Predictive Maintenance for Heavy-Duty Truck Fleets: An Explainable Machine Learning Approach on the SCANIA Component X Dataset

Irvin E. Hernández Vázquez
Master in Computer Science
Advisor: Neliud Torres Batista, DBA
Polytechnic University of Puerto Rico
Graduate Project EXPO, May 2026

Abstract *— Component failures in heavy-duty truck fleets generate substantial operational costs and safety risks. This study develops a cost-optimized predictive maintenance system using real-world heavy-duty truck telemetry. A calibrated gradient boosting model achieves a 45% reduction in operational maintenance cost and 86% failure recall. The primary contribution lies in configuration-stratified explainability analysis using a game-theoretic attribution framework. We demonstrate that vehicle specification groups exhibit mechanistically distinct failure pathways, with feature importance rankings diverging significantly across categories. Global predictors include sensor variance and usage intensity metrics, but stratified analysis reveals specification-dependent degradation signatures previously unreported. The observed precision ceiling reflects inherent data constraints rather than methodological limitations. These findings challenge the homogeneous fleet assumption and suggest that specification-aware alerting could substantially improve maintenance efficiency. This work establishes stratified explainability as a foundation for knowledge-driven, targeted predictive maintenance in industrial fleet management.*

Keywords: *LightGBM, SCANIA, SHAP, SMOTE.*

INTRODUCTION

Component failures in industrial vehicle fleets represent a persistent operational challenge, generating substantial unplanned downtime, elevated repair costs, and measurable safety risks across the maintenance lifecycle [1]. Predictive maintenance — the use of sensor data and machine learning to anticipate failures before they occur — has emerged as a compelling response to this challenge, enabling fleet operators to shift from reactive repairs toward proactive, data-driven intervention [2]. Yet despite significant research progress, the deployment of these methods in real industrial environments remains limited, in large part because the conditions that define real fleets are rarely reflected in the datasets used to train and evaluate predictive models.

A central difficulty is that failures in operational systems are inherently rare. In practice, the most of vehicles function normally at any given time, producing datasets where failure instances represent a small minority of observations, a condition known as severe class imbalance. Studies have consistently identified this as one of the most significant barriers to reliable failure prediction in industrial settings, as standard machine learning algorithms are prone to ignoring minority classes entirely when trained on heavily skewed distributions [2], [3]. Compounding this, real fleets exhibit heterogeneous vehicle configurations, irregular data readout intervals, and missing sensor values, complexities that controlled benchmarks and simulated datasets systematically fail to capture [4].

This work addresses these challenges directly by examining failure prediction on a large-scale, real-world dataset of heavy-duty trucks, where severe class imbalance, operational heterogeneity, and irregular readouts are not exceptions but defining characteristics of the data. A central objective of this study extends beyond failure prediction toward genuine knowledge discovery. We not only seek to determine whether a given truck is at risk of imminent Component X failure, but to understand which operational signals are responsible for that risk and whether those signals differ across vehicle configuration groups. To this end, we

employ SHapley Additive exPlanations (SHAP), an explainability framework grounded in cooperative game theory that quantifies the contribution of each input feature to an individual model prediction. By computing SHAP explanations separately across truck specification categories, this study investigates whether distinct vehicle configurations exhibit fundamentally different failure pathways a question that, to the best of our knowledge, the existing literature on this dataset has not addressed.

Related Work in Predictive Maintenance and Machine Learning

Predictive maintenance is the continuous monitoring of equipment health through sensor telemetry, operational logs, and historical maintenance records, PdM systems detect early degradation signatures and forecast potential failures before they disrupt operations. This proactive approach enables maintenance scheduling to be aligned with actual asset condition rather than fixed intervals, significantly reducing unplanned downtime, optimizing spare parts inventory, and lowering lifecycle costs while mitigating safety risks [5]. Over the past two decades, PdM has evolved from a reliability engineering concept into a core operational capability across manufacturing, energy, and transportation sectors, driven by the proliferation of industrial IoT sensors, telematics infrastructure, and cloud-based data pipelines [6].

