, elucidated various fault detection strategies, providing a valuable resource for understanding and addressing air brake system failures. In the domain of anomaly detection frameworks, Huang et al. [13] introduced the Energy-Efficient and Trustworthy Unsupervised Anomaly Detection Framework (EATU), catering to the Industrial Internet of Things (IIoT). Their work focused on enhancing anomaly detection reliability within sensor networks, presenting a promising avenue for robust system monitoring.

Wolf et al. [14] delved into behavior-based control systems for off-road navigation, particularly in the context of Unimog vehicles. Their study showcased the application of control mechanisms to ensure safe and reliable navigation in challenging terrains, shedding light on specialized navigation techniques for commercial vehicles. Theissler et al. [15] explored the realm of predictive maintenance in the automotive industry, emphasizing the role of machine learning (ML). Their work provided valuable insights into predictive maintenance applications, addressing challenges and presenting potential use cases within the automotive sector. Alam et al. [16] revisited the concept of heavy-duty vehicle platooning, emphasizing cooperative methods for augmenting safety and efficiency in freight transportation. Their study underscored the significance of collaborative driving systems in the pursuit of sustainable freight transportation. Dias [17] and Dias with Peltonen [18] investigated the integration of ML Operations (MLOps) with IoT edge devices, presenting open-sourced integration methods. Their research explored the convergence of ML and edge computing, offering insights into potential applications within IoT frameworks. Nacke and Hirschfeld [19] focused on evaluating the implementation areas of Real-Time Location Systems (RTLS) in production at Scania CV AB Oskarshamn. Their study provided a detailed assessment of RTLS integration in production settings, highlighting its potential benefits and challenges. Scapinakis and Garrison [20] examined communication and positioning systems within the motor carrier industry, emphasizing the role of technology in enhancing industry-wide operations. Their study provided a historical perspective on communication and positioning systems in motor carrier operations. Rettore [21] delved into the fusion of vehicular data space, emphasizing the potential impact on smart mobility. His work explored novel approaches to data fusion within vehicular domains, paving the way for advancements in smart mobility solutions. Lima [22] and Lima [23] investigated optimization-based motion planning and predictive control mechanisms for autonomous driving, particularly in heavy-duty construction trucks. Their studies provided experimental evaluations of control systems tailored for autonomous driving applications, contributing insights into the realm of autonomous heavy-duty vehicles.

3. Methodology

This section encompasses a detailed exposition of the structured methodology developed to integrate diverse sensor data and employ data fusion strategies tailored explicitly for the complexities inherent in the air pressure systems of heavy-duty vehicles. The developed framework strategically applies Principal Component Analysis (PCA) to the dataset pertaining to one specific air system within the Scania trucks. This process involves a systematic reduction of the dataset's dimensionality by identifying and extracting the principal components that encapsulate the maximum variance within the data.

The introduction of a variance-covariance matrix $\mathbf{C}$ facilitates Principal Component Analysis (PCA) by enabling the comprehensive analysis of variance and covariance relationships within the dataset.

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

$$C = \frac{1}{M}\sum_{m=1}^{M}\Delta q_m\ \Delta q_m^T = \langle \Delta q\ \Delta q^T\rangle = \{\langle (q_i - \langle q_i\rangle)(q_j - \langle q_j\rangle)\rangle\}, \quad (1)$$

The division of the variance-covariance matrix into two distinct terms elucidates the contribution of the mean distribution's variance-covariance and the fl weighted average of intra-group variance-covariance, offering a comprehensive insight into the distinct sources shaping the overall matrix structure.

$$C = C^{\text{JAM}} + C^{\text{intra}}$$
$$= \left\{\sum_{l=1}^{L} f_l\langle(\langle q_i\rangle_l - \langle q_i\rangle)(\langle q_j\rangle_l - \langle q_j\rangle)\rangle + \sum_{l=1}^{L} f_l\langle(q_i - \langle q_i\rangle_l)(q_j - \langle q_j\rangle_l)\rangle_l\right\} \quad (2)$$

