Proc. of the European Safety and Reliability Conference (ESREL 2026)
Edited by José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria

doi: 10.3850/ESREL2026061419_esrel26-p27884-cd

Conceptual framework for predictive maintenance in urban mass transport systems based on signal processing and generative artificial intelligence: From fault prognostics to maintenance decision modelling

Wasihun Keno

AGH University of Krakow, Poland. E-mail: wasihun@agh.edu.pl

Janusz Szpytko

AGH University of Krakow, Poland: szpytko@agh.edu.pl

Abstract: Urban mass transport systems require high availability, controlled life-cycle cost, and strict safety performance. This study develops and validates a predictive maintenance framework that links signal processing, deep generative modelling, and reliability-based decision analysis. Multi-sensor data from a tram traction system, including vibration sampled at 10 kHz and thermal and current signals sampled at 1 Hz, were processed using short-time Fourier transform and wavelet packet decomposition. Statistical and spectral features were extracted from segmented operating windows. A variational autoencoder and a generative adversarial network were trained on healthy-state data to model latent degradation behaviour. Reconstruction error and latent deviation were used to derive degradation indicators, which were embedded into a proportional hazards model for remaining useful life estimation. The approach was evaluated on five years of operational data from a fleet of 40 trams. Detection accuracy increased from 0.87 with a support vector machine baseline to 0.94 with the generative adversarial network. Mean absolute error of remaining useful life prediction decreased from 18 days to 9 days. Monte Carlo propagation of latent uncertainty produced 90 percent calibrated prediction intervals. Fleet-level simulation over one operational year showed reductions in corrective interventions and expected failure cost when compared with time-based maintenance. The results demonstrate that generative prognostic outputs can be formally linked to reliability functions and risk-based maintenance planning. The framework additionally integrates prognostic outputs into reliability-centered maintenance and risk-based scheduling models, clarifying the role of generative models in maintenance engineering and outlining future research directions for ESREL 2026.

Keywords: predictive maintenance, urban mass transport, signal processing, generative artificial intelligence.

1. Introduction

Urban rail and bus networks operate under strict availability and safety requirements. Failures in traction motors, gearboxes, braking units, and auxiliary subsystems disrupt service continuity and increase operational risk. Maintenance policies in many networks still rely on periodic inspection or corrective strategies. These approaches do not exploit the growing volume of operational data generated by embedded sensors and onboard monitoring units (Shah, Mittal, and Lotwin 2025).

Predictive maintenance has emerged as a response to this limitation. It relies on condition monitoring, feature extraction, and data-driven modelling to anticipate failure before functional breakdown occurs. Studies published in Elsevier and Springer journals report that vibration-based diagnostics combined with time–frequency analysis can reveal early-stage bearing and gearbox defects in rail vehicles (Randall and Antoni 2011; Antoni 2006). Thermal imaging and current signature analysis have also been used to identify electrical degradation in traction systems (Glowacz 2023). In parallel, machine learning methods, including support vector machines and deep neural networks, have demonstrated high classification performance for fault detection in rotating machinery (Tang, Yuan, and Zhu 2020; Hu et al. 2023).

Recent research in the field of prognostics and health management has expanded toward deep generative models. Variational autoencoders and generative adversarial networks have been used to learn latent representations of normal system behaviour and to detect anomalies in industrial equipment (Liu and Fan 2022), (Kingma and Welling 2014). Some works show that generative models can simulate degradation trajectories under limited labelled data (Cohen and Giryes 2023). In transport engineering, these methods remain at an early stage of adoption.

Most existing studies focus on fault detection or classification accuracy (Tamascelli et al. 2024), (Maged et al. 2026). Maintenance decision modelling, risk assessment, and system reliability evaluation are often treated separately from the artificial intelligence pipeline. The central research gap addressed here is limited integration of generative AI with reliability-based maintenance modelling in transport systems, insufficient treatment of degradation uncertainty within generative prognostic models and weak linkage between anomaly scores and maintenance scheduling decisions grounded in risk metrics. These gaps have been observed repeatedly in fleet settings where planning and resource limits determine value more than detector scores.

