

Article

Multi-Horizon Predictive Maintenance for IoT-Enabled Electric Vehicle Fleets Using a Quantum-Temporal Residual Attention Framework

Mohammad Aldossary ^{1,*} , Jaber Almutairi ² and Ibrahim Alzamil ³ ¹ Department of Software Engineering, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia² Department of Computer Science, College of Computer Science and Engineering, Taibah University, Al-Madinah 42353, Saudi Arabia³ Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Majmaah 11952, Saudi Arabia

* Correspondence: mm.aldossary@psau.edu.sa

Abstract

Predictive maintenance of electric vehicle (EV) fleets requires accurate estimation of Remaining Useful Life (RUL), Time-to-Failure (TTF), and State-of-Health (SOH) from heterogeneous Internet of Things (IoT) telemetry. However, real-world degradation patterns are nonlinear, nonstationary, and highly imbalanced near failure. This study proposes Q-TRACNet, a temporal attention framework that combines causal maintenance-aware preprocessing, adaptive temporal condensation, residual refinement, learnable phase modulation, and hybrid Particle Swarm Optimization–Quantum-Guided Descent parameter tuning. The framework is evaluated on the EV-HLM-RUL dataset and three established prognostics benchmarks: NASA CMAPSS, PHM 2012, and XJTU-SY. Chronological training, validation, and testing partitions are used to preserve temporal causality. On EV-HLM-RUL, Q-TRACNet achieves an MAE of 9.8, an RMSE of 14.7, an R^2 of 0.979, and a Critical Degradation Awareness Index (CDAI) of 0.91. It reduces RMSE by 20.11% relative to the strongest competing baseline and achieves an NRMSE of 0.102 and a Kendall correlation of 0.89 ($p < 10^{-4}$). Cross-dataset experiments demonstrate stable performance for RUL, TTF, and short- and long-horizon SOH prediction. Ablation and sensitivity analyses further confirm the contributions of the temporal and attention components and the stability of degradation-aware evaluation. Q-TRACNet also provides lower training cost and inference latency than competing architectures, supporting practical maintenance planning, inspection prioritization, and resource allocation in connected EV fleets.

Keywords: electric vehicles; predictive maintenance; remaining useful life; state of health; deep learning; Internet of Things

MSC: 68T37; 68T07; 62M10; 90C26; 93C85



Academic Editor: Ioannis Tsoulos

Received: 17 June 2026

Revised: 19 July 2026

Accepted: 23 July 2026

Published: 4 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The global transition toward electric transportation has already accelerated because governments, businesses, and customers want sustainable options to replace gas-powered vehicles. The transportation system, together with public transportation, delivery services, and government vehicle fleets, is using electric vehicles (EVs) because they emit less pollution and cost less to operate and work with renewable energy systems [1]. The EV

adoption process creates problems because it affects how systems work, how safely they operate, and how much money they will need for future upkeep. The vehicle operational capacity and performance efficiency face a major challenge because battery and powertrain parts deteriorate over time [2,3]. The aging process of lithium-ion batteries that power most electric vehicles (EVs) becomes complex because of multiple factors, which include thermal stress, charging behavior, load variations, and environmental conditions. Traditional forecasting methods that rely on rule-based or physics-only models will face difficulties in predicting nonlinear deterioration patterns. The degradation behavior of electric vehicles (EVs) shows different patterns based on the specific vehicle model, the driving conditions, and the geographical areas where they operate. Unscheduled system failures will occur because maintenance planning lacks proper detection methods to find system deterioration that exists between scheduled maintenance checks [4].

Predictive maintenance, which uses data as its foundation, transforms maintenance work from a reactive practice to scheduled maintenance to condition-based maintenance work. Predictive maintenance uses historical and real-time data to forecast component failures, rather than depending entirely on predetermined service intervals [5]. The predictive maintenance process uses Remaining Useful Life (RUL) estimation to forecast the operational duration of a component until it reaches its end-of-life point. The accuracy of RUL estimates enables organizations to implement maintenance work while decreasing equipment downtime and managing their resources effectively [6,7]. The integration of additional sensors with onboard diagnostic systems enables EV systems to use predictive maintenance. Contemporary electric vehicles (EVs) use their Internet of Things (IoT) sensor network system to track their electrical and thermal, mechanical, and operational parameters during their operational time. The sensor system collects extensive multi-variate time-series data, which shows system performance during driving and charging activities. The data enables researchers to study the degradation process which cannot be observed through offline testing and basic testing methods. Real-world telemetry data faces three main challenges, which include noise, missing data, and measurement errors that obstruct analysis and model development [8].