The introduction of an $f\times M$ matrix encompassing the entirety of the dataset allows for a comprehensive representation of the dataset's features across f dimensions and M instances or samples.

$$Q = \{\Delta q_1 \cdots \Delta q_M\}, \quad (3)$$

The matrix **A** can be derived through a matrix product operation, potentially involving the multiplication of matrices or vectors, elucidating a transformation or relationship within the dataset representation.

$$C = \frac{1}{M} QQ^T, \quad (4)$$

The solution involves resolving the standard eigenvalue problem while maintaining the orthonormal condition, where **Q**T represents the transpose of **Q**, ensuring an orthogonal relationship and providing fundamental insights into the dataset's intrinsic characteristics.

$$CV = V\lambda, \quad (5)$$

Upon solving, the outcomes comprise **V**, the eigenvector matrix, **λ**, the eigenvalue matrix, and **I**, the unit matrix, offering essential components that signify the dataset's principal directions, corresponding variances, and an identity matrix, respectively.

$$VV^T = V^T V = I, \quad (6)$$

The comprehensive linear transformation of **Q** utilizing **V** results in a projection matrix **Σ**=(**σ**1⋯**σ**f) that facilitates the projection of data onto the Principal Components (PCs), delineating their respective contributions to the dataset's variance.

$$\sigma_m = V^T \Delta q_m, \quad (7)$$

$$\Sigma = V^T Q. \quad (8)$$

Performing Principal Component Analysis (PCA) via the singular value decomposition (SVD) of **Q** (an $f\times M$ matrix) directly yields a decomposition into three matrices, offering an alternative method to extract the fundamental components of the dataset with inherent relationships.

$$Q = \sqrt{M} V \lambda^{1/2}\, U^T, \quad (9)$$

$$U^T = \frac{1}{\sqrt{M}}\lambda^{-1/2}\Sigma = \frac{1}{\sqrt{M}}\lambda^{-1/2}\ V^T Q. \quad (10)$$

In this context, **λ**1/2 represents an $f\times M$ matrix characterized by zero non-diagonal elements, simplifying the derivation of a specific condition within the analysis.

$$U^T U = I. \quad (11)$$

To gauge the applicability of PCA to a given dataset, the assessment involves examining the contribution of Principal Components (PCs) to the total variance, quantifying their relevance and impact within the dataset through a defined calculation.

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

$$\chi_\alpha = \frac{\lambda_\alpha}{\sum_{\alpha=1}^{f} \lambda_\alpha}. \quad (12)$$

By implementing PCA on the air system data, our framework aims to condense the information while retaining essential patterns and variations present in the dataset. This reduction in dimensionality not only aids in simplifying the dataset but also facilitates a more efficient analysis of the system's behavior and characteristics. The application of PCA to this particular air system data serves as a crucial preprocessing step within our framework, enabling a more streamlined and focused analysis of the system's performance and potential anomalies.

