# Explainable Dynamic Graph Neural Networks for Predictive Maintenance in Vehicle Chassis Systems

**You Wu** r130026160@mail.uic.edu.cn
*Guangdong Provincial/Zhuhai Key Laboratory of IRADS*
*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University*

**Junwei Su** r130026129@mail.uic.edu.cn
*Guangdong Provincial/Zhuhai Key Laboratory of IRADS*
*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University*

**Sirui Zhang** r130026199@mail.uic.edu.cn
*Guangdong Provincial/Zhuhai Key Laboratory of IRADS*
*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University*

**Zhijian Li**[*] zhijianli@uic.edu.cn
*Guangdong Provincial/Zhuhai Key Laboratory of IRADS*
*Department of Statistics and Data Science, Beijing Normal-Hong Kong Baptist University*

**Editors:** Hung-yi Lee and Tongliang Liu

## Abstract

Predictive maintenance is essential for commercial vehicle fleets to reduce unexpected downtime and emergency repair costs. While standardized fault codes (SPN/FMI, representing *Suspect Parameter Numbers* and *Failure Mode Indicators*) assist in diagnosis, their temporal and spatial inconsistency limits the effectiveness of conventional time-series models in identifying high-cost failures. We propose a *Hybrid Node-level Relationship-based Graph Convolutional Network with Random Forest* (**NRP-GCN-RF**), which encodes fault interactions as graphs to capture non-temporal dependencies. Based on a real-world dataset provided by a large commercial vehicle manufacturer, our study follows a dual-task design. (1) We construct a predictive model using graph neural networks (GCN) and random forests (RF) to forecast emergency repair costs and fault categories based on chassis-level fault sequences. (2) In parallel, we apply the Apriori algorithm to mine frequent co-occurring SPN–FMI pairs, revealing interpretable fault patterns and subsystem-level dependencies. This interpretable analysis complements the graph-based model by supporting feature design and failure diagnostics. Experiments show that our approach achieves **98.93%** accuracy, increases high-cost failure precision from **60%** to **95%**, and improves recall by **25%**, offering a robust and explainable solution for predictive maintenance in commercial fleets.

**Keywords:** Fault Prediction, Predictive Maintenance, Graph Neural Networks,Machine Learning, Association Rule Mining, Vehicle Maintenance, Explainable AI

## 1. Introduction

Commercial vehicle fleets operate under tight reliability, safety, and cost constraints. Unplanned mechanical failures—especially those requiring emergency roadside service—cause downtime, elevate risk, and drive unexpected expenses. Although modern vehicles continuously log diagnostic codes, these logs are sparse, irregular, and heterogeneous, which limits the effectiveness of traditional sequence- or rule-based predictors at fleet scale.

Graph neural networks (GNNs) offer a natural way to model the relational structure behind these data. In practice, fault codes rarely occur in isolation: they cluster within a vehicle and recur across vehicles sharing operating conditions. Such patterns reflect temporal propagation, spatial proximity, and subsystem interdependence. Capturing these cross-fault and cross-chassis links improves prediction and supports engineering interpretability.

We propose a hybrid framework that turns raw fault sequences into directed, weighted graphs whose nodes represent chassis–fault events and whose edges encode temporal, spatial, and statistical co-occurrence relations. A GCN aggregates information over local and global neighborhoods to learn fault-propagation dynamics and latent hubs that conventional models overlook. The learned graph representations are passed to a Random Forest classifier, yielding robust predictions under severe class imbalance.

To align with real-world maintenance practice, we evaluate three tasks: predicting whether an emergency repair fee occurs, classifying repair cost into discrete tiers, and identifying the fault category among nine types derived from clustered fault descriptions. Beyond accuracy, the framework emphasizes interpretability by mining Apriori-based co-fault motifs that uncover subsystem dependencies. Unlike methods that treat fault codes independently or rely solely on temporal order, our approach unifies graph topology, operational context, and interpretable rules within a single framework. We develop a decoupled pipeline combining graph construction, GCN-based representation learning, and Random Forest classification, which achieves robust performance under severe class imbalance while preserving interpretability. The proposed graph formulation fuses temporal, spatial, and co-occurrence signals validated through association rules that provide diagnostic insight. Evaluations on multi-year industrial data provided by a large commercial vehicle manufacturer show nearly 99% accuracy and significant gains in high-cost fault precision and recall, underscoring both methodological soundness and practical value.

## 2. Related Work

Research on vehicle fault detection spans three principal methodological directions: statistical approaches, sequence modeling, and graph-based frameworks. Each of these paradigms targets different aspects of the complex, multi-modal data underlying commercial vehicle operation and fault occurrence.

**Statistical methods**, such as frequency analysis and association rule mining, have been widely used to identify frequently occurring or co-occurring fault codes in large datasets (Zhang et al., 2020, 2019). For example, sliding window analysis and threshold-based filtering are employed to monitor changes in fault patterns over time. While these techniques can provide fast, interpretable results, they are fundamentally limited by assumptions of stationarity and regularity in the underlying data (Theissler, 2017). In practical deployments, vehicle operational environments and fault occurrence patterns are dynamic and noisy, often leading to unreliable performance when using static or hand-crafted thresholds. Additionally, statistical approaches typically perform poorly in detecting rare but high-impact failures and may be sensitive to missing or erroneous data.

**Sequence modeling approaches** leverage temporal dependencies between fault events to provide more context-aware prediction and detection. Methods based on recurrent neural networks (RNNs), gated recurrent units (GRUs), and Transformer architectures have demonstrated improvements in tracking evolving fault sequences and extracting semantic relationships among diagnostic codes (Connor et al., 1994; Wen et al., 2023). For instance, GRU-based frameworks capture temporal progression of faults across usage cycles, and embedding-based methods utilize models such as Word2Vec and Transformers to uncover latent associations in code sequences. However, these models are often limited in their ability to represent long-range or non-adjacent dependencies, and may not adequately capture global relationships between system components or across multiple fault events.