The contemporary research indicates that electric vehicle maintenance requires either precise deterioration forecasts or active monitoring of maintenance progress. The timing of expected future deterioration will determine which maintenance activities need to be conducted. Maintenance operations require short-term forecasts that handle immediate inspection requirements and operational changes, while medium- and long-term forecasts serve maintenance operations, inventory control, and cost reduction activities [9]. The accuracy of predictions declines as components experience aging and operations become more unpredictable based on research findings. Maintenance models need assessment through all deterioration stages instead of just evaluating their average performance throughout their lifespan [10]. The predictive maintenance system for electric vehicle (EV) systems needs to handle multiple challenges even though it has made advancements. The previous methods encountered difficulties in making accurate predictions across various driving conditions and vehicle types because the data samples had either insufficient information or an unbalanced distribution. The models develop a bias toward standard operational settings because they lack data to showcase uncommon yet essential safety-related deterioration stages. The interpretability and operational significance of predicted outputs are crucial for EV system maintenance choices, which influence safety, reliability, and user confidence [11]. The research findings demonstrate that predictive maintenance systems need to develop strong frameworks that harness IoT-powered time-series data while handling unpredictable real-world situations to provide effective maintenance support for electric vehicles. The deployment of electric

vehicles in complex operational environments requires additional research to establish safe, reliable, and scalable systems for their operation throughout various situations [12].

This study addresses fleet-level prognostics using degradation signals from the principal EV subsystem categories represented in the EV-HLM-RUL dataset, including battery and electrical state, traction motor and inverter dynamics, thermal-management conditions, and charging-system stress. These telemetry channels are modeled jointly to estimate vehicle-level Remaining Useful Life, Time-to-Failure, and State-of-Health rather than separate subsystem-specific RUL values. The resulting predictions support maintenance decisions such as inspection prioritization, preventive servicing or replacement, temporary vehicle derating, fleet rotation, workshop scheduling, and spare-parts allocation. Because the dataset does not contain audited monetary repair-cost records, failure cost is not modeled as a direct currency-valued target; instead, its operational consequences are characterized through unplanned downtime, roadside recovery and towing, emergency repair, premature component replacement, service interruption, and safety-related losses. Fleet-scale prediction is more difficult than single-vehicle prognostics because differences in vehicle models, driver behavior, routes, climate, charging practices, and maintenance histories produce non-IID and nonstationary degradation distributions. Fleet operation also introduces large telemetry volumes, rare near-failure samples, simultaneous prediction requirements, and limited workshop and maintenance resources. Q-TRACNet is therefore designed to provide degradation-aware, multi-horizon, and computationally efficient predictions under heterogeneous fleet conditions. The main contributions of this work are summarized as follows:

- Quantum–Temporal Residual Attentive Network (Q-TRACNet) is proposed as a unified prognostic architecture that integrates maintenance-aware temporal condensation, residual refinement, and quantum-inspired attentive modulation within a single learning framework. By explicitly preserving degradation continuity while capturing long-range temporal dependencies, the model provides accurate Remaining Useful Life estimation while maintaining computational efficiency suitable for IoT-enabled EV fleet deployments.
- To effectively handle nonstationary operational conditions and severe class imbalance in real-world telemetry, a Maintenance-Aware Temporal Encoding (MTE) mechanism is developed. This mechanism adaptively weights recent operational states and stress indicators, enabling stable representation learning and improved robustness across short-, medium-, and long-horizon Remaining Useful Life and State-of-Health forecasting tasks.
- Training stability in the highly non-convex prognostic optimization space is strengthened through a hybrid PSO–QGD parameter tuning strategy. By combining global swarm-based exploration with probabilistic gradient-guided refinement, the approach mitigates initialization sensitivity and enhances convergence reliability across heterogeneous datasets.
- In addition to conventional regression metrics, a degradation-centric evaluation formulation incorporating the Critical Degradation Awareness Index (CDAI) is introduced to explicitly quantify predictive reliability near end-of-life thresholds. This formulation strengthens safety-oriented IoT maintenance analytics by emphasizing performance where operational decisions are most critical.
- Comprehensive validation on real-world EV telemetry together with three benchmark prognostics datasets demonstrates statistically significant improvements in RMSE, rank correlation, robustness, and computational efficiency, confirming the scalability and practical deployability of Q-TRACNet for connected vehicle predictive maintenance systems.

2. Related Work

Recent improvements in electric vehicle (EV) predictive maintenance use deep learning architectures to understand complicated, nonlinear deterioration patterns from multivariate time-series data. A transformer ensemble architecture was presented in [13] to enhance early-stage RUL prediction and measure uncertainty. Multiple transformer predictors boosted resilience, but their ensemble architecture increased computing complexity and restricted resource-constrained deployments. A Parallel Feature Fusion Network (PFFN) was suggested in [14] to enhance feature use during early deterioration. This study combined handmade aging markers with deep temporal representations to stabilize early RUL estimations. The technique improved prediction accuracy when using restricted historical data, but its requirement for specific character design restricted its use in different electric vehicle driving conditions.