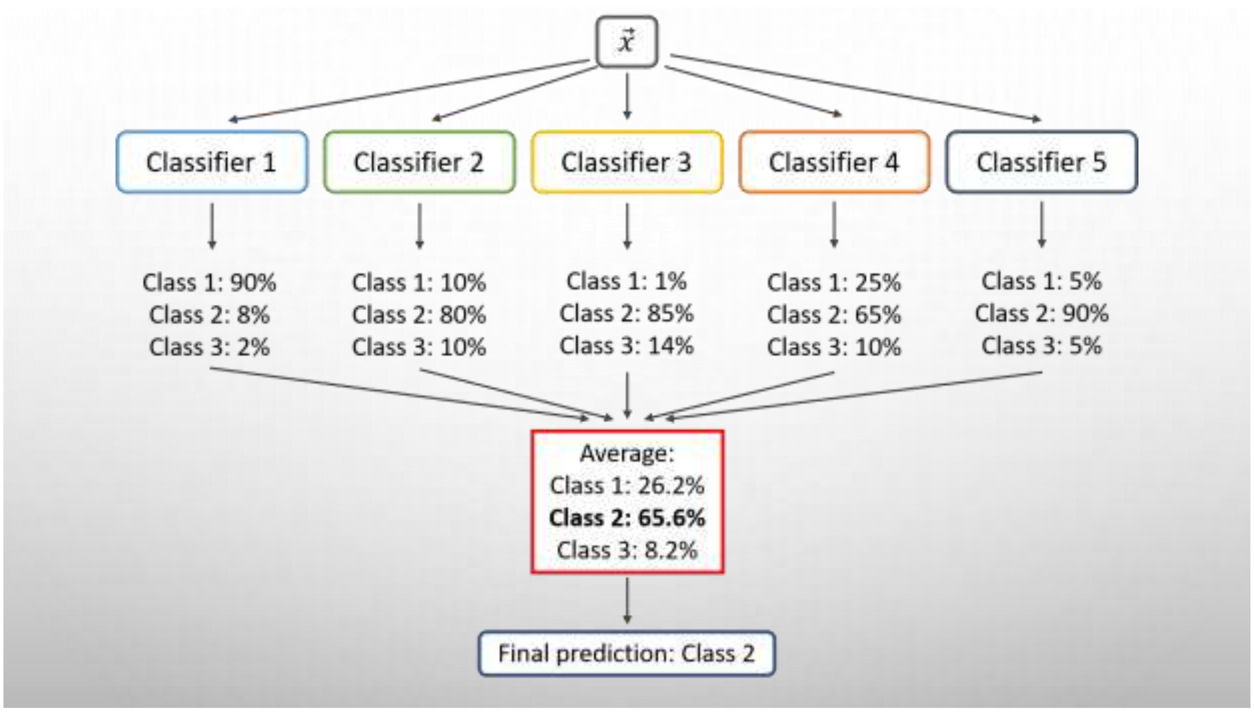

Figure 1: visualization of the concept of voting-based ensemble learning classifier.

In our approach, we leverage ensemble learning techniques via a Voting Classifier to amalgamate the predictive power of diverse base models, including Decision Trees, Random Forests, Naive Bayes, and Linear Regression. The fundamental principle behind ensemble learning involves combining multiple individual models to generate a robust and more accurate prediction collectively. Within the Voting Classifier framework, each base model is trained independently on the dataset, and during the prediction phase, their outputs are aggregated based on various strategies such as majority voting (hard voting) or weighted averaging (soft voting). This amalgamation of predictions from diverse models enhances the overall predictive performance by leveraging the strengths and mitigating the weaknesses of individual models. The steps involved in applying the Voting Classifier encompass:

- Base Model Training: Each individual model—Decision Trees, Random Forests, Naive Bayes, and Linear Regression—is trained on the dataset separately, capturing distinct patterns and relationships within the data.
- Voting Mechanism: During the aggregation phase, the Voting Classifier combines the predictions from these base models using a predetermined strategy. In hard voting, the final prediction is determined by the majority

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

vote among the individual models. In contrast, soft voting incorporates weighted averaging of probabilities assigned by each model, providing a more nuanced prediction.

- Ensemble Prediction: The aggregated predictions derived from the base models via the Voting Classifier are considered the ensemble prediction, which often demonstrates superior performance compared to any single base model.