**Graph-based frameworks** have recently gained traction for their capacity to explicitly model interactions and dependencies within complex systems (Jin et al., 2023, 2022). Heterogeneous and dynamic graph neural networks (GNNs) allow for the integration of spatial, operational, and relational information, improving the expressiveness and flexibility of predictive models. Despite these advances, most graph-based studies are restricted to static graphs or address only single-task prediction problems, limiting their ability to generalize across multiple failure types or operational contexts (Biddle and Fallah, 2021; Theissler, 2017; Choi et al., 2021; Namburu et al., 2019). The challenges of multi-fault interaction, transfer learning across vehicle platforms, and scalability in real-world settings are still largely unsolved.

Efforts to incorporate adaptive thresholding and real-time adjustment mechanisms have further improved model robustness under changing operational conditions (Moura et al., 2011; Namburu et al., 2019). However, these solutions are rarely unified with advanced sequence or graph learning approaches, and thus often fail to deliver a comprehensive answer to the joint challenges of explainability, class imbalance, and dynamic data distributions.

In summary, while significant progress has been made using statistical, sequential, and graph-based techniques, few existing works offer a holistic solution that jointly addresses dynamic, multi-fault, cross-platform, and explainable predictive maintenance for vehicle systems. This motivates our proposed framework, which integrates dynamic GNNs, ensemble learning, and adaptive thresholding in a unified, multi-task setting.

## 3. Problem Formulation & Tasks

**Data and notation.** Let $\mathcal{D} = \{(\mathcal{S}_i, \mathbf{z}_i)\}_{i=1}^{N}$ denote a fleet dataset, where each vehicle $i$ records a time-ordered sequence of fault events $\mathcal{S}_i = \langle (s_{i1}, \tau_{i1}), (s_{i2}, \tau_{i2}), \ldots \rangle$. Each diagnostic entry $s_{ij}$ corresponds to a fault code (e.g., SPN–FMI) with timestamp $\tau_{ij}$. Optionally, $\mathbf{z}_i$ aggregates vehicle-level statistics such as usage, region, or operational metrics. We represent the dataset as a directed attributed graph:

$$G = (V, E, X), \qquad V = \{\, v = (c, s, \tau) \,\}, \quad X \in \mathbb{R}^{|V| \times d},$$

where each node $v$ binds a chassis $c$, a fault code $s$, and an event time $\tau$. The feature matrix $X$ encodes temporal, spatial, and statistical properties. Edges $E$ capture (i) temporal adjacency within a chassis, (ii) spatial proximity across chassis, and (iii) co-occurrence relations mined from maintenance logs. Their weights are fused as

$$w_{uv} = \alpha\, w_{uv}^{\text{temp}} + \beta\, w_{uv}^{\text{spat}} + \gamma\, w_{uv}^{\text{cooc}}, \qquad \alpha, \beta, \gamma \geq 0.$$

**Empirical motivation.** Preliminary data analysis revealed that most emergency fees fall below 2,000 units, showing a strong skew toward lower costs. Several SPN–FMI combinations appear in both zero-fee and non-zero-fee cases, suggesting that structural faults can span multiple cost levels. These observations motivate a joint modeling strategy that predicts both the occurrence and the magnitude of emergency fees, alongside fault categories.

**Learning objectives.** Given $G$, an encoder $F_\theta$ (e.g., a graph neural network) produces node embeddings $\mathbf{H} = F_\theta(G) \in \mathbb{R}^{|V|\times d}$ and a pooled graph representation $\mathbf{g} = \mathrm{Pool}(\mathbf{H})$. We jointly optimize three prediction tasks, each associated with a loss function $\mathcal{L}$.We adopt two standard loss functions: the *binary cross-entropy* (BCE) for binary classification and the *categorical cross-entropy* (CE) for multi-class prediction. Given a graph $G$, an encoder $F_\theta$ (e.g., a graph neural network) produces node embeddings.
**Task 1: Emergency-Fee Existence (binary).** Predict whether a vehicle requires an emergency fee:

$$\hat{y}_1 = f_\theta(G) \in \{0, 1\}, \qquad \mathcal{L}_1 = \mathrm{BCE}(y_1, \hat{y}_1),$$

where BCE (binary cross-entropy) quantifies the discrepancy between the predicted probability $\hat{y}_1$ and the true label $y_1 \in \{0, 1\}$:

$$\mathrm{BCE}(y, \hat{y}) = -\big[y \log \hat{y} + (1 - y) \log(1 - \hat{y})\big].$$

**Task 2: Emergency-Fee Range (4-class).** Predict the cost range of the required fee:

$$\hat{y}_2 = g_\theta(G) \in \{0, 1, 2, 3\}, \qquad \mathcal{L}_2 = \mathrm{CE}(y_2, \hat{y}_2),$$

where CE denotes the standard cross-entropy loss for multi-class classification.
**Task 3: Fault-Category Prediction (9-class node classification).** For each node $v \in V$, predict the subsystem or fault category:

$$\hat{y}_3(v) = h_\theta(v \mid G) \in \{1, \ldots, 9\}, \qquad \mathcal{L}_3 = \frac{1}{|V|} \sum_{v \in V} \mathrm{CE}\big(y_3(v), \hat{y}_3(v)\big).$$

The overall loss function combines all objectives:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_3, \qquad \lambda_1, \lambda_2, \lambda_3 \geq 0.$$

The term *dynamic* emphasizes that graph structures evolve over time, reflecting event sequences and enabling incremental updates during real-time monitoring.

## 4. Data and Labeling

We describe the datasets, preprocessing, labeling, and fault-pattern mining used in this study. We first introduce the fault and maintenance records and their alignment for multi-task prediction, then motivate our cost-aware design through fee distributions and SPN–FMI patterns.

### 4.1. Dataset Overview and Preprocessing

We use two proprietary datasets provided by a large commercial vehicle manufacturer: (i) an automatically generated fault-log dataset and (ii) a maintenance dataset containing repair records and emergency fee information. These industrial datasets capture realistic operating conditions, diverse fault modes, and cost behaviors. Due to confidentiality constraints, the raw data are not publicly available; collaborations may be arranged under appropriate data-sharing agreements.

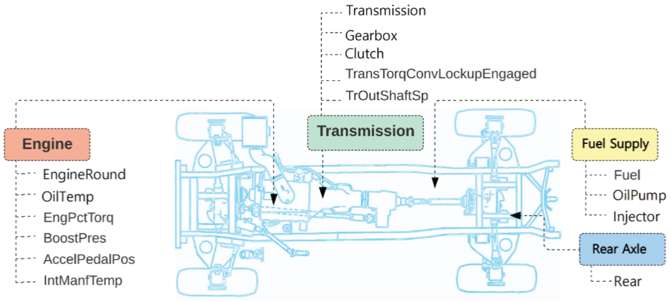

Figure 1: Reference diagram of automotive systems and fault categories.

