Proceedings of the Seventh International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 15, Paper 8.3
Civil-Comp Press, Edinburgh, United Kingdom, 2026
ISSN: 2753-3239, doi: 10.4203/ccc.15.8.3

Risk-Based Predictive Maintenance of EMU Autocouplers: A Hybrid AI Modelling Approach

M. M. Rahman[1], B. Alkali[1], J. P. Gutierrez[1], A. K. Jain[2], C. Mcneil[3] and J. Nelson[3]

[1]School of Science & Engineering, Glasgow Caledonian University, United Kingdom
[2]Department of Mechanical Engineering, Birla Institute of Technology and Science, India
[3]Siemens Mobility Limited, Glasgow, United Kingdom

Abstract

The automated coupler is a safety-critical interface in Electric Multiple Units (EMUs), where failures can lead to significant operational disruptions and safety risks. This study proposes a hybrid data-driven framework for the predictive maintenance of EMU autocouplers, analysing 4.5 years of operational data involving over 17 million fleet miles. The methodology integrates Failure Mode and Effects Analysis (FMEA) with machine learning to enhance risk assessment. A critical step in this framework involves converting text-based failure logs into a structured numerical format using text mining methods to extract meaningful failure patterns. A cost-sensitive Support Vector Machine (SVM) is then employed to classify failure modes into distinct Low, Medium, and High-risk categories, achieving 97% overall accuracy and a critical 100% recall for High-risk failures. Discrete-Time Markov Chain (DTMC) model is used to forecasts the short-term evolution of these risk states over a 7-day horizon. Results indicate that electrical faults contribute most to high-risk predictions, while mechanical and pneumatic categories show more stable, wear-driven risk behaviour. The proposed framework enables maintenance teams to transition from reactive repairs to predictive, risk-prioritised interventions with clear component-level insights.

Keywords: predictive maintenance, autocoupler, support vector machine , Markov chain, failure mode and effects analysis, railway reliability.

1 Introduction

The automated coupler is a safety-critical interface that lets Electric Multiple Units (EMUs) be coupled and uncoupled quickly while in service. This enables flexible train formations and quick operational turnarounds. Modern automatic couplers not only connect vehicles mechanically but also pneumatically and electrically. This lets coupled units work together as a single train. This paper investigates the Desiro UK ScotRail Class 380 autocoupler failures in the operation of 38 Electric Multiple Units (EMUs) in passenger service in Scotland, including 3-car and 4-car formations. The coupler is designed to support automatic coupling under horizontal and vertical misalignment, while accommodating vertical, horizontal, and torsional movements during operation. The assembly integrates the coupler head and shank, uncoupling actuation, pneumatic interfaces, the electrical head and operating gear, and overload protection intended to limit structural damage under high-impact conditions. Figure 1 shows the main parts of the coupler.

