38th International Electric Vehicle Symposium and Exhibition (EVS38) Göteborg, Sweden, June 15-18, 2025

Improving Data-Driven Predictive Maintenance Strategies Through Synthetic Data: A TimeGAN Approach

Batuhan Cinar[1], Serhat Ege Inanc[2], Muhammed Enis Sen[2], Ali Serdar Atalay[2], Mohammad Ghazali[1], Asaf Mustafa Sofu[3], Yunus Karamavus[3], Ahu Ece Hartavi[1,4],

[1]*University of Surrey, Stag Hill, University Campus, Guildford, UK*

[2]*AI4SEC OU, Sepapaja 6, Talinn, Estonia*

[3]*ORTEM, Bilisim Vadisi, Muallimköy Mah. Deniz Cad. No: 143/6 Gebze/KOCAELİ*

[4]*Corresponding Author (a.hartavikarci@surrey.ac.uk)*

Executive Summary

Predictive maintenance using data-driven methods is a leading strategy in smart manufacturing, utilizing industrial big data. Since predictive maintenance research is still largely experimental, existing studies rely on limited open-access datasets, which often fail to represent real-world scenarios, particularly in monitoring battery state-of-health in zero-emission Heavy-Duty Vehicles. The complex behaviour of capacity fade in in zero-emission Heavy-Duty Vehicles batteries makes predicting their remaining useful life challenging. This study presents a novel approach that enhances predictive maintenance battery remaining useful life prediction by combining synthetic data from a custom TimeGAN model with real-world data. Through this the study has improved error rates of the predictive maintenance algorithm 10%.

Keywords: Heavy Duty Electric Vehicles & Busses, AI – Artificial Intelligence for EVs, Energy Storage Systems, Digital Twin Design Tools, Batteries

1 Introduction

The emergence of smart manufacturing and industry has accelerated the adoption of data-driven predictive maintenance (PdM) systems, offering significant potential to reduce unplanned downtime and optimize asset lifecycle management. With the increasing electrification of heavy-duty transportation, zero-emission heavy-duty vehicles (z-HDVs) have become central to sustainability goals. These vehicles rely heavily on lithium-ion battery systems, whose performance degrades nonlinearly over time due to complex electrochemical interactions. As such, the effective monitoring of battery state-of-health (SoH) and accurate estimation of remaining useful life (RUL) are vital to ensuring operational reliability, safety, and cost-efficiency.

PdM has become a cornerstone of intelligent manufacturing and smart fleet management, particularly for z-HDVs, where battery health is crucial for operational reliability. PdM strategies harness sensor-rich environments and machine learning algorithms to predict equipment failure and optimize maintenance schedules. Zhang et al. [1] offer a comprehensive survey on data-driven PdM methods, highlighting the growing role of artificial intelligence (AI) and big data in capturing the complex dynamics of industrial systems. The global PdM market reflects this momentum, expected to grow from $10.6 billion in 2024 to $47.8 billion by 2029 [2], driven by the electrification of transportation and the demand for smarter, more sustainable maintenance strategies. In the context of electric vehicles, Konkimalla [3] demonstrated how AI-based PdM—using prognostics, deep learning, and real-world sensor integration—can significantly reduce costs and enhance fleet reliability.

However, building robust PdM systems for z-HDV batteries remains a formidable challenge, primarily due to the limited availability of high-quality, diverse, and labeled real-world datasets. Current machine learning (ML) approaches for battery RUL prediction often depend on datasets derived from laboratory experiments or small-scale deployments, which may not capture the full range of operating conditions, degradation behaviors, and failure modes encountered in the field. This data scarcity, combined with class imbalance and the rarity of critical degradation events, can lead to overfitting, biased models, and unreliable performance in real-world applications. As Voronov et al. [4] argue, models trained on sparse and imbalanced datasets can struggle to generalize, especially in the presence of rare or extreme degradation scenarios.

Traditional PdM approaches, such as Support Vector Machines (SVMs) [5] and neural networks [6], have been used to extract degradation features from time-series data like voltage, current, and temperature. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have proven effective in modeling sequential dependencies in battery health data [7]. However, recent advancements in deep learning have also introduced hybrid architectures that combine convolutional and recurrent layers for better spatial-temporal modeling. For example, a CNN-LSTM model trained on synthetic data from 3D multiphysics battery simulations has been shown to accurately predict voltage discharge curves and internal temperature hotspots, offering real-time value for thermal safety and energy management in EV batteries [8].