Table 1: Keyword rules for fault clustering

| No. | Keywords |
|---|---|
| 1 | Rear Axle, Axle |
| 2 | Engine, Oil |
| ... | ... |
| 8 | ECU, Electronics, Motor |
| 9 | Other Fault Descriptions |

Each record contains *ChassisNumber, SystemType, FaultName, FaultCount, SPN, FMI, FaultStartTime*, and *FaultEndTime. Each fault event is uniquely identified by its SPN–FMI pair* (e.g., 1761–18), which we also use as the node-level code in downstream analyses. Although *FaultName* includes 700+ unique strings, only 135 distinct SPN–FMI pairs are observed, providing a compact representation. Records are grouped by *ChassisNumber* and sorted by *FaultStartTime* to form vehicle-specific sequences. Very rare SPN–FMI pairs (below a frequency threshold) are removed to reduce noise.

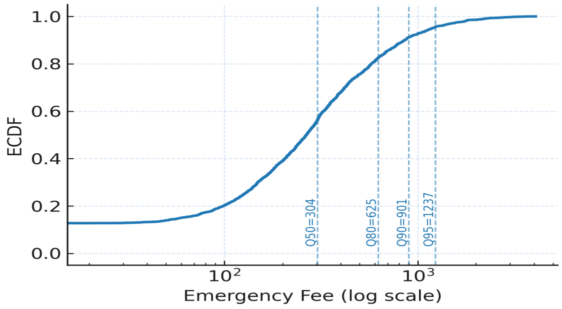

Figure 2: Emergency-fee ECDF with percentile markers.

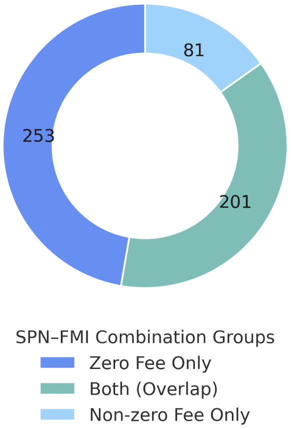

Figure 3: SPN–FMI groups by fee category (zero, non-zero, overlap).

The maintenance dataset includes *FaultDate, CheckDate, FaultCode, FaultDescription, PartName*, and *EmergencyFee*. Because descriptions span 200+ types, a keyword-augmented LSTM maps them into nine subsystem categories (Table 1; Fig. 1).

The ECDF (Empirical Cumulative Distribution Function) of emergency fees (Fig. 2) shows a strong left skew. SPN–FMI analysis (Fig. 3) reveals substantial overlap between zero-fee and non-zero-fee cases, indicating that structural faults span multiple cost levels. These observations motivate three labels: $Y_1$ emergency-fee existence ($y_1 \in \{0, 1\}$) from the zero/non-zero split; $Y_2$ emergency-fee range ($y_2 \in \{0, 1, 2, 3\}$) using four ECDF-derived tiers (e.g., 0–200, 200–1,000, etc.); and $Y_3$ fault category ($y_3(v) \in \{1, \ldots, 9\}$) from the NLP-based clustering (Table 1). This setup links technical patterns ($Y_3$) with economic outcomes ($Y_1$, $Y_2$) in a unified, cost-aware graph.

Table 2: Classification of SPN–FMI occurrence ranges.

| **Range** | **Proportion** | **Characteristics** |
|---|---|---|
| Low | 0–10% | Stable associations with minimal variation |
| Medium | 10–25% | Moderate variation, context-dependent patterns |
| High | >25% | Significant fluctuation, potential transient dependencies |

We extract co-occurring fault patterns via an optimized *Apriori* pipeline: (1) filter low-frequency SPN–FMI pairs; (2) build a sparse chassis–fault Boolean matrix; (3) compress rows/columns to reduce redundancy; (4) mine frequent itemsets at 1% support; and (5) index patterns with a trie for fast retrieval. The resulting combinations are categorized as in Table 2, supporting early warning and resource allocation in predictive maintenance.

## 5. Model Design and Graph Construction

We evaluated three modeling approaches for vehicle fault prediction: Support Vector Machines (SVM), Random Forests (RF), and our proposed hybrid model combining Graph Neural Networks (GNNs) with Random Forests (NRP-GCN-RF). The first two serve as baselines, highlighting the limitations of conventional machine learning on sparse, high-dimensional data. We then present the design of NRP-GCN-RF, which integrates graph-based relational learning with ensemble classification to capture both temporal and non-temporal dependencies in complex fleet fault data.

### 5.1. Support Vector Machines (SVM)

SVM models were trained on one-hot encoded SPN–FMI fault codes, yielding sparse binary matrices with tens of thousands of features. This creates the “large $p$, small $n$” problem, where the feature dimension greatly exceeds the number of samples, and overlap across vehicles is minimal. Inner-product–based kernels fail to capture meaningful similarities, leading to underfitting. Limited improvements were observed when applying feature selection based on frequent co-occurrence combinations, but SVM could not model relational dependencies or long-range interactions among fault codes.