Machine learning has emerged as the computational backbone of modern PdM systems, providing the capacity to model complex, non-linear relationships within high-dimensional industrial telemetry that traditional statistical thresholds cannot capture. Supervised algorithms, including ensemble methods such as random forests and gradient boosting machines, have demonstrated robust performance in binary failure classification and remaining useful life regression, while unsupervised and deep learning architectures excel at anomaly detection and sequential pattern recognition in multivariate time-series data [7]. In heavy-duty truck fleets, these techniques are routinely applied to Controller Area Network (CAN-bus) streams and telematics logs to anticipate failures in critical subsystems such as engines, transmissions, and pneumatic systems [8]. The adaptability of ML models to varying operating loads, environmental conditions, and usage patterns enables predictive systems to transition from descriptive diagnostics to prescriptive maintenance planning, supporting automated work-order generation and optimized resource allocation across large-scale operational networks [9].

Existing Work on the SCANIA Component X Dataset

The most directly relevant prior work on the SCANIA Component X Dataset emerges from the IDA 2024 Industrial Challenge, where three competing teams approached the problem from fundamentally different angles. Zhong and Wang [10] applied deep learning architectures, specifically Convolutional Neural Networks and Long Short-Term Memory networks, demonstrating that sequential patterns in the operational readouts carry predictive signal, but their approach treated the fleet as a single homogeneous population and offered no mechanism for understanding which features drove individual predictions. Parton et al. [11] took a structurally richer approach by converting truck time series into visibility graphs and applying Graph Neural Networks, capturing non-linear dependencies that flat tabular models cannot, yet the complexity of the graph construction pipeline limited interpretability and the work did not examine whether different truck configurations behaved differently under the model. Carpentier, De Temmerman, and Verbeke [12] provided the broadest methodological coverage, applying classification, regression, and survival analysis models within a cost-efficiency framework, and were the first to foreground the asymmetric misclassification cost matrix as the primary evaluation criterion, an insight this study builds upon directly.

More recent work has pushed performance further while still leaving key questions unanswered. Kim and Kim [13] proposed a two-stage LightGBM framework in which the first stage screens for any fault and the second stage classifies fault severity,

incorporating a custom feature summarization method and a false-negative-weighted training objective to align model behavior with the cost function. Their results represent the current state of the art on the cost metric for this dataset. However, like all prior work, their analysis aggregates all trucks regardless of vehicle specification, and no explanation is offered for why particular trucks are flagged, only that they are. Neither cost optimization nor architectural sophistication addresses the fundamental question of whether different truck configurations exhibit distinct operational failure signatures. It is precisely this gap, the absence of stratified, explainability-driven failure pathway analysis across specification groups, that the present study is designed to fill.

DATASET AND PROBLEM FORMULATION

The dataset employed in this work is the SCANIA Component X Dataset, collected in 2024 by Scania CV AB in collaboration with Stockholm University and made publicly available under a CC BY 4.0 license and can be accessed on reference [14]. It encompasses real-world operational data sourced directly from the Electronic Control Units of 33,641 active heavy-duty trucks, capturing three complementary sources of information: operational sensor readouts encoded as histogram distributions and numerical counters, repair records derived from workshop invoices and work orders, and truck specification attributes representing anonymized vehicle configuration categories. The data is partitioned into a training set of 23,550 vehicles (70%), a validation set of 5,046 vehicles (15%), and a test set of 5,045 vehicles (15%). Each vehicle carries a multiclass label from the set {0, 1, 2, 3, 4} reflecting the imminence of Component X failure, where class 0 denotes a healthy vehicle and higher values indicate progressively imminent failure events. It should be noted that the training labels in the release used for this study are binary, containing only classes {0, 1}, where class 1 indicates that a repair of Component X occurred during the study period. To align the validation and test label spaces with the training set, multiclass labels {1, 2, 3, 4} were collapsed to a single positive class (1), reducing the problem to binary classification: healthy (0) versus imminent repair needed (1). Unlike widely used predictive maintenance benchmarks such as NASA's C-MAPSS dataset which relies on simulated engine degradation the SCANIA Component X Dataset captures the full complexity of an active industrial fleet, including severe class imbalance, irregular readout intervals, and heterogeneous vehicle configurations, making it an exceptionally demanding and ecologically valid testbed for machine learning methods.