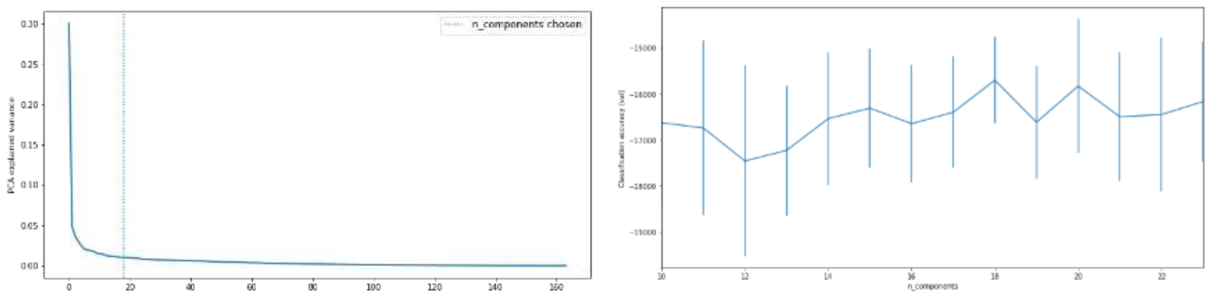

Figure 2: Visualization depicting the Relationship between Principal Component Analysis (PCA) Components and Explained Variance (left) alongside the Impact of Component Numbers on Classification Accuracy (right).

Upon employing this ensemble approach, our framework harnesses the diverse strengths of each model type—ranging from tree-based approaches to regression and probabilistic methods—culminating in a more robust and accurate predictive system capable of handling varying complexities and nuances present within the dataset.

4. Results and Discussion

This section encapsulates the culmination of rigorous experimentation, where the framework's efficacy and performance were rigorously evaluated under various operational conditions and stress scenarios. The results presented herein encapsulate a detailed account of the system's responsiveness, accuracy in fault detection, and ability to provide timely insights into potential anomalies within the air pressure systems.

In Figure 2, an insightful exploration of the Principal Component Analysis (PCA) unfolds, showcasing two distinctive yet interconnected aspects of its application. The left part of the figure meticulously portrays the relationship between the number of principal components utilized and the corresponding explained variance. This graphical representation provides a comprehensive view of how the inclusion of additional components influences the cumulative variance captured within the dataset. Simultaneously, the right segment of Figure 2 meticulously illustrates the impact of varying component numbers on classification accuracy. This graphical depiction offers a nuanced understanding of the relationship between the dimensionality reduction achieved through PCA and its consequent effect on the accuracy of classification algorithms. The juxtaposition of these visualizations not only delineates the trade-offs inherent in dimensionality reduction but also underscores the critical interplay between PCA-driven variance retention and its ultimate influence on the accuracy of classification models employed within the scope of this study.

Figure 3 unveils a detailed dendrogram meticulously constructed to represent the intricate relationships and hierarchical clustering of feature correlations within the dataset. This dendrogram encapsulates a comprehensive visual depiction of the interplay and clustering patterns among the various features under consideration. Through hierarchical clustering, the dendrogram visually dissects the similarities and dissimilarities between features, elucidating clusters of correlated variables and their hierarchical organization. This graphical representation offers invaluable insights into