Table 3: Simulated dataset showing the “large $p$, small $n$” problem with high-dimensional sparse features.

| **Chassis** | **FaultLabel** | **523004-16** | **520415-0** | **792-2** | **523014-8** | ... |
|---|---|---|---|---|---|---|
| C001 | A | 1 | 1 | 0 | 0 | ... |
| C001 | B | 1 | 1 | 0 | 0 | ... |
| C002 | A | 0 | 0 | 1 | 1 | ... |
| C002 | B | 0 | 0 | 1 | 1 | ... |
| C003 | A | 0 | 0 | 0 | 0 | ... |

### 5.2. Random Forests (RF)

Table 4: Restructured Fault Dataset with Tabular Attributes

| **ChassisNumber** | **FaultLabel** | **High Proportion Combinations** | **Counts** | **AvgLongitude** | **AvgLatitude** | **...** |
|---|---|---|---|---|---|---|
| C001 | 1 | [’523004-16’, ’520415-0’] | [5, 5] | 117.2451 | 39.3165 | ... |
| C001 | 2 | [’523004-16’, ’520415-0’] | [7, 6] | 117.2066 | 39.3501 | ... |
| | | ... | | | | |
| C002 | 1 | [’792-2’, ’523014-8’, ...] | [14, 14, ...] | 122.9887 | 41.0847 | ... |
| C003 | 1 | [’518110-5’, ’518118-5’, ...] | [14, 13, ...] | 103.0202 | 43.2948 | ... |

RF models were trained on tabular features including frequent combinations, counts, proportions, and geospatial attributes. While RF captured structured patterns and achieved solid baseline performance, it struggled with rare but high-cost failures due to class imbalance. To mitigate this, we applied SMOTE to synthesize new minority samples:

$$\hat{x}_{\text{new}} = x_{\text{minority}} + \lambda \cdot (x_{\text{neighbor}} - x_{\text{minority}}), \quad \lambda \sim \text{Uniform}(0, 1).$$

This improved minority recall, but RF alone could not model temporal sequences or inter-chassis dependencies.

### 5.3. Proposed Hybrid Model: NRP-GCN-RF

NRP-GCN-RF encodes fleet fault logs as a directed, attributed graph and *decouples* relational representation learning (GNN) from classification (Random Forest). The intent is simple: represent how events unfold within a chassis (temporal order) and how they relate across chassis (spatial proximity and co-occurrence), so we can detect rare, high-cost failures while keeping the model easy to interpret.

For each chassis $c$, the raw log is a sequence $\mathcal{S}_c = \{(s_t, \tau_t, \ell_t)\}_{t=1}^{T_c}$ with SPN–FMI code $s_t$, timestamp $\tau_t$, and location $\ell_t$. We build

$$G = (V, E, X), \qquad V = \{v_i = (c_i, s_i, \tau_i)\}, \qquad X \in \mathbb{R}^{|V| \times d},$$

where $X$ stacks code embeddings, temporal lags (e.g., time since last fault), chassis metadata, and spatial statistics. As shown in Fig. 4, time-sorted sequences are projected into a unified graph; Fig. 5 shows a typical chassis subgraph.

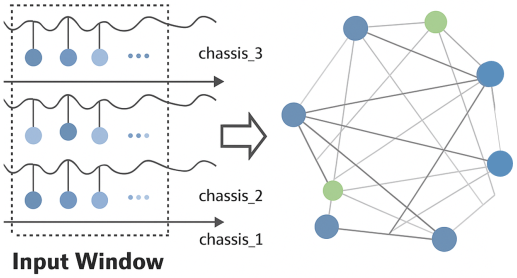

Figure 4: Graph conversion from time-sorted logs to nodes/edges.

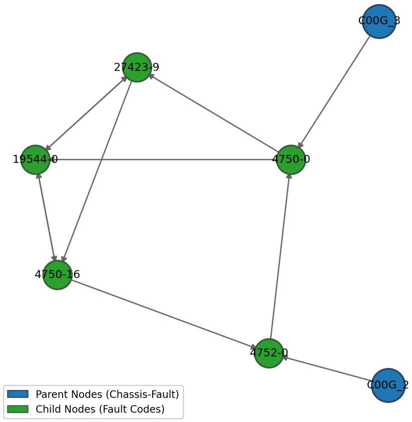

Figure 5: Directed subgraph for a chassis with temporal, spatial, and association links.

A directed edge $i \to j$ is added if at least one condition holds: (i) temporal adjacency within a window $\Delta_t$; (ii) spatial proximity with geodesic distance $d_{\text{geo}}(\ell_i, \ell_j) \le \Delta_s$ (e.g., radius $r$ km or 0.05° in lat/lon); (iii) statistical co-occurrence where $(s_i, s_j)$ satisfies Apriori support/confidence thresholds. We fuse the heterogeneous evidence as

$$w_{ij} = \alpha \, \text{sim}_t(v_i, v_j) + \beta \, \text{sim}_s(v_i, v_j) + \gamma \, \text{assoc}(v_i, v_j), \qquad \alpha, \beta, \gamma \ge 0, \;\; \alpha + \beta + \gamma = 1, \tag{1}$$

where $\text{sim}_t$ is a decayed temporal kernel, $\text{sim}_s$ a spatial kernel (e.g., an RBF of $d_{\text{geo}}$), and assoc a normalized Apriori statistic (confidence or lift). Edges with $w_{ij} \ge \varepsilon$ are kept and the adjacency is row-normalized to obtain $\tilde{A}$. The construction is dynamic: when a new event arrives we insert $v_{\text{new}}$ and update incident edges via (1).

Node embeddings are learned as $H = \text{GNN}(X, \tilde{A})$. For a GCN instantiation,

$$h_v^{(k)} = \sigma\left( \sum_{u \in \mathcal{N}(v)} \frac{1}{\sqrt{d_v d_u}} W^{(k)} h_u^{(k-1)} \right), \qquad h^{(0)} = X, \tag{2}$$

with trainable weights $W^{(k)}$ and nonlinearity $\sigma$. The encoder is interchangeable—GIN, GAT, or a Graph Transformer can be used under the same tuning budget. Decisions are made by a Random Forest on pooled graph embeddings for graph-level outputs and on per-node embeddings for node-level outputs:

$$\hat{y}^{\text{graph}} = \text{RF}(z_G), \qquad z_G = \text{pool}\big(\{h_v\}_{v \in V}\big), \qquad \hat{y}^{\text{node}} = \text{RF}(h_v).$$

This separation improves robustness under class imbalance and label noise, and provides feature importance for diagnostics.

To control density before training, we prune by $\Delta_t$, $\Delta_s$, and the threshold $\varepsilon$ within (chassis, time-window) shards, and parallelize construction across shards. Peak host memory per worker scales as $\tilde{O}(|V_{\text{win}}| + |E_{\text{win}}|)$, and empirically $|E| \approx c\,|V|$ with a small constant

$c$, yielding near-linear build time. With neighborhood sampling (batch size $B$, hidden width $d$, capped neighbors $\bar{k}$, $K$ layers), message passing costs $O(K\,B\,d\,\bar{k})$ per batch; Apriori mining remains tractable using a minimum support (e.g., $\sigma \geq 1\%$), rare-binning of infrequent codes, and an FP-Growth fallback for long tails.