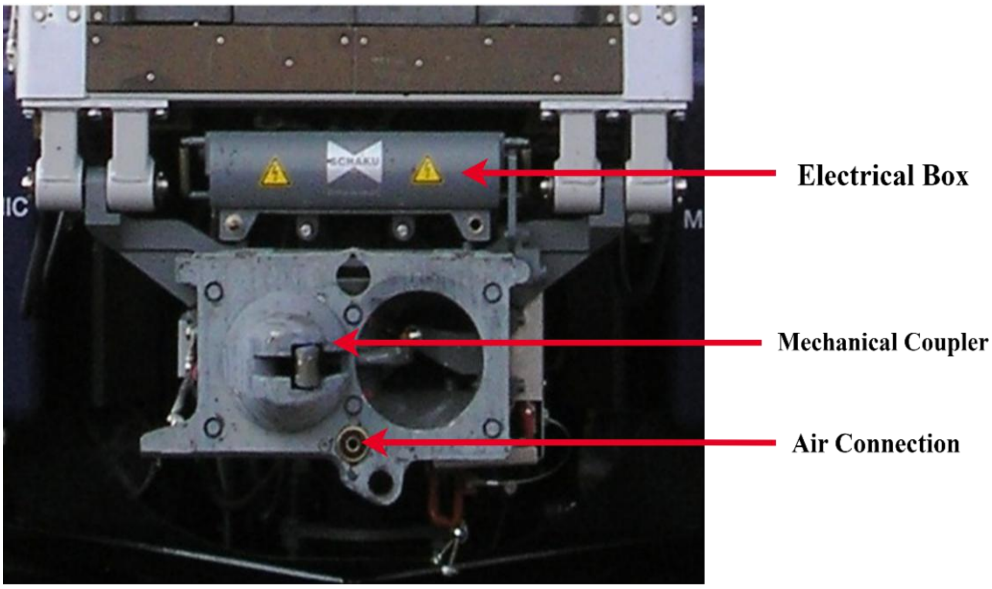

Figure 1: Autocoupler

Autocoupler failures originate from various domains, including mechanical wear, pneumatic failures, electrical contact issues, locking failures, and control monitoring functions, each of which can cause service disruptions. During a 4.5-year (January 2021 to May 2025), the fleet recorded a total mileage of 17,224,547 and approximately 27,000 operational hours. During this period, the reliability of autocouplers became a significant factor influencing operational performance: 471 autocoupler failures were recorded, resulting in 447 minutes of delay, 48 Full cancellations, and 22 partial cancellations. The effects demonstrate the operational sensitivity of coupling-uncoupling functionality and trainline continuity to coupler state, underscoring the necessity for risk assessment methodologies that are both engineering-interpretable and operationally actionable. Failure Modes and Effects Analysis (FMEA) is used to look at these kinds of breakdowns. FMEA helps systematically identify failure modes, their effects, and the most effective ways to fix them. It does this by using structured scores that usually include occurrence (A), severity (B), and detection (E) to apply a Risk Priority Number (RPN). This study bridges the gap by integrating interpretable FMEA-driven risk levels with data-driven modelling to enhance operational planning. Initially, we employ supervised learning to categorise discrete risk levels based on attributes derived from FMEA and then model the evolution of these risk states using a probabilistic state-transition framework to generate short-term (7-day) risk forecasts. The outcome is a framework that maintains technical interpretability while offering decision support more closely connected with daily fleet maintenance and service delivery.

2 Fleet data performance

The pre-processed autocoupler FMEA were grouped into five failure-type categories: Electrical Failure, Control-Protection-monitoring, Mechanical Failure, Pneumatic Coupler, and Buffer-Draft Gear.

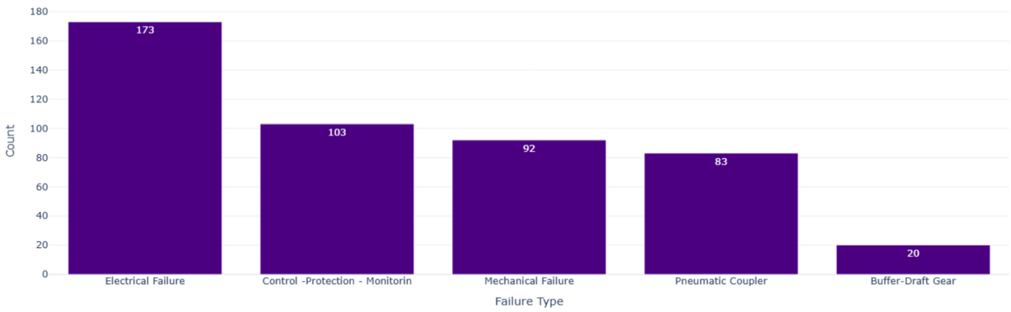

Figure 2: Autocoupler Classification failure

The resulting distribution is summarised in Figure 2. Electrical Failure represents the largest group (173 records), followed by Control-Protection-Monitoring (103), Mechanical Failure (92), Pneumatic Coupler (83), and Buffer-Draft Gear (20), where Electrical Failure accounts for 36.7% of the dataset, Control-Protection-Monitoring for 21.8%, Mechanical Failure for 19.5%, Pneumatic Coupler for 17.6%, and Buffer-Draft Gear for 4.2%. Together, these results indicate that autocoupler risk evidence is dominated by electrical and control-monitoring-related failure types, supporting the need for multi-domain risk modelling.