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

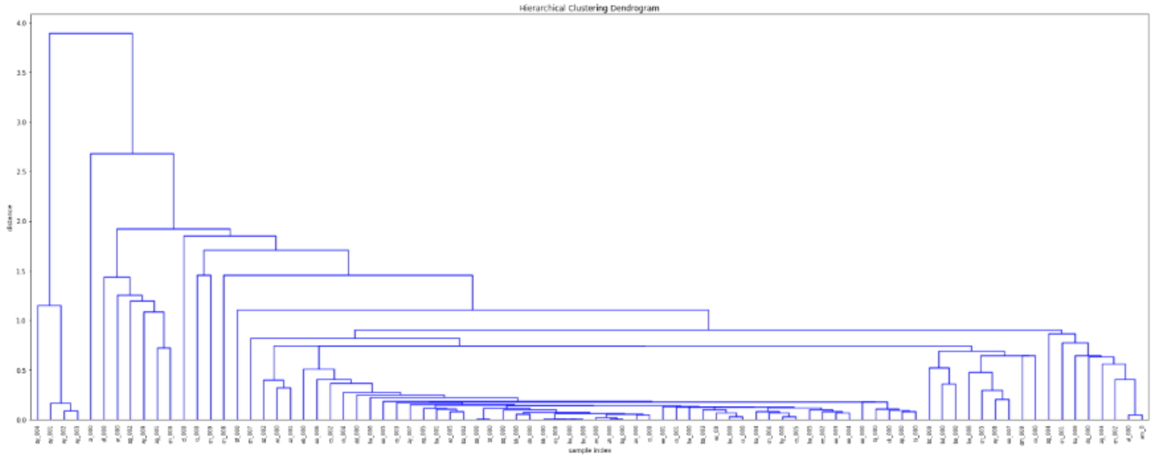

Figure 3: Dendrogram Illustrating Hierarchical Feature Correlations.

the underlying structure of feature correlations, unveiling potential clusters or groups of features exhibiting similar behavioral patterns. The dendrogram showcased in Figure 3 serves as a pivotal tool for comprehending feature relationships and structuring subsequent analyses within the dataset, enabling a more nuanced understanding of feature interdependencies crucial for subsequent modeling and interpretation.

In Table 1, a comprehensive evaluation and comparative analysis of various ML classifiers' performance metrics are meticulously presented. This comparative assessment encapsulates a detailed scrutiny of multiple classifiers, dissecting their performance across diverse evaluation metrics such as accuracy, precision, recall, and F1-score. The table provides a succinct yet comprehensive overview of each classifier's efficacy in handling the dataset, offering insights into their strengths and weaknesses concerning predictive performance. Through this comparative analysis, the table serves as a valuable reference point for understanding the relative merits of each classifier, aiding in the selection of the most suitable model based on specific performance criteria. The comprehensive presentation of multiple classifiers' performance metrics within Table 1 facilitates informed decision-making and highlights the nuances of each model's predictive capabilities within the context of the study's dataset.

Table 1: Comparative Performance Metrics of Various ML Classifiers.

model	Accuracy	f1_score	Precision	Recall	RoC_AUC
LogisticRegression	0.8181	0.8403	0.7487	0.9575	0.8181
GaussianNB	0.9236	0.9198	0.9677	0.8764	0.9236
BernoulliNB()	0.8567	0.8561	0.8599	0.8524	0.8567
KNeighborsClassifier	0.9866	0.9868	0.9749	0.9990	0.9866
DecisionTreeClassifier	0.9906	0.9906	0.9875	0.9937	0.9906
DecisionTreeClassifier	0.9951	0.9951	0.9928	0.9974	0.9951
VotingClassifier	0.9889	0.9890	0.9817	0.9964	0.9889