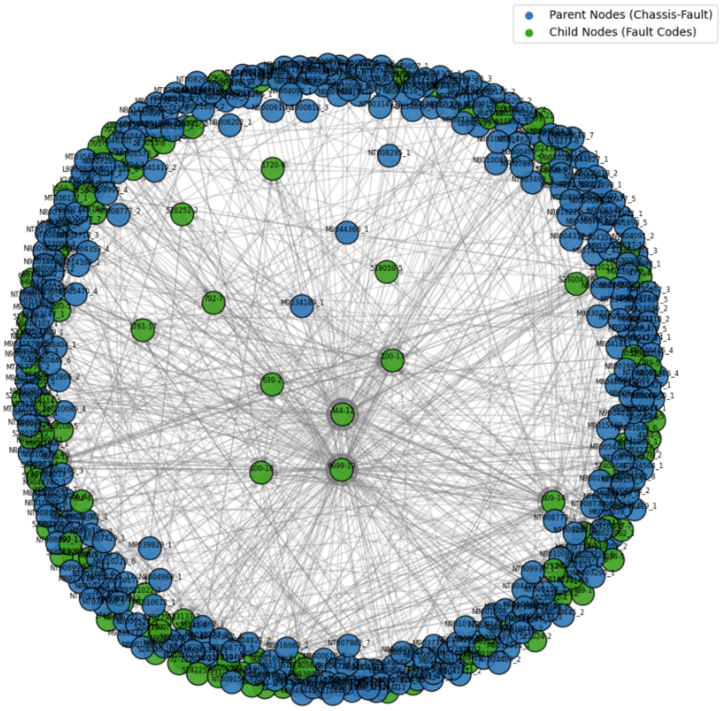

Figure 6: Global fault graph spanning inter-vehicle and temporal relations.

Overall, the formulation yields a multi-factor graph abstraction of fleet faults, couples an encoder-agnostic GNN for representation with a lightweight Random Forest decision head, and is supported by three figures that respectively illustrate the log-to-graph conversion, the mechanics of a local chassis subgraph, and the global fleet structure.

## 6. Experiments and Evaluation

We evaluate NRP-GCN-RF on three tasks: $Y_1$ Emergency-Fee Existence (binary, graph-level), $Y_2$ Emergency-Fee Range (multi-class, graph-level), and $Y_3$ Fault-Category Prediction (multi-class, node-level). Unless noted, we use *vehicle-wise, chronological* $K$-fold CV to avoid temporal and inter-vehicle leakage. We report standard metrics; ROC/PR curves and confusion matrices are in the Appendix.

Hyperparameters (GNN depth/width and edge-weight coefficients $\alpha, \beta, \gamma$ in Eq. (1)) are tuned on validation folds. For $Y_1$ we report AUROC, AUPRC, F1, and Precision@Top-$k$; for $Y_2$ macro/weighted-F1 and QWK; for $Y_3$ micro/macro-F1 and per-class recall. Calibration (ECE) and cost-aware operating points are analyzed in the Appendix.

### 6.1. Results

Table 5 summarizes the headline results across the three prediction targets $Y_1$–$Y_3$. The proposed **NRP-GCN-RF** consistently outperforms classical baselines (SVM, RF) with clear recall and F1-score gains on $Y_1$ and $Y_2$, while maintaining strong calibration across all tasks. This demonstrates that integrating relational representations through a graph encoder, followed by a decoupled Random Forest head, enhances both sensitivity to rare costly faults and generalization across heterogeneous subsystems.

Table 5: Overall performance across three tasks (best results in bold).

| Model | $Y_1$: HasEmergencyFee | | | $Y_2$: EmergencyFeeCategory | | | $Y_3$: FaultCategory | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1-Score | Precision | Recall | F1-Score | Precision | Recall | F1-Score |
| SVM | 0.74 | 0.74 | 0.74 | 0.74 | 0.72 | 0.73 | 0.80 | 0.80 | 0.79 |
| RF | 0.90 | 0.90 | 0.90 | 0.92 | 0.92 | 0.92 | 0.95 | 0.95 | 0.95 |
| **NRP-GCN-RF** | **0.94** | **0.94** | **0.94** | **0.93** | **0.93** | **0.93** | **0.98** | **0.99** | **0.99** |

*Task-wise analysis.* For $Y_1$ (*Emergency-Fee Existence*), recall on the positive class is notably high (Table 6), showing the model's strong ability to identify vehicles likely to incur costly emergency repairs. This property is particularly desirable in predictive maintenance, where false negatives (missed costly cases) are more harmful than false positives.

For $Y_2$ (*Emergency-Fee Range*), all fee intervals perform robustly (Table 7). Class 3.0 achieves perfect recall, and Class 0.0 achieves perfect precision, reflecting the model's *high sensitivity* for expensive cases and its *low false-alarm rate* for benign ones. This result suggests that relational patterns across fault clusters help the model distinguish between high-cost and low-cost maintenance scenarios.

For $Y_3$ (*Fault-Category*), performance remains uniformly high across all nine subsystem classes, with overall accuracy reaching 0.989 (Table 8). Minor variation in Class 1 (with limited support) highlights the data imbalance, yet macro and weighted averages both remain near unity, indicating generalizable fault discrimination capability across the entire chassis network.

Table 6: Classification report for $Y_1$: *Emergency-Fee Existence.*

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.91 | 0.84 | 0.87 | 25 |
| 1 | 0.94 | 0.97 | 0.96 | 67 |
| Accuracy | 0.9348 (on 92 samples) | | | |
| Macro Avg | 0.93 | 0.91 | 0.92 | 92 |
| Weighted Avg | 0.93 | 0.93 | 0.93 | 92 |

Table 7: Classification report for $Y_2$: *Emergency-Fee Range.*

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0.0 | 1.00 | 0.82 | 0.90 | 17 |
| 1.0 | 0.92 | 0.94 | 0.93 | 81 |
| 2.0 | 0.96 | 0.93 | 0.95 | 86 |
| 3.0 | 0.95 | 1.00 | 0.97 | 73 |
| Accuracy | 0.9455 (on 257 samples) | | | |
| Macro Avg | 0.96 | 0.92 | 0.94 | 257 |
| Weighted Avg | 0.95 | 0.95 | 0.95 | 257 |

*Takeaway.* A graph encoder with a decoupled RF head improves recall on costly events while maintaining calibration, demonstrating that structured relational learning can simultaneously enhance early warning capability and preserve interpretability.