To mitigate data scarcity and enhance model robustness, synthetic data generation has emerged as a promising approach. Generative models, especially Generative Adversarial Networks (GANs), have shown promise in augmenting training datasets by producing realistic synthetic samples that preserve the statistical and temporal properties of the original data. Time-series Generative Adversarial Networks (TimeGAN), introduced by Yoon et al. [9], merge supervised learning with adversarial training to retain both temporal patterns and data diversity. TimeGAN has also been successfully used in financial forecasting applications—for instance, to predict ATM cash needs using synthetic time-series data, demonstrating strong generalizability and improved accuracy compared to models trained on limited real-world data [10].

In PdM-specific contexts, recent innovations have explored various generative approaches. Diffusion-based generative models have been proposed to address highly imbalanced datasets, generating synthetic fault data that enhances anomaly detection under data-sparse conditions [11]. Safdari et al. [12] introduced a simulation-based benchmark for synthetic battery degradation data, including both time-series and snapshot features, offering a realistic and publicly available dataset for survival modeling. Similarly, Mihale-Wilson et al. [13] combined clustering and generative AI to inject synthetic anomalies into PdM datasets for smart appliances, facilitating model training in anomaly-scarce environments. Meanwhile, Tondro [14] presented TimeFairGAN, a fairness-aware GAN extension that integrates demographic parity constraints, enabling ethically balanced synthetic time-series generation for PdM use cases.

Despite their potential, the application of TimeGANs to predictive maintenance—particularly in the context of battery SoH and RUL forecasting in z-HDVs—remains largely unexplored.

Recent research has begun to bridge this gap. For example, a study by Zhang et al. [15] investigates how deep generative models can simulate battery aging behavior under different operating conditions, enhancing the flexibility of PdM models in real-world applications. Meanwhile, Esteban et al. [16] provide a broader analysis of generative time-series models, proposing architectures capable of capturing complex temporal dependencies and multi-modal distributions—essential features for accurately modeling battery degradation in z-HDVs.

In this study, a novel framework is proposed that integrates synthetic data generated by a customized TimeGAN model into a deep learning-based PdM pipeline for z-HDV batteries. By augmenting real-world datasets with synthetic data, we aim to enhance the performance, robustness, and generalizability of battery RUL prediction models. The TimeGAN architecture is carefully tailored to capture long-term dependencies, rare degradation events, and cycle-specific effects, making it well-suited for modeling the nonlinear degradation trajectories typical of z-HDV battery systems.

Through comprehensive experimentation and evaluation, this research seeks to answer critical questions around the feasibility and efficacy of synthetic data in improving PdM model outcomes. We test our hybrid approach using both real and combined datasets within a convolutional-recurrent neural network framework, comparing model performance using standard metrics such as accuracy, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). By addressing the data limitations that hinder current PdM systems, our proposed methodology contributes a scalable solution to enhancing battery maintenance strategies in zero-emission transport, ultimately supporting the broader transition to cleaner, more efficient heavy-duty mobility.

2 Methodology

The following section details down the used dataset for the purposes of this study, a PdM algorithm to benchmark the performance of the generated syntehtic data and finally the TimeGAN implementation of the study.

2.1 Dataset Description

The used dataset is provided by [5] which is real world scenario field data captured over a span of 29 months for 20 commercially available electric vehicles. Data is gathered from charger units, utilizing the CAN network of the vehicle when plugged. Figure 1 shows the state-of-charge, voltage, temperature and current curves for a single vehicle.