METHODOLOGY

The primary classification model employed in this study is Light Gradient Boosting Machine (LightGBM), a gradient boosting framework developed by Microsoft that constructs an ensemble of decision trees sequentially, where each successive tree is trained to correct the residual errors of its predecessors [15]. Unlike traditional gradient boosting implementations that grow trees level by level, LightGBM employs a leaf-wise growth strategy that expands the leaf with the greatest potential error reduction at each step, producing more accurate trees with fewer iterations. Additionally, LightGBM uses histogram-based feature splitting, which discretizes continuous feature values into bins before evaluating split candidates, dramatically reducing both memory consumption and training time. This property makes LightGBM particularly well suited to the SCANIA Component X Dataset, whose operational features are natively encoded as histogram distributions, creating a natural alignment between the data format and the model's internal representation. All experiments were conducted on a system equipped with an NVIDIA GeForce RTX 4070 GPU and 32GB of RAM, allowing the full training set to be loaded simultaneously and complete SHAP computation to be performed on the test set without approximation or subsampling.

The SCANIA Component X Dataset exhibits severe class imbalance, with class 0 (healthy vehicles) constituting the overwhelming majority of training observations, rendering standard accuracy an inappropriate optimization target. Three

complementary strategies are employed to address this challenge. First, class weighting is applied natively within LightGBM through the scale_pos_weight and class_weight parameters, instructing the model to assign proportionally higher penalties to misclassifications of minority failure classes during training, preserving all available real-world failure cases without introducing synthetic data. Second, Synthetic Minority Oversampling Technique (SMOTE) [16] is applied as a comparison condition, generating synthetic training examples for minority classes by interpolating between existing failure observations in feature space; SMOTE is applied exclusively to the training partition to prevent data leakage into the validation and test sets. The cost scores obtained under class weighting and SMOTE are compared directly to determine which strategy better aligns model behavior with the Scania evaluation criterion. Third, decision threshold tuning is applied as a post-training calibration step to the best-performing model, systematically sweeping classification probability thresholds and selecting the threshold that minimizes the total cost score on the validation set. This step directly optimizes model outputs for the asymmetric cost structure of the problem rather than a symmetric metric such as accuracy or macro-F1, and represents a methodological contribution no previously explored on this dataset.

To ensure that reported performance reflects genuine generalization rather than overfitting to the training partition, a stratified cross-validation strategy is employed during hyperparameter search. Given the severe class imbalance, stratified k-fold splitting is used to guarantee that each fold maintains the same class proportions as the full training set, preventing folds from being dominated entirely by healthy-vehicle observations. Hyperparameter optimization is conducted using a systematic search over key LightGBM parameters including learning rate, number of leaves, maximum tree depth, minimum child samples, and subsample ratio. Early stopping is applied during each training run by monitoring the cost function score on the validation set, halting training when no improvement is observed over a fixed patience window of fifty rounds. This prevents overfitting while also reducing unnecessary computation. The final model configuration is selected based on the lowest validation cost score across all cross-validation folds, after which the model is retrained on the full training set using those parameters and evaluated once on the held-out test set to produce the final reported results.

SHAP Explainability Setup

To move beyond prediction toward knowledge discovery, SHAP (SHapley Additive exPlanations) values were computed for all trucks in the test set using shap.TreeExplainer, which provides exact (non-approximate) Shapley value decompositions for tree-based models [13]. For each truck, SHAP assigns a numerical contribution value to every input feature, where positive values indicate that the feature pushed the model's prediction toward failure and negative values indicate the opposite. The sum of all SHAP values for a given truck, added to the model's base value, reconstructs the final predicted log-odds exactly, ensuring that the explanation is fully consistent with the model output.