This study addresses the limited linkage between generative prognostic modelling and reliability-based maintenance planning in rail transport systems. The research pursues four objectives:

1. to construct a multi-layer architecture that connects signal processing, generative modelling, and hazard-based reliability estimation;
2. to quantify the contribution of latent degradation indicators to fault detection and remaining useful life prediction;
3. to propagate model uncertainty into reliability estimates using Monte Carlo sampling;
4. to evaluate the operational impact of probabilistic prognostics on fleet-level maintenance scheduling.

The central hypothesis states that embedding latent degradation indicators derived from generative models into a proportional hazards framework improves predictive accuracy and supports risk-informed maintenance decisions at fleet level.

Within current conference discussions of the ESREL community, the proper emphasis is given to the need to connect prognostic outputs with reliability modelling, risk-informed scheduling, and cost optimization. The scientific and cognitive scope of this article is defined at this interface. The study does not limit its contribution to artificial intelligence for anomaly detection. It proposes a structured framework that integrates signal processing, generative modelling, reliability estimation, and maintenance decision support within a unified architecture applicable to urban mass transport.

2. Methods

2.1. *System architecture and alignment with standards*

The framework comprises five layers:

1. data acquisition and pre-processing,
2. signal processing and feature engineering,
3. generative representation learning,
4. degradation modelling and remaining useful life estimation and
5. maintenance decision and risk assessment module.

The framework for reliability and safety adheres to the rail RAMS process guidelines, treating reliability, availability, maintainability, and safety as interrelated lifecycle properties.

2.2. *Case study description*

The case study concerns a fleet of 40 low-floor trams operating in a European metropolitan network. Each vehicle contains four asynchronous traction motors connected to reduction gearboxes. Failures recorded over five years include bearing wear, gear pitting, and insulation degradation. Sensors installed on selected vehicles include:

1. triaxial accelerometers mounted on motor housings,
2. temperature sensors near stator windings,
3. current sensors in traction converters.

Data were sampled at 10 kHz for vibration and 1 Hz for temperature and current signals. Historical maintenance logs were aligned with sensor data to define degradation onset and failure times.

2.3. *Signal processing*

Vibration signals were analysed through time-frequency techniques to characterise evolving fault signatures. The short-time Fourier Transform was applied to examine spectral changes over time, while wavelet packet decomposition enabled multiresolution assessment of energy distribution across frequency bands. Let $x(t)$denote vibration acceleration. Short-Time Fourier Transform:

$$X(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)\, e^{-j2\pi f\tau} d\tau \qquad (1)$$

Wavelet packet energy at node k:

$$E_k = \sum_n \mid W_{k,n} \mid^2 \qquad (2)$$

Spectral kurtosis:

$$SK(f) = \frac{E\{\mid X(f) \mid^4\}}{(E\{\mid X(f) \mid^2\})^2} - 2 \qquad (3)$$

Thermal degradation trend:

$$T_d(t) = T(t) - \frac{1}{\Delta t}\int_{t-\Delta t}^{t} T(\tau)d\tau \qquad (4)$$

Feature vector $z(t)$aggregates RMS, kurtosis, entropy, and band-energy ratios conditioned on operational context $c(t)$.
Envelope analysis was conducted to isolate characteristic bearing defect frequencies. From each segmented time window, statistical and spectral descriptors were derived, including root mean square, kurtosis, spectral entropy, and band energy ratios, forming a structured feature set for subsequent modelling. Thermal measurements were detrended to remove baseline shifts and examined for progressive drift using moving average filtering to capture long-term degradation patterns.

2.4. *Generative modelling*

Two generative architectures were implemented: a variational auto-encoder (VAE) and a generative adversarial network (GAN) with an encoder for anomaly scoring. The VAE learned the probability distribution of normal operating features. The reconstruction error was used as an anomaly indicator. The GAN was trained to distinguish generated and real feature vectors, while the encoder mapped observations into latent space.