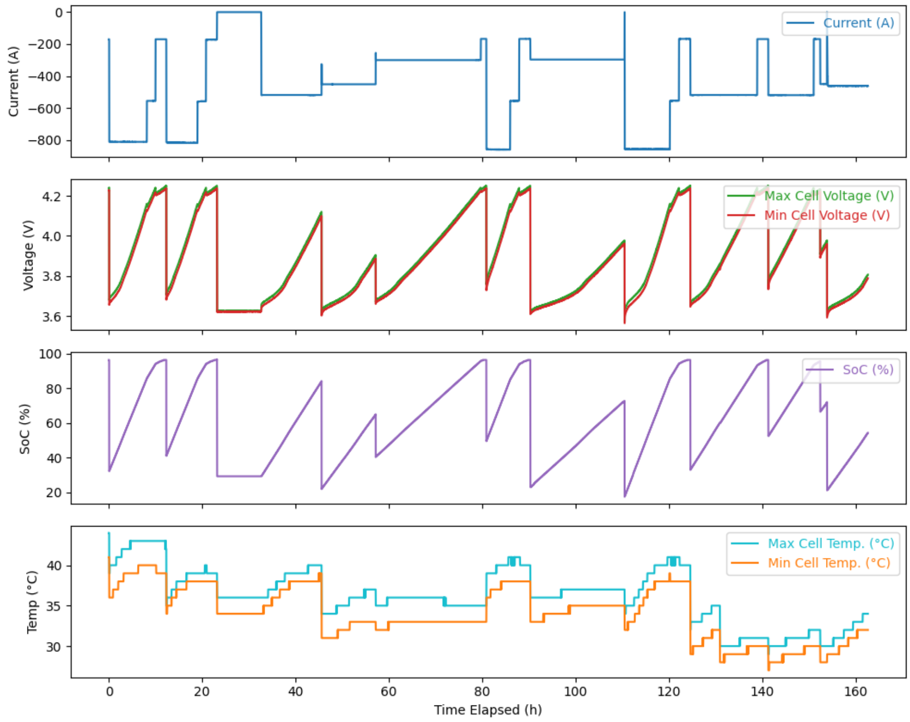


Figure 1: Dataset features for one selected EV truck over time [7]

For the prediction of SoH, labelled capacity available at time t, equation (1), is considered [7].

$$C_a = -\frac{\int_{t_0}^{t} I(t)dt}{SoC_{t_0} - SoC_{\mathrm{t}}} \tag{1}$$

Where C_a is the capacity for single vehicle, $t_{\#}$ is the time period, $I(t)$ is the current at that time and $SoC_{t\#}$ represent the state of charge values as the bounds of the integral

The dataset consists of twenty different CSV files, each representing a single vehicle. All the files contain the same set of features, which are categorized into three main groups: time-dependent parameters, cell-specific parameters, and temperature measurements. The time-dependent parameters include Time, State-of-Charge (SoC) (%), battery pack voltage (V), and charge current (I). The cell-specific parameters consist of the maximum and minimum cell voltages (V). Lastly, the temperature measurements capture the maximum and minimum pack inside temperatures, recorded in degrees Celsius (°C).

These csv files are transferred to Pandas dataframe objects, each having its own dataframe. This approach has been chosen to later on implement Incremental Learning. A sample dataframe from a vehicle is given in Figure 2.

	record_time	soc	pack_voltage (V)	charge_current (A)	max_cell_voltage (V)	min_cell_voltage (V)	max_temperature (°C)	min_temperature (°C)	available_capacity (Ah)
0	2019-07-25 16:28:42	99.6	381.1	-14.80002	4.245	4.223	41	40	135.34
1	2019-07-25 16:28:50	99.6	381.1	-14.80002	4.246	4.223	41	40	135.34
2	2019-07-25 16:28:58	99.6	381.1	-14.80002	4.245	4.222	41	40	135.34
3	2019-07-25 16:29:06	99.6	381.1	-14.80002	4.246	4.223	41	40	135.34
4	2019-07-25 16:29:14	99.6	381.1	-14.80002	4.246	4.224	41	40	135.34
...	...	...	...	...	...	...	...	...	...
647327	2021-04-19 06:48:00	97.2	379.6	-43.39999	4.232	4.200	28	26	117.04
647328	2021-04-19 06:48:10	97.6	379.6	-43.50000	4.234	4.201	28	26	117.15
647329	2021-04-19 06:48:20	97.6	379.7	-43.39999	4.235	4.201	28	26	117.27
647330	2021-04-19 06:48:30	97.6	379.7	-43.50000	4.236	4.203	28	26	117.39
647331	2021-04-19 06:48:40	97.6	379.9	-43.50000	4.237	4.203	28	26	117.49

Figure 2: Single vehicle Pandas data frame object

2.2 Preprocessing

The given dataset undergoes several preprocessing operations to enable more effective analysis and to generate a State of Health (SoH) feature, which is not present in the raw data. The preprocessing begins with data cleaning, as the initial CSV files were not isolated to individual vehicles, leading to data leakage between files. To address this, the data was separated into distinct DataFrames for each vehicle. Additionally, records with missing or anomalous values were manually removed to maintain data integrity. Following this, feature engineering was performed to create a peak capacity feature, which serves as a target variable for prediction. This involved applying peak detection algorithms within specific State of Charge (SoC) ranges, recognizing that not all charging events extend from 0% to 100%. Figure 3 illustrates the trend of this engineered feature over time for Vehicle #2.