Vision-transformer-based models function as battery forecasting tools that are used in battery prognostics. DS-ViT-ESA functions as a dual-stream Vision Transformer that employs efficient self-attention mechanisms to calculate battery lifespan with limited charging data [15]. The research achieved its main outcome by reducing the need for complete life cycle data, which required the vision-based representation to transform time-series information into visual formats that concealed detailed time-based patterns. Generative modeling methods deal with situations where measurements are either incomplete or not taken correctly. Quantum-inspired multimodal fusion has been investigated for integrating heterogeneous IoT and unmanned-aerial-vehicle observations in complex monitoring environments [16]. In the prognostics domain, Luo et al. [17] developed a channel and temporal attention network for Remaining Useful Life prediction of stratospheric airships, demonstrating that jointly emphasizing informative sensing channels and degradation periods can improve lifespan estimation. However, these methods do not jointly address EV-fleet heterogeneity, health-state imbalance, multi-horizon SOH estimation, and reliability near critical degradation stages. The researchers used a conditional generative adversarial network in [18] to reconstruct voltage curves, which improved their assessment of SOH. The method enabled feature-complete extraction from incomplete data, but it created training problems through adversarial attacks and unstable model convergence, which restricted the ability to repeat experiments. Researchers investigate physics-aware learning methods to understand how non-linear aging processes occur. The physics-guided model for knee-point detection developed in [19] helps determine how quickly batteries lose capacity during their aging process. The study improved interpretability and forecast accuracy for late-stage battery performance, although it required knee-point feature identification, which does not suit all battery chemistries and usage patterns.

Recent battery prognostics research has increasingly combined deep learning, multi-scale signal representation, and optimization to improve state estimation under nonlinear degradation conditions. One study in [20] used experimental and simulated aging data with multilayer perceptron and long short-term memory models for battery state-of-health estimation. Another study in [21] introduced a Patch Wavelet Transformer that integrates temporal patching, wavelet decomposition, and attention-based learning to capture multi-scale degradation patterns in electric-vehicle batteries. In [22], an optimized long short-term memory network is combined with adaptive unscented Kalman filtering for lithium-ion battery-state estimation. Although these methods improve battery-level state estimation, they mainly address individual battery indicators and do not jointly consider heterogeneous fleet telemetry, multi-horizon RUL, TTF, and SOH prediction, late-stage degradation sensitivity, and computationally efficient deployment.

Research efforts concentrate on hybrid convolution–transformer architectures to determine SOH estimates. The authors proposed HCMDT-Net in [23] as a network structure which consists of hierarchical convolution combined with multi-channel dense transformer components to enhance local feature extraction before global attention modeling takes place. The system could capture cross-channel interactions, but its high parameter requirements created limits on its ability to handle large fleet operations. The project aims to achieve efficient attention processing. The battery health assessment system uses the separable LogSparse self-attention transformer in [24] to reduce computation costs while sustaining essential long-distance connection tracing abilities needed for battery health assessment. The system became less vulnerable to brief transient degradation problems because its sparse design brought efficiency advancements. Hybrid convolutional–recurrent learning has also been used for temporal fault prediction in industrial equipment. Bai et al. [25] combined a Temporal Convolutional Network with a Bidirectional Long Short-Term Memory network to capture localized fault patterns and longer-range dependencies. Although the architecture provides strong temporal representation, its bidirectional processing depends on past and future context and is therefore less suitable for strictly causal online prognostics than the proposed framework. The research in [26] developed a trustworthy estimate system which operates under industrial limitations through its hierarchical stochastic spatial-temporal transformer design. The team advanced SOH prediction accuracy by developing methods that handled both uncertainty and spatial-temporal dependencies. The use of stochastic modeling increases inference time, which creates challenges for real-time maintenance tasks.

The researchers used an exponential transformer method in [27] to create interpretable temporal elements, which they used for predicting RUL. The technique improved temporal attribution clarity, but operational data with high levels of noise reduced its effectiveness. Prognostics use causal temporal modeling methods that extend beyond battery-specific research domains. The authors introduced CART-Net in [28] to develop a causal framework that predicts remaining useful life across different operational environments. The rigorous validation process for causal assumptions in complex EV telemetry established stable performance across different operational conditions. Learning through mixture-of-experts methods helps handle various types of deterioration. The AttMoE framework, based on expert attention, was introduced in [29] to optimize temporal patterns for RUL estimation. Multi-task learning has also been applied to jointly forecast energy consumption, operational cost, and resource allocation in IoT-cloud environments [30]. This operational perspective supports the use of joint predictive outputs when maintenance decisions involve multiple interconnected objectives. The process of expert selection made training more difficult because it worked with all types of deterioration. The current state of hybrid architectural design has progressed to incorporate advanced sequence modeling technologies. The researchers proposed a denoising-enhanced framework that combines advanced sequence modeling with dilated convolution to achieve better RUL prediction results. The multi-stage system increased the difficulty of implementation while making it harder to understand despite its ability to achieve accurate results.