Table 8: Classification report for $Y_3$: *Fault-Category* (all classes).

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| 0 | 0.96 | 1.00 | 0.98 | 50 |
| 1 | 1.00 | 0.67 | 0.80 | 3 |
| 2 | 1.00 | 0.98 | 0.99 | 62 |
| 3 | 0.99 | 1.00 | 0.99 | 66 |
| 4 | 1.00 | 0.98 | 0.99 | 54 |
| 5 | 0.98 | 1.00 | 0.99 | 55 |
| 6 | 0.98 | 0.96 | 0.97 | 53 |
| 7 | 1.00 | 1.00 | 1.00 | 67 |
| 8 | 1.00 | 1.00 | 1.00 | 57 |
| Accuracy | 0.989 (on 467 samples) | | | |
| Macro Avg | 0.99 | 0.95 | 0.97 | 467 |
| Weighted Avg | 0.99 | 0.99 | 0.99 | 467 |

### 6.2. Learning Capacity and Feature Contributions

Permutation importance is computed on the RF head with out-of-fold predictions under the same cross-validation protocol. Inputs include graph-aggregated operational features (`Count`, `Proportion`), geospatial descriptors (`AvgLatitude`, `AvgLongitude`, `LatitudeDiff`, `LongitudeDiff`), and Apriori-stable combination indicators (e.g., `Combination_6699-19`, `Combination_609-14`, `Combination_444-12`, `Combination_100-18`). The resulting importances are min–max normalized per task and summarized by the median across folds. As shown in Figure 7, `AvgLatitude` dominates ($\approx 0.15$), followed by `LatitudeDiff` and `AvgLongitude`. Operational features rank mid-range, while combination flags contribute mainly through graph connectivity and message passing rather than as isolated tabular indicators.

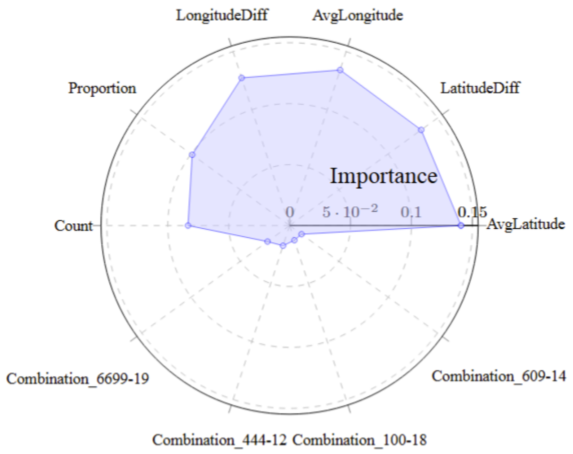

Figure 7: **Permutation-importance radar.** Normalized importances across tasks ($Y_1$–$Y_3$) for geospatial, operational, and combination features.

Learning and validation behaviors are summarized in Figure 8, which plots accuracy versus training size (top) and number of trees (bottom). Solid lines denote fold means, and

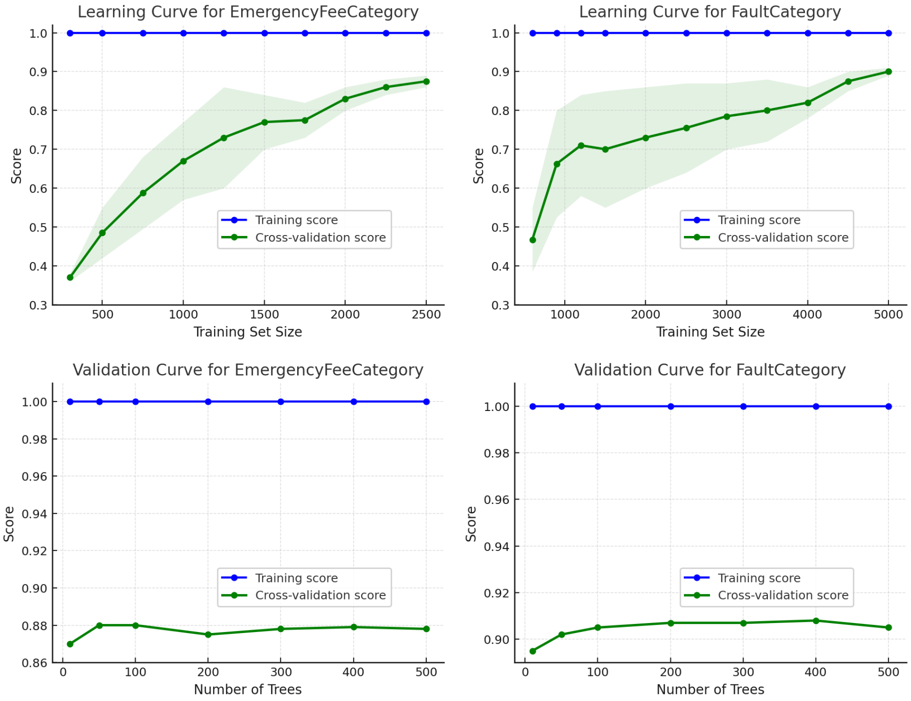

Figure 8: **Learning and validation curves.** Accuracy versus training-set size (top) and number of trees (bottom) for $Y_2$ and $Y_3$ under vehicle-wise chronological CV.

shaded regions indicate the interquartile range (IQR). The results show that $Y_2$ (emergency-fee range) benefits from larger datasets and deeper forests, whereas $Y_3$ (fault-category classification) plateaus earlier, suggesting feature saturation. The remaining train–validation gap indicates potential gains from broader coverage or additional regularization strategies.

Figure 8 shows accuracy vs. training size (top) and vs. number of trees (bottom). Solid lines: fold means; shaded: IQR.

$Y_2$ benefits from more data and deeper forests; $Y_3$ plateaus earlier, hinting at feature saturation. The residual train–validation gap suggests gains from broader coverage or additional regularization.