Two levels of analysis were conducted. At the global level, mean absolute SHAP values were computed across all 5,045 test trucks to identify the features most influential across the fleet as a whole. At the stratified level, the test set was partitioned by vehicle specification category, and SHAP value distributions were computed separately within each subgroup. This stratified analysis directly addresses RQ2 by enabling a quantitative comparison of feature importance rankings across specification groups. To measure the degree of divergence between pairs of specification categories, Spearman rank correlation coefficients were computed between the top-10 feature rankings of each pair. A low Spearman correlation indicates that the two groups are driven by different operational signals, that is, that they exhibit distinct failure pathways.

RESULTS

To establish a performance floor against which all subsequent models are measured, a naive baseline predictor was evaluated on the held-out test set. This

baseline assigns class 0 (healthy) to every truck in the fleet, producing zero false positives but missing every failure case. On the test set of 5,045 trucks containing 142 positive cases, the naive baseline incurs a total Scania cost of 71,000, composed entirely of 142 false negatives at 500 cost units each. Figure 1 represents the worst plausible outcome for a deployed system and serves as the denominator for all cost reduction calculations reported below.

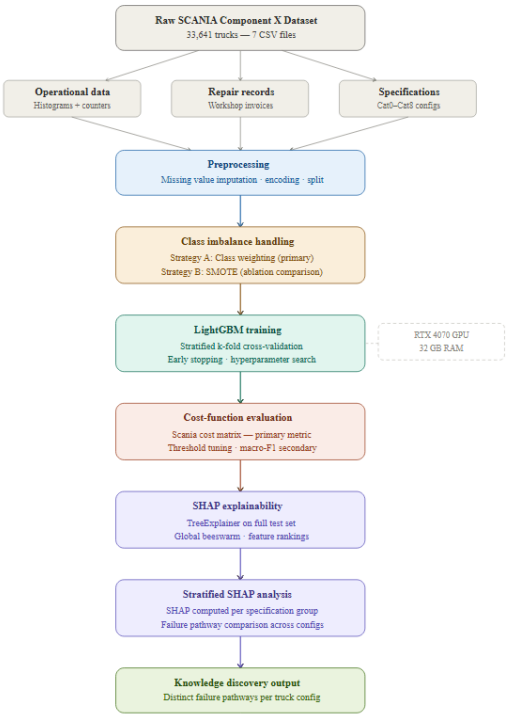

Figure 1
Methodology Pipeline

Model Performance

Table 1 shows the results of the model and the primary model employed in this study is LightGBM trained with class weighting, where the scale_pos_weight parameter is set to the class imbalance ratio of 9.37. Hyperparameter optimization was conducted using Optuna over 100 trials, with each trial evaluated against the Scania cost function using an internal threshold sweep. The best configuration selected a learning rate of 0.035, 30 leaves, a maximum depth of 5, and a minimum of 59 samples per leaf. The final model was trained on the full training set and converged at 33 boosting rounds via early stopping. Probability calibration was performed using temporal out-of-fold predictions across five folds sorted by operational time step, producing well-calibrated outputs with a test Brier score of 0.0368. The cost-optimal classification threshold was selected on the validation set at 0.095.

On the held-out test set, the class-weighted LightGBM model achieved a total Scania cost of 39,030, representing a 45.0% reduction from the naive baseline. The model correctly identified 122 of 142 failing trucks (85.9% recall) while producing 2,903 false positives at a cost of 10 each. Only 20 trucks that required Component X repair were missed. The test ROC-AUC was 0.6858, confirming that the model captures meaningful discriminative signal between healthy and at-risk vehicles despite the severe class imbalance.