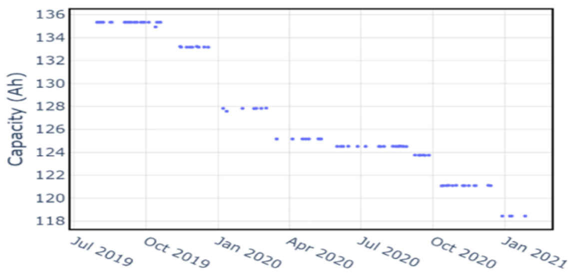


Figure 3: Engineered battery capacity target feature for vehicle #2

To further refine the dataset, interpolation and smoothing techniques were applied. The engineered target variable, along with some predictive features, contained gaps that were addressed using linear interpolation, which estimates missing values based on the existing trends in the data. Since battery degradation is a gradual physical process and abrupt changes are not expected, a Savitzky-Golay filter was subsequently applied to smooth the data. This filtering technique helps reduce noise and mitigates sudden variations in the gradient. Figure 4 illustrates the effect of interpolation and smoothing on the target variable for Vehicle #2.

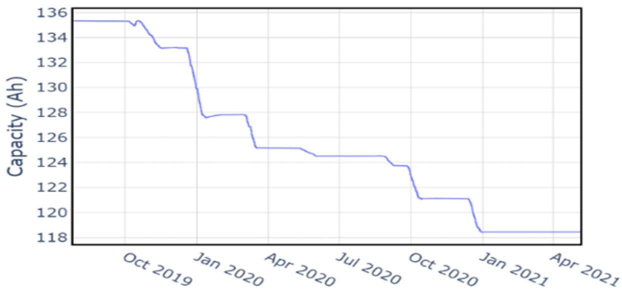


Figure 4: Interpolated and smoothed battery capacity target feature for vehicle #2

All numerical features are normalized to have a mean of 0 and a standard deviation of 1. This normalization step is essential for the neural network model which is sensitive to feature scaling

2.3 The Predictive Maintenance Algorithm

In order to address the challenges of limited and imbalanced real-world datasets in battery RUL prediction for z-HDV, this paper proposes a PdM framework that integrates synthetic data from a TimeGAN model into a deep learning pipeline. The core PdM algorithm is based on a hybrid architecture combining Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for capturing temporal dependencies in time-series sensor data.

The proposed PdM algorithm is visualized in Figure 5: A 1D CNN with multiple convolutional and max-pooling layers is used to extract high-level representations from each multivariate time window. This reduces dimensionality while preserving local temporal patterns, such as abrupt voltage drops or cyclical temperature changes. The CNN output is fed into a stack of LSTM layers designed to learn long-term dependencies and degradation trends across time. LSTMs are chosen due to their ability to capture vanishing or exploding gradient issues common in standard RNNs, and to effectively model battery aging processes over operational cycles Finally for RUL regression, a fully connected layer outputs a continuous value representing estimated remaining life.

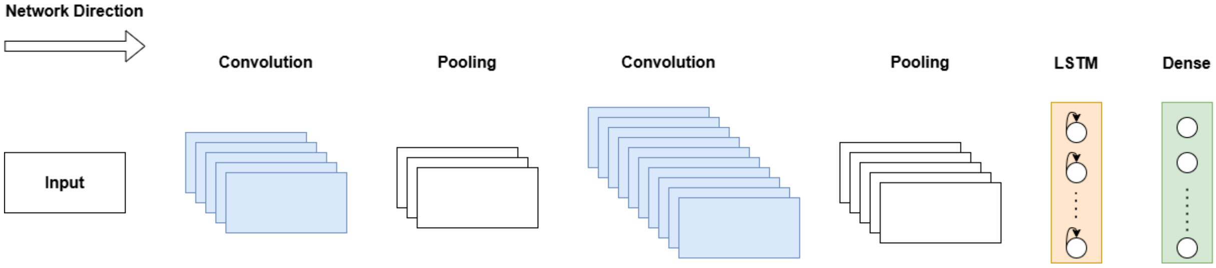


Figure 5: PdM Architecture

2.3.1 Training Strategy

The model is trained under two experimental settings to evaluate its performance under different data conditions. In the first setting, the model is trained using only real-world data. In the second setting, a combined dataset comprising both real data and synthetic data generated by TimeGAN is used. This comparison allows for assessing the impact of incorporating synthetic data on the model's predictive capabilities.

RMSE is employed for the regression task as the loss of the training network. An Adam optimizer [] is applied with an initial learning rate of 0.001. Early stopping is implemented based on validation loss to prevent overfitting.

Training was conducted on a 12th Gen Intel® Core™ i7-12700K CPU, reflecting the resource-constrained settings often found in practical industrial deployment environments. Model performance is assessed using accuracy, RMSE and MAE. These metrics were computed across both real-only and hybrid datasets to quantify the impact of synthetic data on predictive performance.