Reliable telemetry acquisition also depends on the continuous operation of embedded monitoring sensors. Wireless power transfer for railway IoT bolts has demonstrated autonomous sensor powering under challenging conductive and ferromagnetic infrastructure conditions [31,32]. Although this work addresses sensing hardware rather than prognostic modeling, it provides a complementary deployment context for maintaining continuous condition-monitoring data in harsh transportation environments.

Multi-horizon prediction has also been strengthened through explicit time–frequency decomposition. Researchers in [33] introduced a wavelet-transform framework for multi-horizon flight-trajectory prediction, showing that multiscale frequency information can reduce accumulated errors across longer forecasting horizons. Although the application differs from component prognostics, the study supports the use of multiscale representations for maintaining prediction stability over increasing horizons.

Optimization-assisted multimodal temporal learning has been applied to lithium-battery life prediction, demonstrating the benefit of combining heterogeneous degradation features with adaptive parameter search [34]. Efficient model transfer has also been investigated through knowledge distillation. A classification and error-correction knowledge-distillation framework is developed for bearing RUL prediction in an air-turbine starter, improving predictive accuracy and generalization while reducing deployment complexity. Nevertheless, these methods remain focused on individual components and do not jointly address fleet-level RUL, TTF, SOH, and imbalanced health-state prediction [35]. Table 1 summarizes the capabilities and limitations of representative predictive maintenance methods relative to Q-TRACNet.

Table 1. Comparison of representative predictive maintenance methods and Q-TRACNet.

Method	Temporal TS Modeling	Long-Range Dependency	Hybrid Architecture	Multi-Horizon Support	Class Imbalance Handling	IoT-Deployable
Transformer Ensemble [13]	✓	✓	✗	✗	✗	✗
PFFN [14]	✓	✗	✓	✗	✗	✓
DS-ViT-ESA [15]	✓	✓	✓	✗	✗	✗
GAN-Assisted Prognostics [18]	✓	✗	✓	✗	✗	✗
Physics-Guided ML [19]	✓	✗	✓	✓	✗	✓
CNN–Transformer Hybrid [23]	✓	✓	✓	✗	✗	✗
Mixture-of-Experts [29]	✓	✗	✓	✓	✗	✗
Proposed Q-TRACNet	✓	✓	✓	✓	✓	✓

3. Proposed Methodology

The proposed predictive maintenance framework is organized as a structured and interconnected processing pipeline that enables operational functions for IoT-enabled electric vehicle fleets. The methodology follows a fixed sequence of processes that develop from telemetry acquisition through maintenance-aware preprocessing to degradation-sensitive feature representation and prognostic modeling through Q-TRACNet and hybrid parameter optimization with the PSO-QGD strategy. The process begins with multivariate operational data that onboard EV sensors and fleet management systems collect to capture all electrical, thermal, mechanical, and usage-related indicators that show how components age over time. The system uses adaptive normalization together with temporal alignment and stress-aware encoding to process raw telemetry data using these three methods to maintain performance during nonstationary operating conditions. The system creates degradation representations that preserve aging continuity through time while reducing transient operational noise. Q-TRACNet processes these representations by using three mechanisms, which include temporal condensation with residual refinement and quantum-inspired attentive modulation to create an efficient model for nonlinear degradation patterns. The hybrid PSO-QGD optimization process refines architectural parameters together with training parameters to improve system performance, which leads to enhanced stability during multiple-dataset tests. The framework as shown in Figure 1 delivers prognostic accuracy results that maintain operational performance in real-world conditions while using computational resources efficiently to support large-scale IoT fleet operations.