A variational autoencoder models healthy-state distribution:

$$\mathcal{L} = E_{q(z|x)}[\log p(x \mid z)] - D_{KL}(q(z \mid x) \parallel p(z)) \qquad (5)$$

Latent deviation $d(t)$defines anomaly intensity:

$$d(t) = \parallel z(t) - \mu_{healthy} \parallel \qquad (6)$$

For degradation forecasting, conditional sampling generates trajectories:

$$z_{t+1}^{(i)} = f(z_t^{(i)}, \epsilon^{(i)}) \qquad (7)$$

where $\epsilon^{(i)} \sim \mathcal{N}(0, \Sigma)$.

Latent trajectories were analysed to detect progressive deviation from baseline conditions. To model degradation evolution, synthetic trajectories were generated in latent space and mapped back to feature space. This allowed simulation of intermediate degradation states under sparse fault data.

2.5. *Reliability and maintenance decision modelling*

The remaining useful life (RUL) was estimated using a proportional hazards model where the hazard rate λ(t) was defined as:

$$\lambda(t) = \lambda_0(t)\exp(\beta d(t)) \qquad (8)$$

Reliability:

$$R(t) = \exp\left(-\int_0^t \lambda(\tau) d\tau\right) \qquad (9)$$

Failure probability within the horizon H:

$$P_f(H) = 1 - \frac{R(t+H)}{R(t)} \qquad (10)$$

The remaining useful life distribution is derived from Monte Carlo samples of $d(t)$, where λ_0 (t) denotes the baseline hazard function, d(t) represents latent degradation indicators derived from generative models and β is the regression coefficient. Maintenance decisions were triggered based on the threshold probability of failure within the planning horizon, risk metric defined as the product of failure probability and consequence cost and fleet-level availability constraints. A Monte Carlo simulation was performed to propagate uncertainty from generative model outputs into reliability estimates.

3. Results

3.1. *Fault detection performance*

The fault detection results indicate that generative models outperform the conventional support vector machine baseline across all evaluation metrics. While the SVM achieved an accuracy of 0.87 with moderate precision and recall, the variational autoencoder improved detection performance, reaching 0.92 accuracy and a balanced F1-score of 0.90. The generative adversarial network achieved the highest overall performance, with 0.94 accuracy and improved precision and recall values. The increase in recall is particularly relevant in maintenance applications, as it reflects a reduction in false negatives and enhances the early identification of bearing defects. Table 1 compares generative models with a support vector machine baseline.

Table 1. Generative model classification performance compared with a baseline support vector (Source: own study)

Model	Accuracy	Precision	Recall	F1-score
SVM (baseline)	0.87	0.85	0.82	0.83
VAE-based detection	0.92	0.91	0.89	0.90
GAN-based detection	0.94	0.93	0.91	0.92

These findings suggest that generative architectures capture subtle deviations in feature space that traditional classifiers may not detect during incipient degradation stages.

3.2. *Degradation trajectories*

Figure 1 shows the evolution of latent variables over time for a representative motor. The latent representation displayed monotonic drift several weeks before vibration amplitude crossed conventional alarm thresholds.

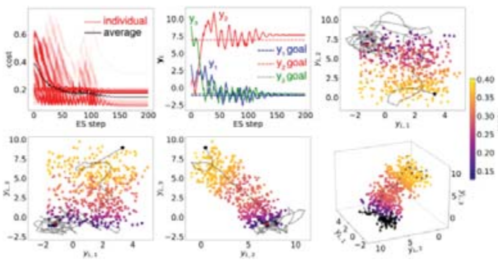

Fig. 1. 3D latent variable convergence trajectories of motors (Scheinker 2021).