3 Modelling and Analysis Framework

The main application of this approach is the automation of RPN estimation. ML systems can predict FMEA values and RPN results by learning from historical FMEA data rather than relying on expert ratings. Sader et al. [1] developed an innovative FMEA enhancement for agricultural machinery by training a classifier to predict FMEA ratings from failure descriptions and related features. The system used four classification models to predict Severity, Occurrence, Detection, and failure category ratings, achieving high accuracy. Naranjo et al. [2] employed Random Forest regression to predict RPN values from process-FMEA data, achieving precise predictions. Peddi et al. [3] created a "modified FMEA" for food supply chain applications that employs a machine learning model to forecast RPN values in real time. Xiaojun et al. [4] present an FMEA model utilising a unified probabilistic framework that integrates both fuzzy and stochastic uncertainty. Use Markov Chain Monte Carlo (MCMC) to estimate uncertain parameters and improve the reliability of RPN rankings. The modelling workflow framework is illustrated in the steps in Figure 4.

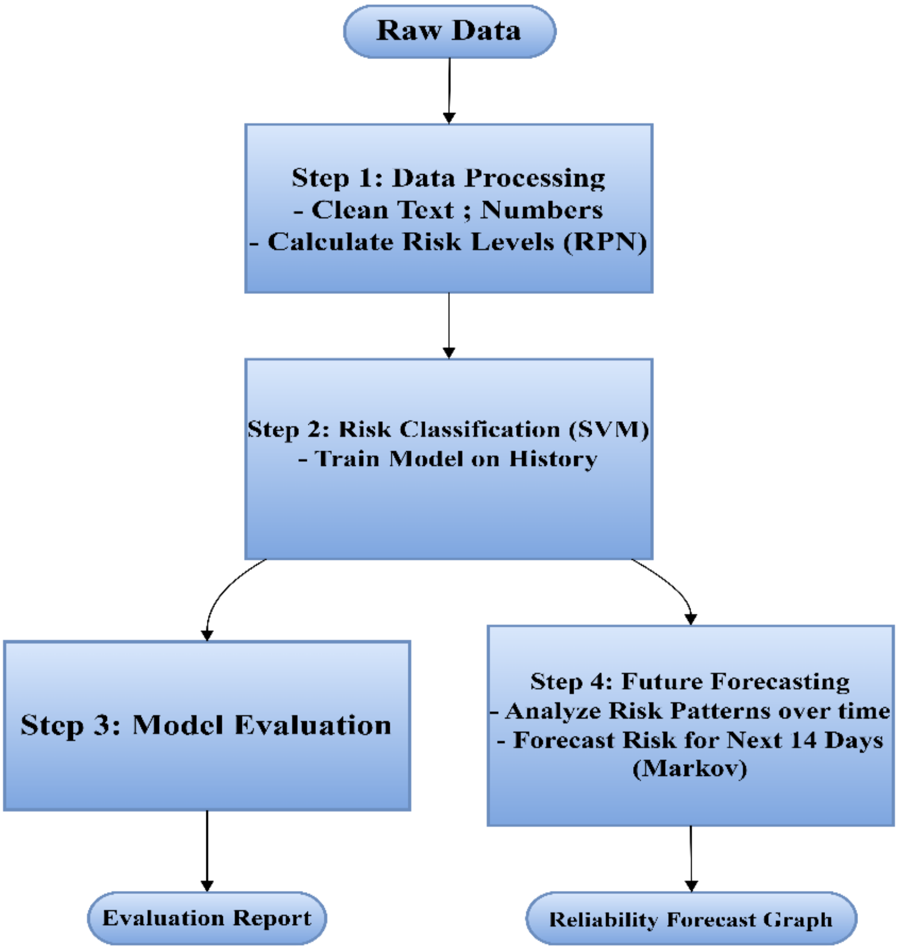

Figure 4: Flow chart of autocoupler failure model

Yang et al. [5] enhance the precision of goods volume forecasting by integrating a Markov chain with the Grey GM(1,1) model. The Markov model corrects forecast errors by modelling state transitions, while the Grey model identifies trends in small datasets. It is useful for transportation planning because it makes predictions more accurate, especially when the data are random and change quickly. This paper proposes a hybrid data-driven framework for the predictive maintenance of autocoupler. The approach integrates Natural Language Processing (NLP) for unstructured log analysis, a Cost-Sensitive Support Vector Machine (SVM) for risk classification, and Discrete-Time Markov Chains (DTMC) for short-term reliability forecasting.

3.1 Data Preprocessing

Failure Mode and Effects Criticality Analysis (FMECA) data of the auto-coupler system is used further analysis. A multi-stage preprocessing pipeline was implemented using the following steps:

1. **Data Cleaning:** Null observations were removed to ensure data integrity.
2. **Feature Engineering:** A composite textual feature X_{text} constructed by concatenating the Component, Function, Failure Mode, and Manifestation fields. This ensures that the semantic context of each failure is captured.