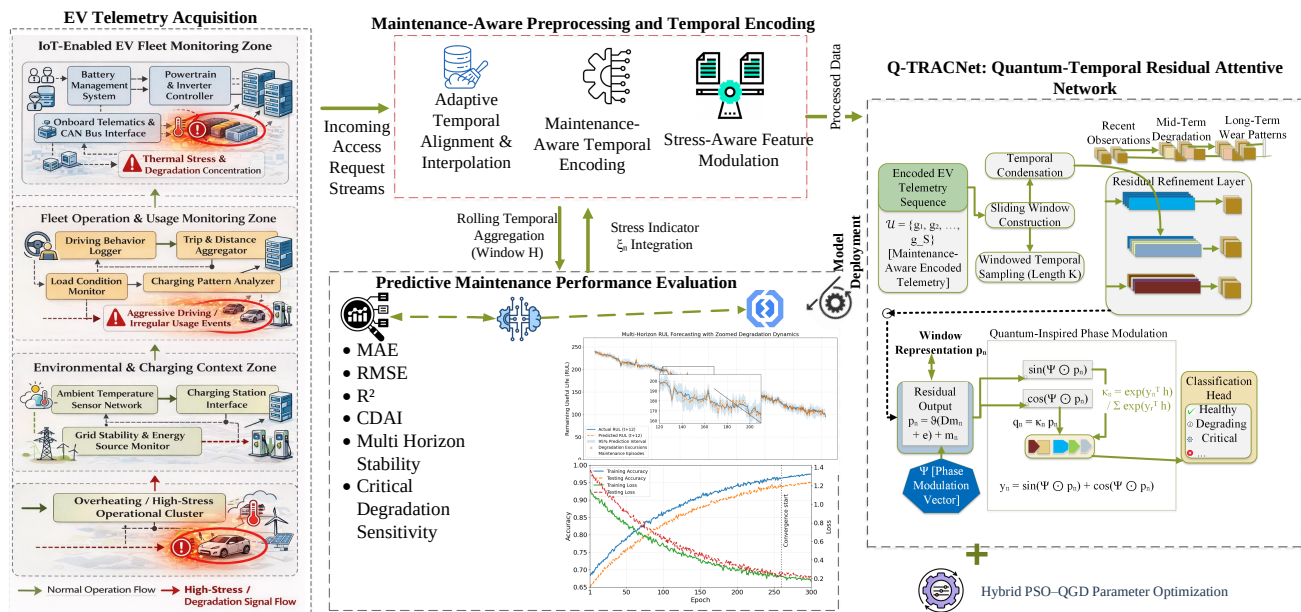


Figure 1. Pictorial view of the proposed framework.

3.1. Dataset Description

The proposed study uses the EV-HLM-RUL dataset, which consists of IoT telemetry data collected through the continuous monitoring of an electric vehicle fleet operating in different urban and highway driving conditions. The dataset includes synchronized multivariate data that safeguards communication between battery management systems, powertrain controllers, charging interfaces, environmental sensors, and fleet management systems. The system collects data through embedded telematics units that connect with vehicle control networks and their cloud-based monitoring system, which allows for continuous condition tracking during all vehicle modes of idle, driving, and charging. From 3 February 2022 to 3 April 2025, telemetry data recording occurs at 15-min intervals, which results in a detailed time-series database that supports degradation-aware modeling [36]. The dataset demonstrates how operational conditions naturally change through dynamic load conditions, driving behavior changes, temperature changes, and charging frequency patterns and maintenance activities. The feature distributions show three patterns: data points extend toward one side, the distribution contains many zero values, and the different operational modes exhibit different usage rates, which match the typical behavior of large-scale connected vehicle monitoring systems.

Table 2 summarizes the monitored attributes and target variables. The behavioral features were calculated over each 15-min telemetry interval. In the proposed preprocessing framework, the Driving Aggressiveness Index (DAI) is formulated as a weighted combination of rapid-acceleration, hard-braking, speeding, and high-jerk events:

$$DAI_n = 100(0.35A_n + 0.35B_n + 0.20S_n + 0.10J_n),$$

where A_n , B_n , S_n , and J_n denote the proportions of valid samples in interval n satisfying acceleration $\geq 2.5 \text{ m/s}^2$, deceleration magnitude $\geq 3.0 \text{ m/s}^2$, speed exceeding the route-specific limit by more than 10 km/h, and absolute jerk $\geq 2.0 \text{ m/s}^3$, respectively. The weights sum to one and assign greater importance to acceleration and braking events because of their direct influence on battery-current demand, regenerative braking, and powertrain stress. DAI values below 20 indicate normal driving, values from 20 to below 40 indicate moderately aggressive driving, and values of 40 or greater indicate aggressive driving. The continuous DAI value, rather than its categorical interpretation, is supplied

to Q-TRACNet. The remaining attributes include battery, powertrain, thermal, charging, environmental, and maintenance-history indicators. The primary target is continuous Remaining Useful Life (RUL), while the derived health states are Healthy, Degrading, and Critical. The EV-HLM-RUL dataset is publicly available on Zenodo for replication and benchmarking [36].

Table 2. High-level organization of dataset attributes and target variables.