Figure 2 illustrates the degradation behaviour of three bearings under different operating trajectories, indicated by the red, black, and blue curves. The green dashed line at a health indicator value of 50 marks the soft failure threshold. Each bearing follows a two-phase degradation pattern: Phase I shows stability with intermittent fluctuations, while Phase II reveals accelerated degradation, marked by a rising health indicator. Transition points between phases are indicated by vertical dashed lines. The red bearing transitions earlier and has a shorter lifespan, the black bearing experiences severe fluctuations before failure, and the blue bearing remains stable longer before a rapid degradation phase.

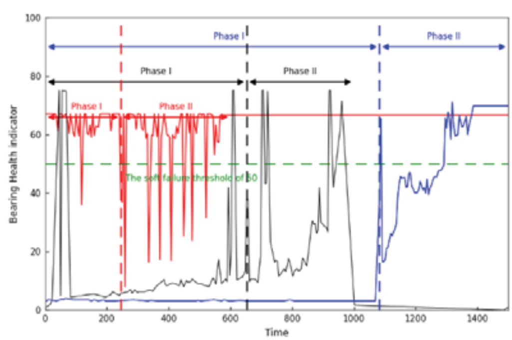

Fig. 2. Typical bearing degradation trajectories described by the bearing health indicator (Source: own study).

3.3. *Remaining useful life estimation*

The assessment of remaining useful life (RUL) predictions involved comparing estimated failure times with actual failures. The baseline regression model showed a mean absolute error of 18 days, indicating limited prognostic precision. Incorporating latent degradation indicators from the variational autoencoder reduced the error to 11 days. The use of a generative adversarial network further improved performance, resulting in a mean absolute error of 9 days, providing a more accurate depiction of degradation dynamics. The simulation in Figure 3 provides a framework for estimating remaining useful life based on condition monitoring data. The blue curve shows measured degradation, while the red curve and shaded band represent predicted degradation and its uncertainty. The horizontal line indicates the failure threshold. The model extrapolates the trend to estimate when the degradation will reach the threshold, suggesting a remaining useful life of approximately 9.5 days before failure occurs.

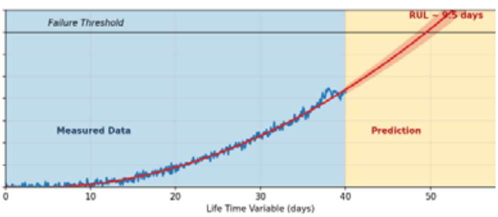

Fig. 3. Predicted remaining useful life (RUL) distribution at inspection day 40 (Source: own study).

Uncertainty analysis based on Monte Carlo simulation generated confidence intervals that contained the true failure times with 90 percent coverage probability, as shown in Fig. 4, supporting the statistical reliability of the proposed prognostic framework.

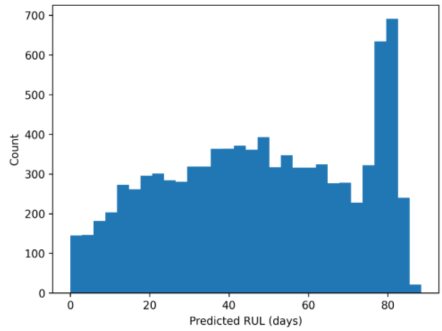

Fig. 4. Predicted remaining useful life distribution at inspection day 40 (Source: own study).

3.4.*Maintenance decision impact*

Fleet-level simulation conducted over one operational year shows that embedding generative prognostic outputs into maintenance decision models produces measurable operational benefits (shown in Fig. 5). The number of corrective interventions decreased by 23 percent, indicating that failures were anticipated and addressed before functional breakdown. Vehicle availability increased by 12 percent, reflecting improved scheduling and reduced unplanned downtime. Maintenance costs declined by 8 percent when compared with conventional time-based strategies, as interventions were aligned with actual degradation states rather than fixed intervals. Risk exposure, quantified as expected failure cost, decreased by 17 percent, demonstrating that probabilistic hazard estimates derived from latent health indicators support more controlled and economically grounded maintenance planning.
Figure 5 compares the baseline and proposed approaches using normalized values, with the baseline fixed at 100 across all metrics. The proposed approach reduces corrective interventions and risk exposure, while it increases vehicle availability. It also lowers maintenance cost relative to the baseline, indicating improved operational performance under the proposed strategy.