2.4 TimeGAN for Synthetic Time-Series Data Generation

To overcome limitations posed by scarce and imbalanced real-world battery datasets, this study integrates a Time-series Generative Adversarial Network (TimeGAN) to synthesize realistic battery degradation data. Unlike traditional GANs, which are typically unsuitable for sequential data, TimeGAN uniquely combines generative modeling with recurrent neural networks to capture both temporal dynamics and data distributions, making it particularly well-suited for time-series applications such as battery RUL monitoring.

2.4.1 TimeGAN Architecture

TimeGAN consists of four key components, as illustrated in Figure 6, each playing a critical role in generating realistic synthetic time-series data. The **Embedding Network** encodes high-dimensional input time-series into a lower-dimensional latent space while preserving temporal information, allowing both the generator and discriminator to operate in a compressed yet informative space. The **Recovery Network** then decodes this latent representation back into the original data format and is trained jointly with the embedding network to ensure accurate reconstruction and capture of temporal patterns. The **Generator** creates new time-series sequences from random noise within the latent space, learning to replicate the real data distribution and key temporal behaviors observed in battery degradation, such as gradual capacity loss or sudden voltage drops. Finally, the **Discriminator** attempts to distinguish between real and synthetic sequences in both latent and original feature spaces, thereby pushing the generator to produce more realistic and high-quality data through adversarial training.

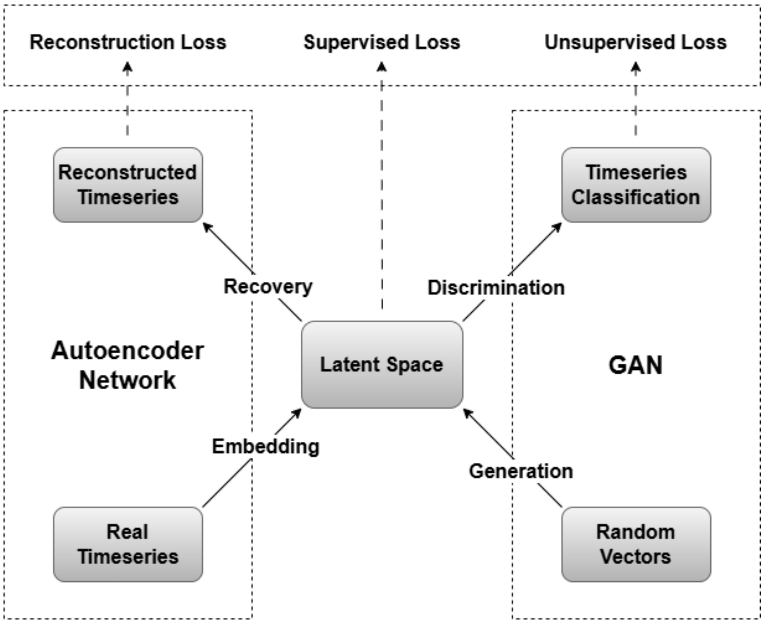


Figure 6: TimeGAN Architecture

2.4.2 Training Objective of the TimeGAN

The TimeGAN model is trained using a composite loss function that integrates multiple objectives to enhance the quality and realism of the generated time-series data. This includes a **supervised loss** that aligns temporal dynamics within the latent space, a **reconstruction loss** that ensures accurate recovery of the original time-series data, and an **adversarial loss** that encourages the generator to produce synthetic sequences indistinguishable from real data by fooling the discriminator.

This multi-component objective ensures that the generated data maintains sequence continuity, statistical similarity, and temporal coherence, addressing key challenges in synthetic data generation for PdM applications. These three losses combined generate the main losses of the two components, the autoencoder loss and the discriminator and generator losses which this study monitors. The reconstruction loss, forming the core of the autoencoder loss, is unsupervised and ensures the embedder and recovery networks can accurately reconstruct the original input from its latent representation-preserving the data's feature-level details. It is defined as:

$$L_{Recon} = \mathbb{E}\left[\left|\left|x - x_{hat}\right|\right|_2^2\right], \qquad \text{where } x_{hat} = R\big(E(x)\big) \tag{2}$$

Where, x is the original timeseries data, $E(.)$ is embedding function application, $R(.)$ is the recovery function application, $\mathbb{E}[.]$ is the expected value operation and $\left|\left|.\right|\right|_2^2$ represents the squared L2 norm.