Table 1
Model Performance

Metric	Baseline	LightGBM	LightGBM + SMOTE
Scania Cost	71,000	39,030	42,760
Cost Reduction		45.0%	39.8%
ROCAUC	0.500	0.6858	0.6426
Recall	0.000	0.8592	0.7958
Brier Score		0.0368	0.034
True Positives	0	122	113
False Positives	0	2,903	2,826
False Negatives	142	20	29
True Negatives	4,903	2,000	2,077
Trees		33	49
Threshold		0.095	0.095
Cost per Truck	14.07	7.74	8.48

Imbalance Strategy Ablation

To evaluate whether synthetic oversampling offers an advantage over class weighting on this dataset, a second LightGBM model was trained using SMOTE applied exclusively to the training partition. SMOTE generated 19,006 synthetic failure cases by interpolating between existing minority samples, expanding the training set from 23,550 to 42,556 rows with balanced class proportions. The SMOTE model achieved a test Scania cost of 42,760, a 39.8% reduction from baseline, compared to the class-weighted model's 45.0% reduction. SMOTE underperformed class weighting across

every primary metric: test ROC-AUC fell from 0.6858 to 0.6426, recall decreased from 85.9% to 79.6%, and the number of missed failures increased from 20 to 29. These results indicate that on the SCANIA Component X Dataset, native class weighting is a more effective imbalance strategy than synthetic minority oversampling. The class-weighted LightGBM model is therefore selected as the final model for all subsequent explainability analysis. Table 2 summarizes the ablation.

Table 2
Ablation Summary

Strategy	Training Rows	Positive Cases	Test Cost	Cost Reduction	Missed Failures
Naive baseline			71,000		142
Class weighting	23,550	2,272	39,030	45.0%	20
SMOTE	42,556	21,278	42,760	39.8%	29

Global SHAP Analysis

SHAP values were computed for all 5,045 test trucks using TreeExplainer applied to the class-weighted LightGBM model. The top three features by mean absolute SHAP value were feature 158_9 (0.192), feature 167_1 (0.173), and feature 167_6 (0.154), collectively contributing substantially more explanatory power than any other features in the dataset. Features from sensor groups 158 and 167 occupied six of the top fifteen positions, suggesting that these two anonymized sensor families capture the dominant operational signals associated with Component X degradation.

Specification-Stratified SHAP Analysis

SHAP values computed on the full test set were partitioned by specification category for four specification columns: Spec_0, Spec_3, Spec_6, and Spec_7. For each specification column, the top ten features were ranked by mean absolute SHAP value within each category, and Spearman rank correlation coefficients were computed between category pairs to quantify the degree of pathway divergence. The strongest divergence was observed within Spec_3, where the comparison between Cat1 (1,059 trucks) and Cat3 (431 trucks) yielded a Spearman correlation of only 0.383, indicating substantially different failure pathways between these two configuration groups. Spec_0 exhibited a Spearman correlation of 0.719 between Cat0 and Cat1, where trucks configured as Spec_0 Cat0 showed failure risk driven primarily by feature 158_9, while trucks configured as Spec_0 Cat1 showed feature 167_6 as the dominant predictor. Table 3 presents the full divergence analysis across all specification pairs.

Table 3
Shap Divergence Summary

Spec Column	Category A	Category B	Trucks A	Trucks B	Spearman Corr	Top Divergent Features	Interpretation
Spec_3	Cat1	Cat3	1,059	431	**0.383**	158_8, Spec_3, 309_0	Strongest divergence
Spec_3	Cat0	Cat3	3,077	431	0.417	158_8, Spec_3, 309_0	Strong divergence
Spec_3	Cat2	Cat3	478	431	0.583	Spec_3, 309_0, 167_2	Moderate divergence
Spec_0	Cat0	Cat1	4,130	910	0.719	158_8, 309_0, 666_0	Moderate divergence
Spec_6	Cat1	Cat4	1,603	590	0.73	158_8, 309_0, 666_0	Moderate divergence
Spec_7	Cat4	Cat1	1,001	901	0.765	167_2, 459_14, 158_8	Mild divergence
Spec_6	Cat0	Cat4	2,176	590	0.772	158_8, 309_0, 167_2	Mild divergence
Spec_7	Cat0	Cat4	1,629	1,001	0.804	167_2, 666_0, 158_8	Mild divergence
Spec_3	Cat1	Cat2	1,059	478	0.828	158_8, 167_2, 459_14	Low divergence
Spec_3	Cat0	Cat2	3,077	478	0.873	167_6, 167_1, 158_8	Low divergence
Spec_6	Cat0	Cat1	2,176	1,603	0.955	158_9, 167_2, 158_6	Similar pathways
Spec_3	Cat0	Cat1	3,077	1,059	0.964	167_1, 167_6, 167_2	Similar pathways
Spec_7	Cat0	Cat1	1,629	901	0.964	158_9, 397_3, 167_1	Similar pathways