3. **Vectorization:** The textual data was transformed into a numerical vector space using Term Frequency-Inverse Document Frequency (TF-IDF) to penalize frequently occurring but less informative words. Simultaneously, the numerical indices (Severity, Occurrence, Detection) were normalized using Standard Scaling to ensure parity with the high-dimensional text vectors.

3.2 Risk Stratification

A critical step in this approach is articulating the continuous Risk Priority Number (RPN) into categorical risk levels to be incorporated into the classification model. The RPN is calculated as the product of the Severity (B) × Occurrence (A) × Detection (E). The failure target variable is denoted as Y_{risk} defines the severity of the failure state. It is determined using a piecewise function based on standard maintenance thresholds:

$$\mathrm{Y_{risk}} = \begin{cases} \text{Low Risk} & \text{if RPN} < 40 \\ \text{Medium Risk} & \text{if } 40 \leq \text{RPN} < 90 \\ \text{High Risk} & \text{if RPN} \geq 90 \end{cases} \tag{1}$$

Where Low Risk represents minor faults requiring routine monitoring. Medium Risk: Represents moderate faults requiring scheduled maintenance. High Risk: Represents critical faults requiring immediate intervention or design modification.

3.3 Classification Model: Support Vector Machine (SVM)

In this paper, we further adopted a Support Vector Machine (SVM) approach with a linear kernel to classify failure risk. The SVM is a supervised learning algorithm that constructs a high-dimensional space to separate data classes. The objective is to find the optimal hyperplane defined by the equation:

$$f(x) = w^T x + b \tag{2}$$

Where, w is the weight vector orthogonal to the hyperplane, b is the bias term. The optimization problem seeks to maximize the margin between the classes while minimizing classification errors. This is formulated mathematically as minimizing the cost function $J(w, \varepsilon)$:

$$\min_{w,b}\left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\varepsilon_i\right) \quad (3)$$

Subject to the constraints, $y_i\,(w^T x_i + b) \geq 1 - \varepsilon_i\,,\ \varepsilon_i \geq 0$ (4)

where, $\|w\|^2$ the Euclidean norm of the weight vector (minimizing this maximizes the margin). C is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the misclassification of training and ε_i represent the slack variables that allow for misclassification of non-linearly separable datasets.