The **supervised loss** contributes to the **generator loss**, guiding the model to learn **temporal transitions** in the latent space by minimizing the error between predicted and actual future hidden states, thus ensuring temporal consistency. It is defined as:

$$L_{sup} = \mathbb{E}\left[\left|\left|H_{t+1} - H_{(hat)\ t+1}\right|\right|_2^2\right], \tag{3}$$

Where, H_{t+1} is the actual next step hidden-state produced by embedding of real data, $H_{(hat)t+1}$, is the next hidden state predicted by generator from previous latent state and $\mathbb{E}[.]$ is the expected value operation

Meanwhile, the **adversarial loss** appears in both the **generator and discriminator losses**: the generator seeks to produce latent sequences that appear realistic to the discriminator, while the discriminator learns to distinguish between real and synthetic sequences. The adversarial loss for the generator is provided in equation (3).

$$L_{adv(gen)} = \mathbb{E}\left[\log\left(1 - D(H_{hat})\right)\right] \tag{3}$$

And the discriminator loss is:

$$L_{disc} = \mathbb{E}[\log D\,(H)] \; - \mathbb{E}\left[\log\left(1 - D(H_{hat})\right)\right], \tag{4}$$

Where, $D(.)$ is the probability of latent sequence being real output from discriminator, H is the latent space representation of real data and H_{hat} is the latent space representation of synthetic data generated by generator

This **unsupervised adversarial objective** enforces statistical similarity and sequence realism. Together, these losses govern the two main components of TimeGAN-the **autoencoder** (embedding and recovery networks) and the **adversarial training loop** (generator and discriminator)-ensuring that synthetic data is not only accurate and coherent but also temporally plausible.

2.5 Integration into PdM Pipeline

Once trained on real z-HDV battery datasets, the TimeGAN model generates synthetic sequences that effectively replicate a variety of long-term trends in RUL degradation behaviors. These synthetic sequences are then combined with real data to augment the overall dataset size and enrich the training corpus with underrepresented degradation patterns. This enhanced dataset improves the learning capacity of the downstream CNN-LSTM PdM model, leading to better generalization and increased predictive accuracy in RUL estimation.

3 Results

To establish a baseline, the CNN-LSTM PdM model was first trained and evaluated using only real-world battery time-series data from z-HDV systems. This dataset, while rich in temporal sensor information such as voltage, temperature, and current cycles, was relatively limited in size and suffered from class imbalance; particularly underrepresenting critical degradation stages.

Despite these limitations, the model was able to achieve reasonably strong predictive accuracy. The RUL regression model yielded an RMSE of 1.0617 and MAE of 0.8078. However, model generalizability was constrained, particularly in edge cases of long-duration degradation scenarios. The RUL predictions showed higher variance in late-stage battery aging.

Following the generation of synthetic time-series data via the TimeGAN model, an augmented dataset was created by merging the synthetic and real data. The synthetic sequences expanded the training distribution to include rare and complex degradation behaviors that are underrepresented or absent in the original dataset. The adversarial, supervised and construction losses of the TimeGAN network is given below in Table 1.

Table 1: TimeGAN Training Losses

	MSE
Autoencoder	0.0054
Generator	2.0665
Discriminator	2.3532

An instance of one of the generated target variable of PdM network is given below.