Feature Category	Representative Features	Distribution Characteristics
Battery and Electrical State	Voltage, Current, State of Charge, State of Health, Internal Resistance, Charge Cycle Count	Skewed numerical with long-tail degradation trends
Powertrain and Inverter Dynamics	Motor Speed, Torque, Temperature, Vibration Index, Inverter Efficiency, Power Loss Ratio	Imbalanced numerical with stress-induced spikes
Thermal and Environmental Conditions	Battery Temperature, Ambient Temperature, Thermal Gradient, Cooling Load, Overheat Events	Seasonally varying and highly imbalanced numerical
Driving Behavior and Usage Patterns	Vehicle Speed, Acceleration, Braking Intensity, Trip Distance, Driving Aggressiveness Index	Zero-inflated and skewed numerical distributions
Charging and Energy Context	Charging Power, Charging Duration, Fast-Charging Indicator, Charging Frequency, Renewable Share	Highly imbalanced numerical and binary distributions
Operational and Maintenance Context	Operating Mode, Load Condition, Maintenance History, Time Since Last Service	Imbalanced categorical and irregular temporal patterns
Primary Target Label	Remaining Useful Life (RUL)	Continuous regression target with monotonic decay
Derived Target Label	Health Class (Healthy, Degrading, Critical)	Imbalanced categorical derived from RUL thresholds

3.2. Data Preprocessing and Maintenance-Aware Temporal Encoding

The telemetry data gathered from IoT-connected electric vehicles produces a multi-variate time-series which inherently contains measurement errors, operational condition changes, and unexpected user behavior [37]. The signals undergo initial alignment to a 15-min sampling interval because it establishes a standardized timeframe for subsequent observation throughout the entire observation period. Forward interpolation fills short gaps which occur due to intermittent communication or sensor dropouts within specific time periods, while percentile-based clipping reduces the impact of abnormal spikes which happen at irregular intervals. The time series which has been preprocessed, can be expressed as

$$\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}, \quad (1)$$

where $\mathbf{u}_n \in \mathbb{R}^d$ denotes the d -dimensional observation vector recorded at time index n . To accommodate nonstationary operating conditions without introducing future information, each feature is normalized using a strictly causal rolling window of $L = 48$ preceding observations, corresponding to 12 h at the 15-min sampling interval. For feature dimension q and time index $n > L$, the causal mean, sample standard deviation, and normalized value are defined as:

$$\begin{aligned} \eta_{n,q} &= \frac{1}{L} \sum_{s=n-L}^{n-1} u_{s,q}, \\ \rho_{n,q} &= \sqrt{\frac{1}{L-1} \sum_{s=n-L}^{n-1} (u_{s,q} - \eta_{n,q})^2}, \\ \bar{u}_{n,q} &= \frac{u_{n,q} - \eta_{n,q}}{\rho_{n,q} + \delta}, \quad L = 48, \end{aligned} \quad (2)$$

where $\eta_{n,q}$ is the rolling arithmetic mean, $\rho_{n,q}$ is the unbiased sample standard deviation, and $\delta = 10^{-6}$ prevents division by zero for nearly constant signals. The current observation $u_{n,q}$ is transformed using statistics computed only from indices $n - L, \dots, n - 1$; therefore, neither the current target nor any future telemetry value contributes to the normalization parameters. Vehicle records are first ordered chronologically and divided into training, validation, and testing periods before model fitting. Rolling statistics are then generated sequentially within each vehicle trajectory, and no validation or test observation is used to transform an earlier training sample. The first L observations of each vehicle are treated as a causal warm-up period and are excluded from prediction-window construction. This procedure preserves local operating-scale information while preventing look-ahead leakage from future degradation states. Equation (2) is based on standard z-score normalization applied within a causal rolling window for prognostic signals under time-varying operating conditions [37]. Cloud-based forecasting has also been investigated for coordinating electric-vehicle charging with renewable-energy availability and operational demand [38].

The proposed Maintenance-Aware Temporal Encoding mechanism represents recent operational history through a contextual sequence of length H . For each time index n , the sequence is formulated as

$$\mathcal{C}_n = \{\bar{\mathbf{u}}_{n-H+1}, \bar{\mathbf{u}}_{n-H+2}, \dots, \bar{\mathbf{u}}_n\}. \quad (3)$$

The context window elements are not treated with equal weight because the recent observations receive higher weight through the decay-controlled weighting function, which assigns them greater importance,

$$\pi_h = \frac{\exp(-\kappa(H-h))}{\sum_{m=1}^H \exp(-\kappa(H-m))}, \quad h = 1, \dots, H, \quad (4)$$

where κ governs the rate at which historical influence diminishes. The system computes the temporally encoded representation through the following process,

$$\mathbf{v}_n = \sum_{h=1}^H \pi_h \bar{\mathbf{u}}_{n-H+h}. \quad (5)$$

The encoding system reduces the effect of temporary operational changes while preserving vital data about system degradation through time.

The model incorporates cumulative operational stress information to demonstrate how challenging operational conditions impact equipment deterioration. The composite stress indicator ξ_n provides a single value that represents all thermal excursions and aggressive operation and charging irregularities that occur during time period n . The final maintenance-aware state is expressed as

$$\tilde{\mathbf{v}}_n = \mathbf{v}_n \circ (1 + \beta \xi_n), \quad (6)$$

where β controls the sensitivity to stress modulation and $\circ$ denotes element-wise scaling. The representation of encoded information remains proportional to stress levels that occur during specific periods.