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

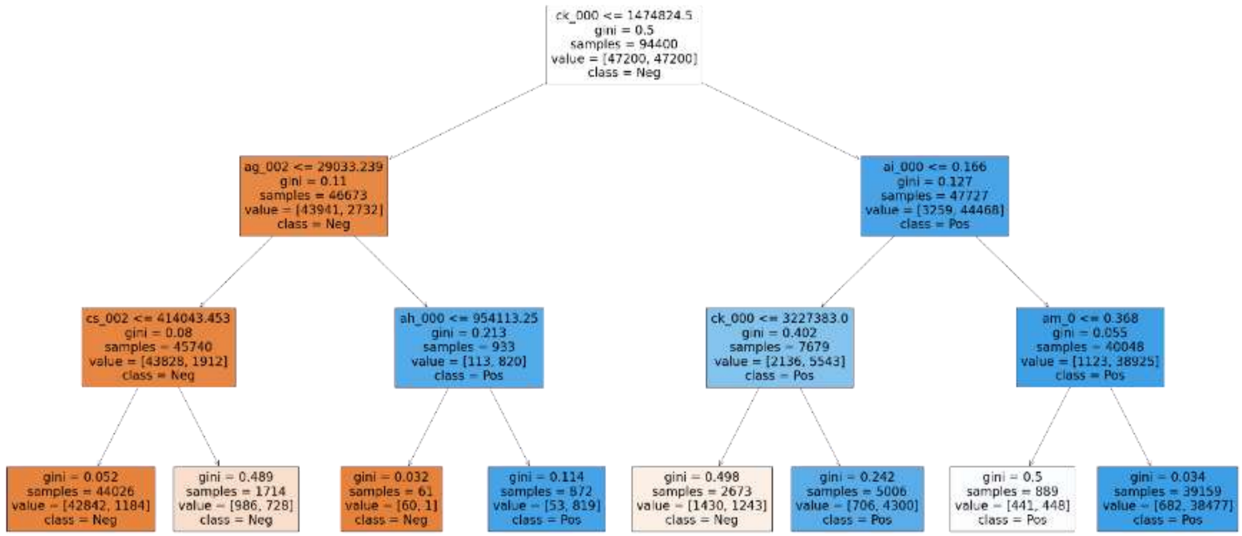

Figure 4: Visualization of Decision Tree Branching.

Figure 4 offers a comprehensive visualization portraying the branching structure of a Decision Tree (DT) model. This intricate visual representation meticulously maps out the hierarchical decision-making process employed by the DT algorithm. Each node within the tree signifies a decision point based on specific features, with branches extending to subsequent nodes representing further attribute splits or decision pathways. Through this graphical depiction, the branching patterns of the DT model are elucidated, showcasing the sequence of feature conditions and their hierarchical importance in determining the final predictions or classifications made by the model. This visualization in Figure 4 serves as a valuable tool for comprehending the decision logic employed by the DT algorithm, providing an intuitive and transparent illustration of its rule-based approach to predictive modeling within the context of the study's dataset.

5. Conclusion

This study marks a pivotal step forward in the realm of real-time monitoring for air pressure systems in heavy-duty vehicles, specifically addressing the challenges encountered within Scania trucks. Through the development and empirical validation of a dedicated Data Fusion Framework, this research showcases a promising approach to enhancing operational safety and reliability. The comparative performance analysis of multiple machine learning classifiers emphasized the superiority of our framework in handling the complexities of air pressure system monitoring. By amalgamating these methodologies, this study not only contributes a robust framework for proactive anomaly detection but also underscores the potential for fostering safer and more efficient heavy-duty vehicle operations. As the automotive industry progresses towards predictive maintenance and real-time monitoring paradigms, this research stands as a foundation, paving the way for continued advancements in enhancing the reliability and safety standards of heavy-duty vehicle systems.

References

[1] Jan, Zohaib, Farhad Ahamed, Wolfgang Mayer, Niki Patel, Georg Grossmann, Markus Stumptner, and Ana Kuusk. 2022. "Artificial Intelligence for Industry 4.0: Systematic Review of Applications, Challenges, and Opportunities." Expert Systems with Applications, 119456.

[2] Katreddi, Sasanka, Sujan Kasani, and Arvind Thiruvengadam. 2022. "A Review of Applications of Artificial Intelligence in Heavy Duty Trucks." Energies 15 (20): 7457.

[3] Wang, Qihang, Tianci Gao, Haichuan Tang, Yifeng Wang, Zhengxing Chen, Jianhui Wang, Ping Wang, and Qing He. 2021. "A Feature Engineering Framework for Online Fault Diagnosis of Freight Train Air Brakes." Measurement 182: 109672.