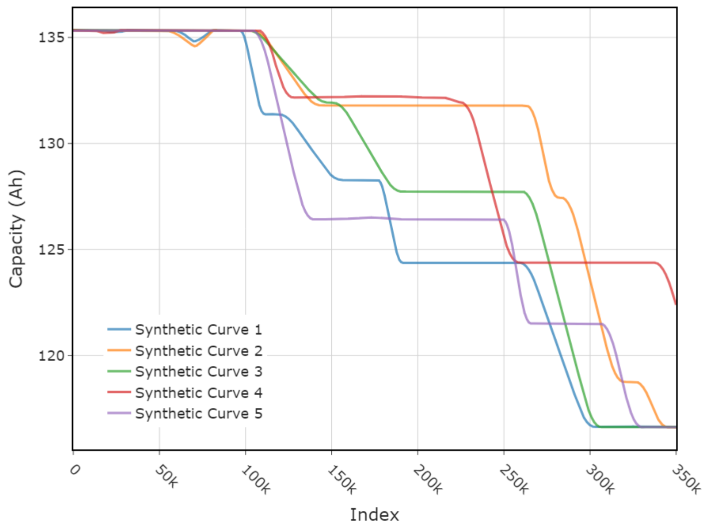


Figure 7: TimeGAN Generated Synthetic Target Variable

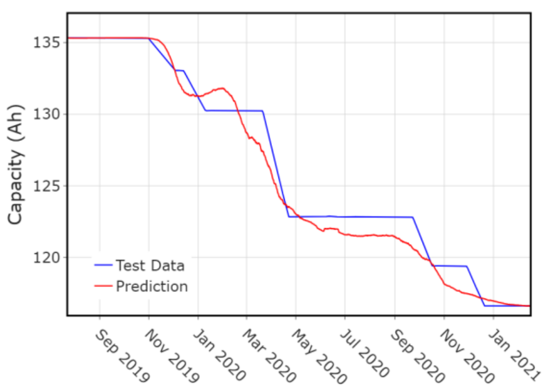
When the CNN-LSTM model has been retrained on this hybrid dataset, substantial improvements were observed across all performance metrics: with RUL regression RMSE dropping to 1.0153, and MAE reducing to 0.6087, indicating enhanced precision in life expectancy estimation.

Given below in Table 2 is a comparative analysis of the two different scenarios tested

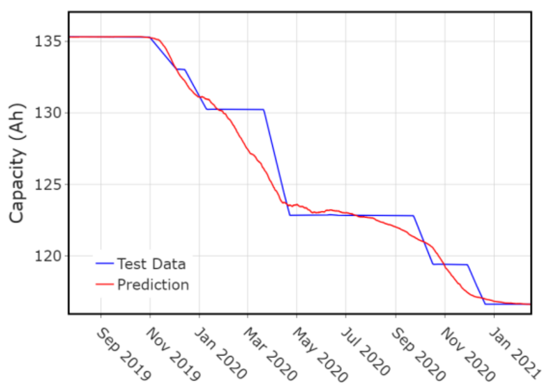
Table 2: Real and Real + Synthetic Errors

	RMSE	MAE
Real Data Only	1.0617	0.8078
Real + Synthetic	1.0153	0.6087

Also, the predictions of the two different trained models on the same real target variable are provided below in Figure 8. It is evident that the second strategy with the synthetic data inclusion improved model performance, by emphasizing the degradation pattern rather than local patterns, decreasing the oscillatory behaviour and providing a smoother transition between constant values.



(a)



(b)

Figure 8: PdM network predictions (a) real data only, (b) synthetic and real data

4 Conclusion

This study introduced a novel data-driven PdM framework tailored for battery RUL regression in zero-emission Heavy-Duty Vehicles. The proposed approach addresses key challenges inherent in industrial battery monitoring, including limited real-world data availability, temporal complexity of degradation patterns, and the need for robust model generalization under data sparsity conditions.

To mitigate these limitations, this study incorporated a Time-series Generative Adversarial Network (TimeGAN) to generate high-fidelity synthetic battery degradation sequences. The TimeGAN model was explicitly trained to preserve both temporal dynamics and feature-level distributions observed in real-world operational data, producing synthetic sequences with statistically and structurally consistent SoH behavior.

The predictive model leveraged a hybrid deep learning architecture consisting of 1D Convolutional Neural Networks (CNNs) for spatial feature extraction and stacked Long Short-Term Memory (LSTM) layers for temporal sequence modeling. Comparative experiments demonstrated that while the CNN-LSTM model trained solely on real data achieved baseline performance levels of RMSE equal to 1.197427, the TimeGAN-augmented variant significantly outperformed it, reducing RMSE to 1.0153. These improvements are attributed to the model's improved ability to capture minority-class dynamics and rare degradation phenomena, thereby increasing predictive resolution and robustness.

The proposed framework offers a scalable, hardware-efficient, and generalizable solution for PdM in electrified heavy-duty transport. It also demonstrates the viability of synthetic data generation in industrial contexts where labeled historical data is scarce, expensive to collect, or insufficiently diverse.

Future work could be focused on: (i) incorporating probabilistic and Bayesian deep learning approaches to quantify prediction uncertainty, and (ii) deploying the pipeline in a real-time embedded edge setting to assess latency, energy efficiency, and on-board inference stability.

References

[1] W. Zhang, D. Yang, H. Wang "Data-Driven Methods for Predictive Maintenance of Industrial Equipment: A Survey" IEEE Systems Journal, Vol. 13, Sep. 2019, doi: 10.1109/JSYST.2019.2905565

[2] MarketsandMarkets. "Predictive Maintenance Market Size, by Technology (Analytics, Data Management, AI, IoT Platform, Sensors), Technique (Vibration Analysis, Infrared Thermography, Oil analysis, Motor Circuit Analysis, Acoustic Monitoring) - Global Forecast to 2029". Mar. 2024, URL: https://www.marketsandmarkets.com/Market-Reports/operational-predictive-maintenance-market-8656856.html

[3] S. Konkimalla, "AI-Based Predictive Maintenance for Electric Vehicles: Enhancing Reliability and Performance," International Journal of Engineering and Computer Science, vol. 11, no. 12, pp. 25647–25661, Nov. 2024. DOI: 10.18535/ijecs/v11i12.4713.