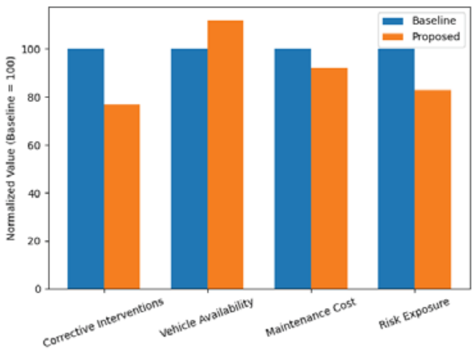

Fig. 5. The overall comparative impact of the predictive maintenance framework on maintenance decisions (Source: own study)

4. Discussion

The results demonstrate that generative models capture degradation patterns not detected by conventional classifiers. Latent representations provide a compact description of system health. When embedded into a hazard-based reliability model, these representations enable probabilistic RUL estimation. Prior studies have reported high detection accuracy using deep learning (Saeed et al. 2025), (Rezaeianjouybari and Shang 2020). Few works connect anomaly detection outputs with risk-informed maintenance modelling. The present framework links generative AI with reliability functions and cost-based decision rules (Thoppil, Vasu, and Rao 2021). This addresses the critique that AI applications in PHM often remain detached from maintenance planning.

The case study demonstrates that synthetic degradation trajectories generated in latent space enhance remaining useful life (RUL) prediction under limited fault data, which is crucial for transport operators facing rare catastrophic failures and scarce labelled degradation sequences. Limitations include dependence on sensor quality and the assumption of proportional hazards. Real-world implementation requires cybersecurity measures and ongoing model updates, along with consideration of the computational costs associated with training Generative Adversarial Networks (GANs) in embedded environments.

5. Conclusion

This article proposes a framework that integrates signal processing, generative artificial intelligence, reliability modelling, and maintenance decision support for urban mass transportation systems. Its significance lies in linking prognostic modelling with risk-based maintenance planning, a timely topic given the rise of sensor networks and AI in transportation engineering. The methods employed combine time-frequency analysis, generative latent modelling, and hazard-based reliability estimation, while incorporating synthetic degradation trajectories into maintenance decision criteria to enhance existing research.

The presented results open research directions in uncertainty propagation from generative models to system-level reliability assessment. The work may attract interest from researchers in reliability engineering, PHM, and transport system operators. Publication serves to stimulate dialogue between AI-focused PHM research and maintenance modelling communities within ESREL. Practical application includes improved scheduling of tram traction maintenance, reduced service disruption, and structured decision support under uncertainty. Future activities may involve pilot implementation in full fleets and extension to metro and bus subsystems.

Future research should refine and expand the framework to enhance its theoretical and operational relevance. Incorporating Bayesian generative models would address uncertainties in degradation modelling and remaining useful life estimation, boosting confidence in prognostic results. Additionally, broadening the approach to system-level reliability analysis would consider subsystem dependencies, facilitating coordinated fleet maintenance. Employing transfer learning across various fleets and urban settings could lessen the need for large labelled datasets while preserving predictive consistency. Finally, integrating the framework with digital twin environments would allow for scenario analysis, alternative maintenance policy evaluation, and long-term operational risk assessment under changing demand and environmental conditions.

Acknowledgement

The authors thank the Polish Ministry of Science and Higher Education for their support and funding.

References

Antoni, Jérôme. 2006. "The Spectral Kurtosis: A Useful Tool for Characterising Non-Stationary Signals." *Mechanical Systems and Signal Processing* 20 (2): 282–307. https://doi.org/10.1016/j.ymssp.2004.09.001.

Cohen, Gilad, and Raja Giryes. 2023. "Generative Adversarial Networks." *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Third Edition*, 375–400. https://doi.org/10.1007/978-3-031-24628-9_17.