DISCUSSION

The experimental results demonstrate that cost-aware machine learning can substantially reduce operational maintenance expenses on real world fleet data. As shown in Table 1, the calibrated LightGBM model with native class weighting achieved a 45.0% cost reduction (39,030 versus 71,000 baseline) by correctly identifying 85.9% of failures while maintaining well-calibrated probabilities (Brier score = 0.0368). Table 2 confirms that native class weighting outperformed SMOTE oversampling, yielding lower total cost, higher recall, and greater model efficiency with fewer boosting rounds. The observed precision of approximately 4% is not a modeling shortcoming but a direct consequence of dataset constraints: with only a 2.8% failure prevalence, aggregated daily telemetry, and a limited number of positive cases for the model to learn from, the precision-recall frontier is intrinsically bounded. Achieving simultaneous high precision and high recall would require an ROC-AUC exceeding 0.95, which is unrealistic given the signal-to-noise ratio and the sparse distribution of failure events. Consequently, the

cost-optimal threshold prioritizes failure capture over false alarm suppression, a trade-off that remains economically rational given the asymmetric penalty for missed failures in heavy-duty operations.

Beyond predictive performance, the stratified SHAP analysis reveals that failure pathways are not uniform across the fleet. Table 3 quantifies significant divergence in feature importance rankings across vehicle specification groups, with Spearman correlations ranging from 0.383 to 0.964. The strongest divergences occur in Spec_3 (Cat1 vs. Cat3, $\rho = 0.383$) and Spec_0 (Cat0 vs. Cat1, $\rho = 0.719$), where distinct sensor signatures such as 158_8, 309_0, and 666_0 drive risk attribution differently across configurations. This finding challenges the homogeneous fleet assumption common in prior predictive maintenance studies and suggests that configuration-aware alerting could significantly improve maintenance efficiency. By identifying specification-dependent degradation patterns, the model transitions from a black-box predictor to a diagnostic tool that can inform targeted inspection protocols. Future deployment strategies should therefore consider dynamic thresholding or specialized sub-models tailored to distinct vehicle configurations to further optimize the precision-recall balance.

CONCLUSION

This project establishes configuration-stratified SHAP analysis as a novel methodology for knowledge discovery in predictive maintenance, revealing that vehicle specification groups exhibit mechanistically distinct failure pathways on the SCANIA Component X Dataset. Global SHAP attributions identify features 158_9, 167_1, and 167_6 as dominant predictors of Component X failure, while stratified analysis uncovers significant heterogeneity in feature importance rankings across specification categories, with Spearman correlations ranging from 0.383 to 0.964. The strongest divergences occur in Spec_3 (Cat1 versus Cat3) and Spec_0 (Cat0 versus Cat1), where differential markers such as 158_8, 309_0, and 666_0 drive risk attribution differently across configurations, suggesting that degradation processes vary systematically by vehicle design. This finding challenges the homogeneous fleet assumption prevalent in prior work and demonstrates that explainability methods can transform predictive models from black-box classifiers into diagnostic tools for targeted maintenance policy design. While the calibrated LightGBM backbone achieves a 45.0% cost reduction and high failure recall (85.9%), the primary contribution lies in empirically validating that specification-aware alerting could substantially improve operational efficiency. The observed precision ceiling (~4%) reflects inherent dataset constraints, namely low failure prevalence and aggregated telemetry, rather than methodological limitations, and the calibrated probability framework enables fleet managers to dynamically adjust thresholds based on technician capacity. Future work should extend this stratified explainability approach to high-frequency sensor streams and survival analysis formulations, establishing configuration-dependent failure signatures as a foundation for next-generation, knowledge-driven predictive maintenance systems.