[4] S. Voronov, M. Krysander, and E. Frisk, "Predictive Maintenance of Lead-Acid Batteries with Sparse Vehicle Operational Data," International journal of prognostics and health management, vol. 11, no. 1, Jun. 2023, doi: https://doi.org/10.36001/ijphm.2020.v11i1.2608.

[5] F.K. Wang, and T. Mamo. "A hybrid model based on support vector regression and differential evolution for remaining useful lifetime prediction of lithium-ion batteries." Journal of Power Sources 401 (2018): 49-54.

[6] J. Wu, C. Zhang, and Z. Chen, "An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks," Applied Energy, vol. 173, pp. 134–140, Jul. 2016, doi: https://doi.org/10.1016/j.apenergy.2016.04.057.

[7] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, "Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles," Applied Energy, vol. 339, pp. 120954–120954, Jun. 2023, doi: https://doi.org/10.1016/j.apenergy.2023.120954.

[8] Y. Pang, A. Dong, Y. Wang, and Z. Niu, "Deep learning from three-dimensional Lithium-ion battery multiphysics model Part II: Convolutional neural network and long short-term memory integration," Energy and AI, vol. 17, p. 100398, 2024. doi: 10.1016/j.egyai.2024.100398.

[9] J. Yoon, D. Jarrett, and M. Van der Schaar. "Time-series generative adversarial networks." , Advances in neural information processing systems, 2019, 32

[10] F. Ranja, E. B. Nababan and A. Candra, "Synthetic Data Generation Using Time-Generative Adversarial Network (Time-GAN) to Predict Cash ATM," 2023 International Conference on Computer, Control, Informatics and its Applications (IC3INA), Bandung, Indonesia, 2023, pp. 418-423, doi: 10.1109/IC3INA60834.2023.10285809. keywords: {Training;Measurement;Analytical models;Training data;Predictive models;Prediction algorithms;Data models;Synthetic data generation;Time-GAN;XGBoost},

[11] A. Thompson, A. Sommers, A. Russell-Gilbert, L. Cummins, S. Mittal, S. Rahimi, M. Seale, J. Jaboure, T. Arnold, and J. Church, "Multivariate Data Augmentation for Predictive Maintenance using Diffusion," arXiv preprint arXiv:2411.05848, Nov. 2024.

[12] A. Safdari, E. Frisk, O. Holmer, and M. Krysander, "Synthetic Generation of Streamed and Snapshot Data for Predictive Maintenance," IFAC-PapersOnLine, vol. 58, no. 4, pp. 270–275, 2024, doi: 10.1016/j.ifacol.2024.07.229.

[13] C. Mihale-Wilson, A. Cordes, and M. Lowin, "Synthetic Data Generation for Predictive Maintenance Services," in Proceedings of the SIG SVC Pre-ICIS Workshop 2024, Bangkok, Thailand, 2024. [Online]. Available: https://aisel.aisnet.org/sprouts_proceedings_sigsvc_2024/5

[14] F. Tondro, "Fair Synthetic Time Series Data for Predictive Maintenance," M.S. thesis, Dept. of Information Technology, Uppsala Univ., Uppsala, Sweden, 2024. [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1899960/FULLTEXT01.pdf

[15] Zhang, Z., Yang, F., Ren, D., Liu, J., Wang, Y., Wang, Z., & Jiang, J. (2024). Data-driven approaches for improving the predictive maintenance of lithium-ion batteries using synthetic data. Journal of Power Sources. https://www.sciencedirect.com/science/article/abs/pii/S0378775325005580?dgcid=rss_sd_all

[16] Esteban, C., Hyland, S. L., & Rätsch, G. (2021). TGANs: Enhancing generative models for time-series with physical and medical constraints. arXiv preprint arXiv:2101.03678. https://arxiv.org/abs/2101.03678

Presenter Biography: Batuhan Cinar had both his Bachelor and Master of Engineering degrees from Imperial College London Electrical and Electronic Engineering Department. He is currently a Research Engineer at University of Surrey working on developing AI Predictive Maintenance algorithms for ESCALATE (Grant Agreement No: 101096598 and 10063997) and FASTEST (Grant Agreement No: grant agreement 101103755 and 10078013)