The predictive maintenance model uses the sequence $\{\tilde{\mathbf{v}}_n\}_{n=1}^N$ as input data. The preprocessing pipeline achieves efficient operations while maintaining its ability to model both immediate operational changes and extended degradation patterns through its combination of adaptive normalization, stress-aware modulation, and temporally weighted aggregation methods which handle actual degradation prediction from EV time-series data.

3.3. Feature Representation, Temporal Aggregation, and Imbalance Handling

The multivariate telemetry streams collected from IoT-enabled electric vehicles must be transformed into a learning-ready representation that shows progressive degradation while maintaining its ability to handle all changes in operational conditions. After the time series undergo preprocessing together with a maintenance-aware temporal encoding process, the time series is expressed as an ordered sequence [39,40]:

$$\mathcal{H} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_M\}, \quad (7)$$

each vector $\mathbf{r}_m \in \mathbb{R}^q$ provides an overview of the system condition that has been encoded at time index m .

To enable steady learning throughout different time periods, the system extracts fixed-length temporal segments of size W through its sliding-window mechanism. Building on weighted time-series condensation principles [41], each segment is converted into a degradation-oriented representation as

$$\mathbf{c}_m = \sum_{j=0}^{W-1} \psi_j \mathbf{r}_{m-j}, \quad (8)$$

with ψ_j representing the temporal coefficient which decreases over time and satisfies $\sum_{j=0}^{W-1} \psi_j = 1$. The weighting system gives preference to more current operational states while still keeping the historical data, which connects to aging processes that occur over extended periods.

To reduce redundant telemetry without prematurely discarding potentially nonlinear degradation indicators, the feature refinement stage applies a lightweight, target-independent dispersion filter. For lightweight removal of near-constant telemetry channels, we define the normalized dispersion score of feature dimension k as

$$\theta_k = \frac{\text{Var}(c_{:,k})}{\sum_{l=1}^q \text{Var}(c_{:,l})} \quad (9)$$

Here, $c_{:,k}$ denotes the k th aggregated feature across all training samples. Equation (9) is not interpreted as a complete relevance measure; it is used only to remove near-constant dimensions satisfying $\theta_k < \varepsilon_v$, where ε_v is selected using the training partition. Mutual information was not adopted as the primary filtering criterion because its estimation in high-dimensional and temporally correlated telemetry is sensitive to discretization, neighborhood size, and estimator configuration, while repeated training-only estimation would increase computational cost. To avoid discarding nonlinear prognostic indicators, all features with non-negligible variation are retained, and their nonlinear relationships with degradation are subsequently learned through the residual transformation, phase modulation, and attention components of Q-TRACNet. Thus, the variance-based stage removes only uninformative channels rather than serving as a complete measure of feature relevance.

To examine whether variance filtering could exclude low-variance but failure-relevant precursors, Equation (9) was compared with mutual-information and SHAP-based feature filtering. Mutual information between each aggregated feature and RUL was estimated using a k -nearest-neighbor estimator with $k = 5$, whereas SHAP importance was calculated as the mean absolute SHAP value over a stratified sample of 10,000 training windows. For a fair comparison, each method retained the same number of features as the variance-based configuration. All rankings, estimator settings, and filtering thresholds were determined using the training partition only; validation and test observations were not used for feature selection.

The distribution of predictive maintenance data becomes imbalanced because most vehicle operational time occurs when vehicles remain in either a healthy condition or experience mild hardware degradation, but vehicles only reach near-failure state infrequently. To address the imbalance among the health states, inverse-frequency class weighting is applied following established imbalanced-learning practice [42]. The weight assigned to class ℓ is computed as

$$\omega_\ell = \frac{M}{|\mathcal{L}| \cdot M_\ell}, \quad (10)$$

with M representing the complete sample count and M_ℓ indicating the samples that belong to class ℓ and $|\mathcal{L}|$ denoting the total number of health categories. The method enables balanced learning to occur while maintaining the authenticity of the operational data.

The system resolves imbalance when estimating Remaining Useful Life through its optimization process, which concentrates on late-stage degradation. The system uses horizon-aware weighting in the regression objective to assign greater importance to samples with lower life expectancy because these samples between maintenance thresholds need precise prediction.

The resulting representation combines three components, which are temporally aggregated degradation indicators together with relevance-filtered features and imbalance-aware learning signals. This formulation provides a robust and interpretable foundation for Remaining Useful Life prediction and health state assessment from real-world EV telemetry, while remaining computationally efficient and suitable for large-scale deployment.