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023

[4] Mårtensson, Jonas, Assad Alam, Sagar Behere, Muhammad Altamash Ahmed Khan, Joakim Kjellberg, Kuo-Yun Liang, Henrik Pettersson, and Dennis Sundman. 2012. "The Development of a Cooperative Heavy-Duty Vehicle for the GCDC 2011: Team Scoop." IEEE Transactions on Intelligent Transportation Systems 13 (3): 1033–49.
[5] Refors, Michael. 2016. "Information Filter Based Sensor Fusion for Estimation of Vehicle Velocity."
[6] Nilsson, Sanna. 2012. "Sensor Fusion for Heavy Duty Vehicle Platooning."
[7] Bellavista, Paolo, Roberto Della Penna, Luca Foschini, and Domenico Scotece. 2020. "Machine Learning for Predictive Diagnostics at the Edge: An IIoT Practical Example." In ICC 2020-2020 IEEE International Conference on Communications (ICC), 1–7.
[8] Kafunah, Jefkine, Muhammad Intizar Ali, and John G Breslin. 2023. "Uncertainty-Aware Ensemble Combination Method for Quality Monitoring Fault Diagnosis in Safety-Related Products." IEEE Transactions on Industrial Informatics.
[9] Stoichkov, Radoslav. 2013. "Android Smartphone Application for Driving Style Recognition." Department of Electrical Engineering and Information Technology Institute for Media Technology 20.
[10] Alam, Assad. 2014. "Fuel-Efficient Heavy-Duty Vehicle Platooning." KTH Royal Institute of Technology.
[11] Hou, Zhefan, C K M Lee, Yaqiong Lv, and K L Keung. 2023. "Fault Detection and Diagnosis of Air Brake System: A Systematic Review." Journal of Manufacturing Systems 71: 34–58.
[12] Alam, Assad. 2011. "Fuel-Efficient Distributed Control for Heavy Duty Vehicle Platooning." KTH Royal Institute of Technology.
[13] Huang, Zijie, Yulei Wu, Niccolò Tempini, Hui Lin, and Hao Yin. 2022. "An Energy-Efficient and Trustworthy Unsupervised Anomaly Detection Framework (EATU) for IIoT." ACM Transactions on Sensor Networks 18 (4): 1–18.
[14] Wolf, Patrick, Thorsten Ropertz, Karsten Berns, Martin Thul, Peter Wetzel, and Achim Vogt. 2018. "Behavior-Based Control for Safe and Robust Navigation of an Unimog in off-Road Environments." In Commercial Vehicle Technology 2018: Proceedings of the 5th Commercial Vehicle Technology Symposium-CVT 2018, 63–76.
[15] Theissler, Andreas, Judith Pérez-Velázquez, Marcel Kettelgerdes, and Gordon Elger. 2021. "Predictive Maintenance Enabled by Machine Learning: Use Cases and Challenges in the Automotive Industry." Reliability Engineering \& System Safety 215: 107864.
[16] Alam, Assad, Bart Besselink, Valerio Turri, Jonas Mårtensson, and Karl H Johansson. 2015. "Heavy-Duty Vehicle Platooning for Sustainable Freight Transportation: A Cooperative Method to Enhance Safety and Efficiency." IEEE Control Systems Magazine 35 (6): 34–56.
[17] Dias, Savidu. 2023. "Integration of MLOps with IoT Edge." S. Dias.
[18] Dias, Savidu, and Ella Peltonen. 2023. "LinkEdge: Open-Sourced MLOps Integration with IoT Edge." In ESAAM 2023: 3rd Eclipse Security, AI, Architecture and Modelling Conference on Cloud to Edge Continuum.
[19] Nacke, Gustav, and Samuel Hirschfeld. 2021. "Evaluating Implementation Areas of Real-Time Location System (RTLS) in the Production at Scania CV AB Oskarshamn."
[20] Scapinakis, Dimitris A, and William L Garrison. 1991. "Communications and Positioning Systems in the Motor Carrier Industry."
[21] Rettore, Paulo Henrique Lopes. 2019. "Fusion on Vehicular Data Space: An Approach to Smart Mobility."
[22] Lima, Pedro F. 2018. "Optimization-Based Motion Planning and Model Predictive Control for Autonomous Driving: With Experimental Evaluation on a Heavy-Duty Construction Truck." KTH Royal Institute of Technology.
[23] Lima, Pedro. 2016. "Predictive Control for Autonomous Driving: With Experimental Evaluation on a Heavy-Duty Construction Truck." KTH Royal Institute of Technology.

Doi: https://doi.org/10.54216/NIF.020203
Received: June 19, 2023 Accepted: November 03, 2023