3.4 Classification performance metrics of SVM

$$\text{Precision} = \frac{\mathrm{TP_C}}{\mathrm{TP_C}+\mathrm{FP_C}} \quad (5)$$

$$\text{Recall} = \frac{\mathrm{TP_C}}{\mathrm{TP_C}+\mathrm{FN_c}} \quad (6)$$

$$\text{F1-score} = \frac{2\times\mathrm{Precision_c}\times\mathrm{Recall_c}}{\mathrm{Precision_c}+\mathrm{Recall_c}} \quad (7)$$

$$\text{Accuracy} = \frac{\sum_{\mathrm{c}}\mathrm{TP_c}}{\mathrm{N}} \quad (8)$$

Where, TP_C true positives, FP_C false positives, FN_c false negatives for class c and N is the total number of test instances.

3.5 Predictive Forecasting: Discrete-Time Markov Chain (DTMC)

To transition from static classification to temporal reliability analysis, the system's degradation behaviour is modelled as a stochastic process. We employ a Discrete-Time Markov Chain (DTMC), which relies on the Markov Property (memoryless property): the future risk state of a component depends solely on its current state, independent of the sequence of events that preceded it. Let X_t denote the discrete risk state at time t, where: $X_t \in \{\text{Low, Medium, High}\}$. A first-order Markov chain expression is given as:

$$P(X_{t+1} = j \mid X_t = i, X_{t-1\ldots\ldots}) = P(X_{t+1} = j \mid X_t = i) \quad (9)$$

3.5.1 State Space and Transition Matrix

The state space S is defined by the three risk levels identified in the previous stage $S = \{L\text{ow}, \text{Medium}, \text{High}\}$. The core of the model is the Transition Probability Matrix (TPM), denoted as P which governs the dynamics of the system:

$$P = \begin{bmatrix} P_{LL} & P_{LM} & P_{LH} \\ P_{ML} & P_{MM} & P_{MH} \\ P_{HL} & P_{HM} & P_{HH} \end{bmatrix} \quad (10)$$

The elements p_{ij} represent the conditional probability of transitioning from state i to state j over a single time interval Δt (one day). These probabilities are estimated empirically from the historical failure logs using Maximum Likelihood Estimation (MLE):

$$p_{ij} = \frac{N_{ij}}{\sum_{k \in S} N_{ik}} \tag{11}$$

Where, N_{ij} is the total count of observed transitions from state i to state j. The denominator $\sum N_{ik}$ represents the total residence time in state i. The constraint $\sum_{j \epsilon S} p_{ij} = 1$ ensures the matrix is stochastic.

3.5.2 Forecasting

To predict the risk profile over a future horizon (n = 7 days), we utilize the Chapman-Kolmogorov equations. Let $\pi_t = [\pi_{Low}, \pi_{Med}, \pi_{High}]$ be the row vector representing the probability distribution of the system's state at day t. The probability distribution for the next time step $(t + 1)$ is calculated via vector-matrix multiplication:

$$\pi_{t+1} = \pi_t \times P \tag{12}$$

By recursive application, the distribution for n days into the future is derived as:

$$\pi_{t+n} = \pi_t \times P^n \tag{13}$$

4 Analysis and results

4.1 SVM Risk Classification Performance

The proposed Cost-Sensitive Support Vector Machine (SVM) was evaluated on a test dataset of 142 failure modes to assess its ability to categorize faults into Low, Medium, and High-risk levels. The classification performance metrics are summarized in Table 1.

Risk Level	Precision	Recall	F1-Score	Support
High	0.96	1.00	0.98	25
Low	1.00	0.96	0.98	82
Medium	0.92	0.97	0.94	35
Accuracy			0.97	142

Table 1: SVM Classification Performance

Table 1 illustrates that the model attained an overall accuracy of 97%. Crucially, for a safety-critical system, the model attained a Recall of 1.00 for the High-Risk category. This demonstrates that all important failures in the test set (25 out of 25) were accurately identified, confirming that the cost-sensitive weighting technique effectively eradicated false negatives for dangerous faults. Figure 5 presents the confusion matrix for the 3-level risk classification model. The diagonal cells show correct classifications, while off-diagonal cells indicate misclassifications. The model correctly identified 79 Low, 34 Medium, and 25 High risk cases. Misclassifications were limited to 4 Low cases predicted as Medium and 1 Medium case predicted as High. No High-risk cases were predicted as Medium or Low, and no Low-risk cases were predicted as High, indicating strong separation between the extreme risk classes.