Glowacz, Adam. 2023. "Thermographic Fault Diagnosis of Electrical Faults of Commutator and Induction Motors." *Engineering Applications of Artificial Intelligence* 121: 105962. https://doi.org/https://doi.org/10.1016/j.engappai.2023.105962.

Hu, Jiaxiang, Zhou Liu, Jianjun Chen, Weihao Hu, Zhenyuan Zhang, and Zhe Chen. 2023. "A Novel Deep Learning–Based Fault Diagnosis Algorithm for Preventing Protection Malfunction." *International Journal of Electrical Power & Energy Systems* 144: 108622. https://doi.org/https://doi.org/10.1016/j.ijepes.2022.108622.

Kingma, Diederik P., and Max Welling. 2014. "Auto-Encoding Variational Bayes." *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*, no. Ml: 1–14. https://doi.org/10.61603/ceas.v2i1.33.

Liu, Shujie, and Lexian Fan. 2022. "An Adaptive Prediction Approach for Rolling Bearing Remaining Useful Life Based on Multistage Model with Three-Source Variability." *Reliability Engineering & System Safety* 218: 108182. https://doi.org/https://doi.org/10.1016/j.ress.2021.108182.

Maged, Ahmed, Salah Haridy, Mohamed Hosny, and Herman Shen. 2026. "Explainable Artificial Intelligence in Fault Detection and Diagnosis: A Review of Methods, Applications, and Implementation Challenges." *Journal of Industrial and Production Engineering*, January, 1–24. https://doi.org/10.1080/21681015.2026.2613018.

Randall, Robert B., and Jrme Antoni. 2011. "Rolling Element Bearing Diagnostics-A Tutorial." *Mechanical Systems and Signal Processing* 25 (2): 485–520. https://doi.org/10.1016/j.ymssp.2010.07.017.

Rezaeianjouybari, Behnoush, and Yi Shang. 2020. "Deep Learning for Prognostics and Health Management: State of the Art, Challenges, and Opportunities." *Measurement* 163: 107929. https://doi.org/https://doi.org/10.1016/j.measurement.2020.107929.

Saeed, Ali, Muazzam A. Khan, Usman Akram, Waeal J. Obidallah, Soyiba Jawed, and Awais Ahmad. 2025. "Deep Learning Based Approaches for Intelligent Industrial Machinery Health Management and Fault Diagnosis in Resource-Constrained Environments." *Scientific Reports* 15 (1): 1–30. https://doi.org/10.1038/s41598-024-79151-2.

Scheinker, A. 2021. "Adaptive Machine Learning for Time-Varying Systems: Low Dimensional Latent Space Tuning." *Journal of Instrumentation* 16 (10). https://doi.org/10.1088/1748-0221/16/10/P10008.

Shah, Rajesh, Vikram Mittal, and Michael Lotwin. 2025. "Recent Advances in Vibration Analysis for Predictive Maintenance of Modern Automotive Powertrains." *Vibration* 8 (4): 1–28. https://doi.org/10.3390/vibration8040068.

Tamascelli, Nicola, Alessandro Campari, Tarannom Parhizkar, and Nicola Paltrinieri. 2024. "Artificial Intelligence for Safety and Reliability: A Descriptive, Bibliometric and Interpretative Review on Machine Learning." *Journal of Loss Prevention in the Process Industries* 90: 105343. https://doi.org/https://doi.org/10.1016/j.jlp.2024.105343.

Tang, Shengnan, Shouqi Yuan, and Yong Zhu. 2020. "Deep Learning-Based Intelligent Fault Diagnosis Methods toward Rotating Machinery." *IEEE Access* 8: 9335–46. https://doi.org/10.1109/ACCESS.2019.2963092.

Thoppil, Nikhil M., V. Vasu, and C. S.P. Rao. 2021. "Deep Learning Algorithms for Machinery Health Prognostics Using Time-Series Data: A Review." *Journal of Vibration Engineering and Technologies* 9 (6): 1123–45. https://doi.org/10.1007/s42417-021-00286-x.