## 6.3. Fault Code Association Analysis

To complement supervised prediction with interpretable structure, we apply the Apriori algorithm to mine frequently co-occurring SPN–FMI pairs from per-chassis transaction windows. These association rules not only inform graph edge design (Sec. 5.3) but also support subsystem-level diagnostics, enabling interpretable connections between fault patterns and underlying vehicle systems.

Given a transaction set $T$, Apriori extracts frequent itemsets that satisfy a minimum support threshold and generates rules $A \Rightarrow B$ that meet minimum confidence requirements:

$$\mathrm{supp}(A) = \frac{|T_A|}{|T|}, \quad \mathrm{conf}(A \Rightarrow B) = \frac{\mathrm{supp}(A \cup B)}{\mathrm{supp}(A)}, \quad \mathrm{lift}(A \Rightarrow B) = \frac{\mathrm{supp}(A \cup B)}{\mathrm{supp}(A)\,\mathrm{supp}(B)}.$$

To avoid spurious high lift on rare items, we impose domain-informed minimum support thresholds and verify rule stability across folds. This ensures that the extracted associations represent genuine subsystem relationships rather than statistical artifacts.

Representative results are summarized in Table 9. High-lift rules with perfect confidence indicate strong functional linkages within the same subsystem, while moderate-lift cross-module rules capture looser correlations that may reflect indirect dependencies or cascading effects. For instance, (1761-18) ⇒ (1241-1) achieves a lift of 98, revealing a strong co-occurrence pattern between two key SPN–FMI combinations.

Table 9: Association rules discovered by Apriori.

Strongly associated rules.

| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
|---|---|---|---|---|
| (1761-18) | (1241-1) | 0.0102 | 1.00 | 98.0 |
| (520243-0) | (639-5) | 0.0102 | 1.00 | 98.0 |
| (523053-8) | (523056-8) | 0.0204 | 1.00 | 49.0 |
| (518112-5) | (131-7) | 0.0204 | 1.00 | 49.0 |
| (4794-1) | (523006-4) | 0.0102 | 1.00 | 32.7 |

Moderately associated rules.

| **Antecedents** | **Consequents** | **Support** | **Confidence** | **Lift** |
|---|---|---|---|---|
| (524040-20) | (597-13) | 0.0196 | 1.00 | 51.0 |
| (518108-5) | (790-5) | 0.0196 | 1.00 | 34.0 |
| (521022-2) | (520252-2) | 0.0588 | 0.86 | 14.6 |
| (1720-9) | (1624-2) | 0.0882 | 0.90 | 10.2 |
| (518114-5) | (4750-16) | 0.0294 | 0.75 | 6.38 |

High-lift pairs reveal subsystem interdependencies—for example, links between the Electronic Engine Control Unit (EECU) and induction-system faults—while moderate-lift cross-module rules suggest potential cascading or co-occurring behaviors. These interpretable associations reinforce the soundness of our graph-construction design and provide actionable cues for fault tracing and preventive maintenance scheduling.

### 6.4. Cross-Validation Performance

Table 10 reports fold-wise accuracies, indicating stable generalization under vehicle-wise chronological splits.

Table 10: Cross-validation accuracy across tasks with NRP-GCN-RF.

| **Task** | **Fold Accuracies** | **Mean Accuracy** | **Std. Dev.** |
|---|---|---|---|
| $Y_1$: Emergency-Fee Existence | 0.9130, 0.9565, 0.8913, 0.9333, 0.8889 | 0.9166 | 0.0257 |
| $Y_2$: Emergency-Fee Range | 0.9690, 0.9297, 0.9609, 0.9844, 0.9375 | 0.9563 | 0.0202 |
| $Y_3$: FaultCategory | 0.9968, 0.9838, 0.9903, 0.9870, 0.9773 | 0.9870 | 0.0065 |

*Summary.* NRP-GCN-RF attains high accuracy with low variance, strong sensitivity to costly events, and interpretable associations that aid downstream diagnostics.

## 7. Conclusion and Future Work

This paper presents **NRP-GCN-RF**, a hybrid framework that combines graph representation learning with ensemble classification for predictive maintenance in commercial vehicle fleets. By transforming chassis-level fault logs into directed, weighted graphs and fusing operational context, the model captures structural dependencies and statistical regularities. In parallel, Apriori-based rule mining exposes frequent co-fault motifs that support interpretability. Evaluated on three targets—*Emergency fee existence* ($Y_1$), *fee range* ($Y_2$), and *fault category* ($Y_3$)—NRP-GCN-RF outperforms baselines under class imbalance and cost-sensitive conditions.

Future work will extend the framework to dynamic graphs for evolving fault propagation, investigate explainable GNN modules (e.g., attention-based rationales) for traceable decisions, and study cross-fleet generalization by applying the framework to multi-vendor datasets and real-time telemetry. Future research will also explore validation and deployment of the proposed model across a broader range of industry vehicle data in collaboration with commercial partners, supporting real-world predictive maintenance applications.

## 8. Acknowledgement

Our work was supported in part by the Guangdong Provincial Key Laboratory of IRADS (2022B1212010006) and in part by the Guangdong Higher Education Upgrading Plan (2021–2025) with No. UICR0400032-22 at Beijing Normal University–Hong Kong Baptist University United International College, Zhuhai, PR China. We also gratefully acknowledge the industry partner who provided the real-world vehicle datasets used in this paper.

## References

Liam Biddle and Saber Fallah. A novel fault detection, identification and prediction approach for autonomous vehicle controllers using SVM. *Automotive Innovation*, 4(4): 301–314, 2021. doi: 10.1007/s42154-021-00138-0. URL https://doi.org/10.1007/s42154-021-00138-0.

B. Biller and B. L. Nelson. Modeling and generating multivariate time-series input processes using a vector autoregressive technique. *ACM Transactions on Modeling and Computer Simulation*, 13(3):211–237, 2003. doi: 10.1145/858527.858528. URL https://doi.org/10.1145/858527.858528.

G. E. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970. doi: 10.1080/01621459.1970.10481199. URL https://doi.org/10.1080/01621459.1970.10481199.

L.-J. Cao and F. E. H. Tay. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 14(6):1506–1518, 2003. doi: 10.1109/TNN.2003.819501. URL https://doi.org/10.1109/TNN.2003.819501.