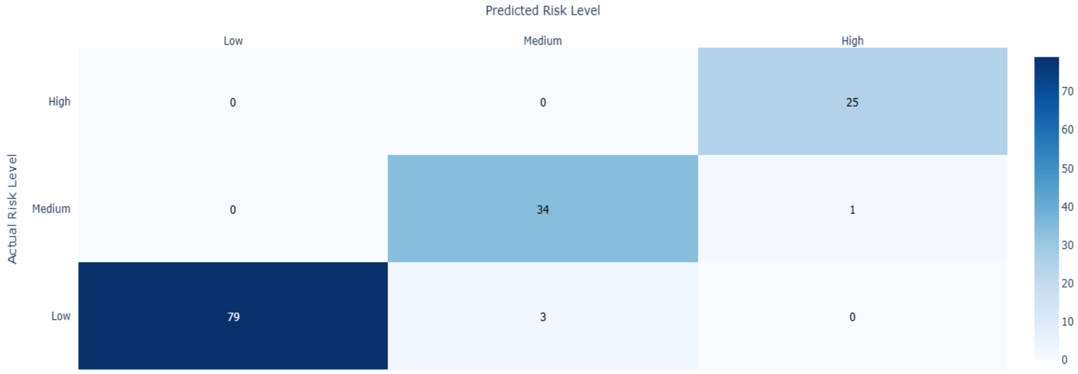

Figure 5: Confusion Matrix of SVM

Figure 6 shows the distribution of the predicted risk levels by failure type. Electrical failures account for the largest number of predicted cases across all risk categories (Low = 25, Medium = 15, High = 12), indicating this category contributes most to overall risk exposure. Mechanical failures are mainly predicted as Low (20) with fewer Medium (8) and High (4) cases. Pneumatic coupler failures show a mixed profile (Low = 12, Medium = 9, High = 3). Control-Protection-Monitoring failures are mostly predicted as Low (20), but include a noticeable High-risk portion (6). Buffer-Draft Gear has the smallest volume overall (Low = 2, Medium = 4, High = 1). Overall, higher predicted risk is concentrated in Electrical failures, while other failure types are dominated by Low-risk predictions.

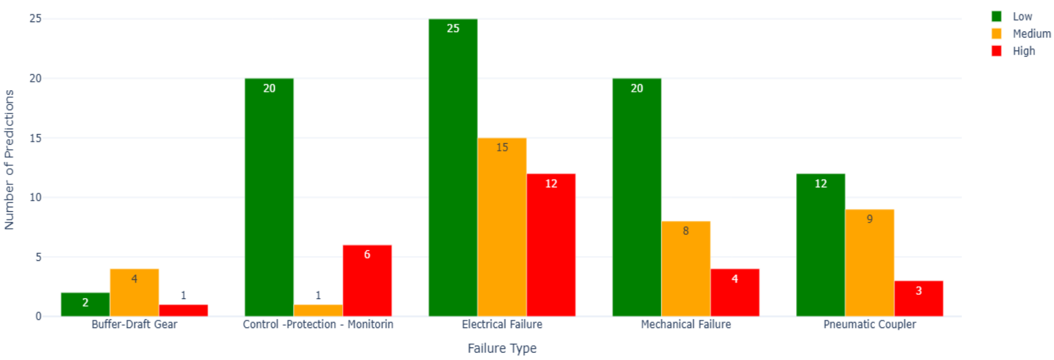

Figure 6: Predicted Risk Level Distribution of Failure Type

4.2 Markov Chain Risk Forecasting

Following the SVM-based risk stratification presented earlier in (Section 4.1), the forecasting stage is the Electrical failure category because it is the largest failure predicted High-risk cases and therefore represents the dominant short-term operational exposure. A Discrete-Time Markov Chain (DTMC) was trained using the observed transitions among the three discrete risk states {Low, Medium, High} within the Electrical subset. The trained DTMC was then used to forecast the Electrical risk-state distribution over a 7-day horizon.