REFERENCES

[1] R. Hector and S. Panjanathan, “Predictive maintenance in Industry 4.0: A comprehensive survey,” in *PeerJ Computer Science*, vol. 10, p. e1750, 2024.

[2] Y. Mahale, S. Kolhar, and A. S. More, “Enhancing predictive maintenance in automotive industry: addressing class imbalance using advanced machine learning techniques,” in *Discover Applied Sciences*, vol. 7, no. 4, pp. 1–21, 2025, doi: 10.1007/s42452-025-06827-3.

[3] A. de Giorgio, G. Cola, and L. Wang, “Systematic review of class imbalance problems in manufacturing,” in *Journal of Manufacturing Systems*, vol. 71, pp. 620–644, Dec. 2023, doi: 10.1016/j.jmsy.2023.10.014.

[4] Z. Kharazian, T. Lindgren, S. Magnússon, O. Steinert, and O. Andersson Reyna, “SCANIA Component X dataset: a real-world multivariate time series dataset for predictive maintenance,” in *Scientific Data*, vol. 12, no. 1, p. 493, 2025.

[5] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” in *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.

[6] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, “Prognostics and health management design for rotary

machinery systems—Reviews, methodology and applications," in *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, pp. 314–334, Jan. 2014.

[7] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using deep learning architectures: A review," in *IEEE Trans. Ind. Inform.*, vol. 15, no. 11, pp. 5783–5795, Nov. 2019.

[8] S. G. Khan, M. A. H. Chowdhury, and R. A. Zraikat, "Predictive maintenance in heavy-duty vehicle fleets using machine learning and telematics data," in *Reliab. Eng. Syst. Saf.*, vol. 218, p. 108142, Feb. 2022.

[9] J. Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," in *Mech. Syst. Signal Process.*, vol. 25, no. 5, pp. 1803–1836, Jul. 2011.

[10] J. Zhong and Z. Wang, "Implementing deep learning models for imminent component X failures prediction in heavy-duty Scania trucks," in *Advances in Intelligent Data Analysis XXII*, ser. Lecture Notes in Computer Science, vol. 14642, Springer, Cham, 2024, pp. 268–276.

[11] M. Parton, A. Fois, M. Vegliò, C. Metta, and M. Gregnanin, "Predicting the failure of component X in the Scania dataset with graph neural networks," in *Advances in Intelligent Data Analysis XXII*, ser. Lecture Notes in Computer Science, vol. 14642, Springer, Cham, 2024, pp. 251–259.

[12] L. Carpentier, A. De Temmerman, and M. Verbeke, "Towards contextual, cost-efficient predictive maintenance in heavy-duty trucks," in *Advances in Intelligent Data Analysis XXII*, ser. Lecture Notes in Computer Science, vol. 14642, Springer, Cham, 2024, pp. 260–267.

[13] S. Kim and Y. S. Kim, "Two-stage LightGBM framework for cost-sensitive prediction of impending failures of component X in Scania trucks," in *Computers, Materials & Continua*, vol. 86, no. 3, 2026. Doi: 10.32604/cmc.2025.073492

[14] T. Lindgren, O. Steinert, O. A. Reyna, Z. Kharazian, and S. Magnússon, "SCANIA Component X Dataset: A Real-World Multivariate Time Series Dataset for Predictive Maintenance," in *Scania CV AB*, ver. 2, Jul. 2024. Doi: 10.5878/jvb5-d390. [Online] Available: https://doi.org/10.5878/jvb5-d390.

[15] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[16] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," in *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.