F. Chen et al. Vehicle maintenance demand prediction: A survey. *Applied Sciences*, 15(20):11095, 2025. doi: 10.3390/app152011095. URL https://doi.org/10.3390/app152011095.

Kukjin Choi, Jihun Yi, Changhwa Park, and Sungroh Yoon. Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines. *IEEE Access*, 9:120020–120035, 2021. doi: 10.1109/ACCESS.2021.3107975. URL https://doi.org/10.1109/ACCESS.2021.3107975.

J. T. Connor, R. D. Martin, and L. E. Atlas. Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2):240–254, 1994. doi: 10.1109/72.279181. URL https://doi.org/10.1109/72.279181.

Apostolos Giannoulidis and Anastasios Gounaris. A context-aware unsupervised predictive maintenance solution for fleet management. *Journal of Intelligent Information Systems*, 60(2):521–547, 2023. doi: 10.1007/s10844-022-00744-2. URL https://link.springer.com/article/10.1007/s10844-022-00744-2.

Ark Ifeanyi. A graph neural network approach to system-level health index and remaining useful life estimation. In *Annual Conference of the PHM Society*, 2024. doi: 10.36001/phmconf.2024.v16i1.4159. URL https://doi.org/10.36001/phmconf.2024.v16i1.4159.

G. Jin, Y. Liang, Y. Fang, J. Huang, J. Zhang, and Y. Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. *arXiv preprint arXiv:2303.14483*, 2023. URL https://arxiv.org/abs/2303.14483.

M. Jin, Y. Zheng, Y.-F. Li, S. Chen, B. Yang, and S. Pan. Multivariate time series forecasting with dynamic graph neural ODEs. *IEEE Transactions on Knowledge and Data Engineering*, 34(2):557–570, 2022. doi: 10.1109/TKDE.2021.3057342. URL https://doi.org/10.1109/TKDE.2021.3057342.

Y. Mahale et al. A comprehensive review on artificial intelligence driven vehicle maintenance strategies and diagnostics. *SN Applied Sciences*, 2025. doi: 10.1007/s42452-025-06681-3. URL https://link.springer.com/article/10.1007/s42452-025-06681-3.

M. C. Moura, E. Zio, I. D. Lins, and E. Droguett. Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(11):1527–1534, 2011. doi: 10.1016/j.ress.2011.06.004. URL https://doi.org/10.1016/j.ress.2011.06.004.

S. M. Namburu, M. Wilcutts, S. Chigusa, L. Qiao, K. Choi, and K. Pattipati. Systematic data-driven approach to real-time fault detection and diagnosis in automotive engines. *Proceedings of the IEEE*, 107(3):522–536, 2019. doi: 10.1109/JPROC.2018.2871491. URL https://doi.org/10.1109/JPROC.2018.2871491.

Subash Neupane, Ivan A. Fernandez, Wilson Patterson, Sudip Mittal, and Shahram Rahimi. A temporal anomaly detection system for vehicles utilizing functional working groups and sensor channels. In *Proceedings of the 2022 IEEE 8th International Conference on Collaboration and Internet Computing (CIC)*, page 24, 2022. doi: 10.1109/CIC56439.2022.00024. URL https://ieeexplore.ieee.org/document/9962393.

Moirangthem T. Singh, Rabinder K. Prasad, Gurumayum Robert Michael, N. Hemarjit Singh, and N. K. Kaphungkui. Spatial-temporal bearing fault detection using graph attention networks and LSTM. *Sensors*, 2024. URL https://arxiv.org/abs/2410.11923. preprint at arXiv:2410.11923; full journal details pending.

Andreas Theissler. Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection. *Knowledge-Based Systems*, 123:163–173, 2017. doi: 10.1016/j.knosys.2017.02.023. URL https://doi.org/10.1016/j.knosys.2017.02.023.

J. Wang, C. Zhang, X. Ma, Z. Wang, Y. Xu, and R. Cattley. A multivariate statistics-based approach for detecting diesel engine faults with weak signatures. *IEEE Transactions on Industrial Electronics*, 67(5):3923–3931, 2020. doi: 10.1109/TIE.2019.2958450. URL https://doi.org/10.1109/TIE.2019.2958450.

Min Wang and Xijun Zhu. Association rule analysis based on improved apriori algorithm. *Journal of Qingdao University of Science and Technology*, 42(6):413–416, 2021. URL https://kns.cnki.net/kcms/detail/detail.aspx?FileName=QKXY202106016&DbName=CJFQ2021.

Xiaoyang Wang, Zhiqiang Wei, Wenjuan Zhu, and Qing Li. A graph neural network approach for fault diagnosis and prognosis in industrial systems. *IEEE Transactions on Industrial Informatics*, 18(2):1023–1033, 2022. doi: 10.1109/TII.2021.3099806. URL https://doi.org/10.1109/TII.2021.3099806.

Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun. Transformers in time series: A survey. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2023. URL https://www.ijcai.org/proceedings/2023/109.

Y. Xia and J. Chen. Traffic flow forecasting method based on gradient boosting decision tree. In *Proceedings of the International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT)*, pages 413–416, 2017. doi: 10.2991/fmsmt-17.2017.87. URL https://doi.org/10.2991/fmsmt-17.2017.87.

Chuxu Zhang, Dongjin Song, Yuncong Chen, Xinyang Feng, Cristian Lumezanu, Wei Cheng, Jingchao Ni, Bo Zong, Haifeng Chen, and Nitesh V. Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019. doi: 10.1609/aaai.v33i01.33010429. URL https://ojs.aaai.org/index.php/AAAI/article/view/3942.

Minghu Zhang, Jianwen Guo, Xin Li, and Rui Jin. Data-driven anomaly detection approach for time-series streaming data. *Sensors*, 20(19):5646, 2020. doi: 10.3390/s20195646. URL https://doi.org/10.3390/s20195646.

B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu. Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1):162–169, 2017. doi: 10.1109/JSEE.2017.00036. URL https://doi.org/10.1109/JSEE.2017.00036.

Geng Zheng and M. Tong. Vehicle fault code analysis based on association rule mining. *Automotive Engineering*, 45(4):699–707, 2023. doi: 10.1145/3672758.3672791. URL https://doi.org/10.1145/3672758.3672791.