4.2.1 Electrical Failure Forecast

Table 2 shows a clear rebalancing of Electrical risk over time. The forecast indicates that Electrical risk is initially dominated by High-risk percentage but progressively shifts toward Low-risk dominance as the process moves toward a near-stationary regime.

Day	Low (%)	Medium(%)	High(%)
1	10.26	23.08	66.67
2	21.79	26.74	51.48
3	30.90	26.52	42.57
4	37.58	25.75	36.67
5	42.36	25.06	32.58
6	45.76	24.55	29.69
7	48.18	24.17	27.65

Table 2: Forecast of Electrical risk-state probabilities (7 days)

The near-stabilisation after approximately Day-6 is consistent with a Markov process approaching its stationary distribution, and it is operationally meaningful: it identifies (i) a critical early window where Electrical risk is disproportionately high and maintenance resources should be front-loaded, and (ii) a persistent baseline regime where a non-zero high-risk probability remains and therefore requires continued monitoring rather than a single static risk label.

4.2.2 Component-level interpretation using FMEA factors and RPN

To convert the aggregate state probabilities into actionable maintenance insights, we decomposed the forecast into component-level contributions and associated FMEA factor values Ocurrence (A), Severity (B), and Detection (E) reported through the Risk Priority Number RPN = A × B × E. Table 3 reports the dominant contributors together with their forecast probability contribution and FMEA factor values. High-risk dominated window (front-load inspection and controls). On Day-1, Electrical risk is strongly High (66.67%). This High-risk mass is primarily driven by Electrical Head (10.00%, A=4.8, B=7.3, E=4.0, RPN=141.3) and Electrical Coupler Head (8.33%, A=3.8, B=8.0, E=4.2, RPN=129.6). In addition, several high-criticality contributors appear in the High-risk set, including DMOSA Coupling (3.33%, A=4.0, B=8.0, E=5.0, RPN=160) and the 5/2-Way Directional Valve (1.67%, A=6.0, B=9.0, E=5.0, RPN=270). Although some of these may have smaller percentage shares than the dominant "electrical head" items, their very high RPN values indicate a safety-operational consequence. Therefore, immediate actions should prioritise inspection and functional checks of electrical interfaces, harness-connectors, contact surfaces, and targeted tests for the most critical valve-actuation contributors.

Risk level	Components Probability with RPN
High	Electrical Head (10.00%; RPN(141.3), Electrical Coupler Head (8.33%; RPN(129.6), DMOSA Coupling (3.33%; RPN (160)), 5/2-Way Directional Valve (1.67%; RPN (270))
Medium	Inductive Sensor (1.7-2.0%; RPN (73.0)), Electrical Head Deployment (1.15–1.34%; RPN (69.0)), Coupling system (0.58–0.67%; RPN (84.0))
Low	Electrical Coupler Head (18.65%; RPN (19.2)), Electrical Head (5.18%; RPN (23.8)), Electrical Coupler Sensor (1.04%; RPN (24.0))

Table 3: Example Component with FMEA factors and RPN

As the forecast evolves over the 7-day horizon, a critical divergence emerges between failure percentage and failure severity. The model indicates a progressive shift toward Low-risk dominance, which reaches 48.18% by Day 7. This low-risk category is saturated with routine components such as the Electrical Coupler Head (18.65%), which, despite its high frequency, carries a low RPN of 19.2. However, the High-risk percentage stabilizes at a non-zero baseline of 27.65%. A granular analysis of this "high-risk" reveals a dangerous safety paradox: while the probability of these specific failures is low, their safety consequences are extreme. As detailed in the forecast breakdown, the 5/2-Way Directional Valve accounts for a negligible 0.69% of the forecast yet carries the highest severity in the dataset with an RPN of 270.0. Similarly, the DMOSA Coupling accounts for only 1.38% but maintains a critical RPN of 160.0, and other critical elements like the Coupler Operated Sensor contribute only 0.69% but hold an RPN of 140.0. This trend demonstrates that, over the 7-day horizon, a low failure rate does not equate to low passenger risk. If we designed our maintenance protocols based solely on the most likely failures, we would inevitably overlook the low-frequency components that pose the highest safety risks. This insight points to a necessary change in approach: we need to shift from managing based on frequency in the early stages to strictly prioritizing severity in the later horizon.

5 Conclusions and Contributions

This paper presented an integrated predictive maintenance framework for train autocouplers by combining FMEA-based Risk Priority Number (RPN) scoring, cost-sensitive SVM risk stratification, and Discrete-Time Markov Chains (DTMC) based short-term risk forecasting. Using 471 maintenance records, the proposed approach produced risk labels (Low-Medium-High) and provided a dynamic view of how risk evolves over time for operational decision-making. The application of a Cost-

Sensitive Support Vector Machine (SVM) to classify risk proved highly effective, achieving an overall accuracy of 97% and, most critically, a perfect recall of 1.00 for high-risk failures. This ensures that dangerous faults, such as those in locking mechanisms or electrical heads, are correctly prioritised, thereby validating the model's suitability for safety-critical systems where false negatives cannot be tolerated. Short-term forecasting using DTMC revealed distinct degradation profiles across the autocoupler subsystems, indicating electrical failures as the primary source of operational risk. Crucially, the analysis over the 7-day horizon revealed a significant divergence between failure percentage and failure severity. While the system statistically trends toward a Low-risk state (reaching 48.18% by Day 7) dominated by routine wear, a persistent "high-risk" of 27.65% remains. Detailed component analysis showed that specific elements within this tail, such as the 5/2-Way Directional Valve (RPN 270.0) and DMOSA Coupling (RPN 160.0), possessed extreme severity scores despite their negligible forecast percentage.

This research makes a significant contribution to railway maintenance strategies by empirically demonstrating that relying solely on failure percentage is insufficient for safety planning. If maintenance protocols were based solely on the most probable failures, these low-frequency but high-severity components would be overlooked. Therefore, the results support a risk-prioritised maintenance strategy: shifting from managing frequency in the early window (front-loading checks for common faults) to strictly prioritizing severity in the later horizon, enforcing mandatory checks for high-RPN components.

Acknowledgements

The authors would like to thank Siemens Mobility Limited Glasgow, UK for funding this research project.

References

[1] Sader, S., Husti, I. and Daróczi, M., 2020. Enhancing failure mode and effects analysis using auto machine learning: A case study of the agricultural machinery industry. *Processes*, *8*(2), p.224.

[2] Naranjo, J.E., Alban, J.S., Balseca, M.S., Bustamante Villagómez, D.F., Mancheno Falconi, M.G. and Garcia, M.V., 2025. Enhancing Institutional Sustainability Through Process Optimization: A Hybrid Approach Using FMEA and Machine Learning. *Sustainability*, *17*(4), p.1357.

[3] Peddi, S., Lanka, K. and Gopal, P.R.C., 2023. Modified FMEA using machine learning for food supply chain. *Materials Today: Proceedings*.

[4] Wu, X. and Wu, J., 2021. The risk priority number evaluation of FMEA analysis based on random uncertainty and fuzzy uncertainty. *Complexity*, *2021*(1), p.8817667.

[5] Yang, F., Tang, X., Gan, Y., Zhang, X., Li, J. and Han, X., 2021. Forecast of freight volume in Xi'an based on gray GM (1, 1) model and Markov forecasting model. *journal of mathematics*, *2021*(1), p.6686786.