3.4. Q-TRACNet: Quantum-Temporal Residual Attentive Classification Network

The health state classification task operates directly on multivariate time-series data collected from IoT-enabled electric vehicles. Following preprocessing and maintenance-aware temporal encoding, the telemetry is represented as an ordered sequence of encoded observations,

$$\mathcal{U} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_S\}, \quad (11)$$

where each vector $\mathbf{g}_s \in \mathbb{R}^f$ captures the system condition at the s th time instant. To preserve degradation continuity, overlapping temporal windows of fixed length K are constructed from $\mathcal{U}$, yielding

$$\mathbf{G}_n = \{\mathbf{g}_{n-K+1}, \mathbf{g}_{n-K+2}, \dots, \mathbf{g}_n\}. \quad (12)$$

Each window $\mathbf{G}_n$ constitutes a single health evaluation sample and is processed independently by Q-TRACNet. The framework pictorial view is shown in Figure 2.

Motivated by exponential temporal weighting [41], Q-TRACNet generates adaptive decay coefficients from a trainable unconstrained parameter a through the following normalized function:

$$\begin{aligned} \lambda &= \text{softplus}(a), \\ \xi_j &= \frac{\exp[-\lambda(K-j)/(K-1)]}{\sum_{\ell=1}^K \exp[-\lambda(K-\ell)/(K-1)]}, \quad j = 1, \dots, K, \\ \mathbf{m}_n &= \sum_{j=1}^K \xi_j \mathbf{g}_{n-K+j}. \end{aligned} \quad (13)$$

The raw decay parameter is initialized as $a^{(0)} = 0$, which yields $\lambda^{(0)} = \log 2$ and provides a mild recency-oriented initialization without suppressing long-term degradation information. The softplus and normalized exponential operations guarantee $\xi_j > 0$ and $\sum_{j=1}^K \xi_j = 1$, while preserving a monotonic increase in importance from older to more recent observations. During training, a is optimized jointly with the remaining Q-TRACNet

parameters according to $a^{(t+1)} = a^{(t)} - \eta_t \partial \mathcal{E}_{\text{cls}} / \partial a^{(t)}$, and the normalized decay weights are recomputed at each forward pass. Consequently, the effective temporal memory adapts to the degradation dynamics observed in the training data rather than remaining manually fixed. The operation detects short-term degradation signals while maintaining long-term wear-related information.

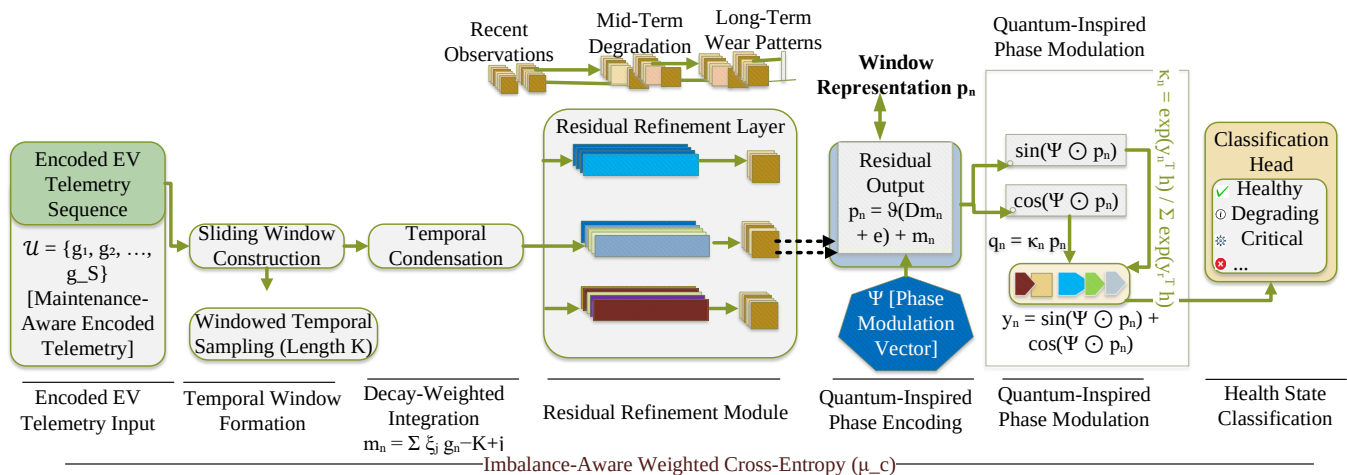


Figure 2. Pictorial view of the proposed Q-TRACNet architecture.

Following the residual-learning principle [43], Q-TRACNet refines the condensed degradation representation while preserving the original information through an identity connection:

$$p_n = \vartheta(Dm_n + e) + m_n, \quad (14)$$

where D and e are trainable parameters and $\vartheta(\cdot)$ denotes a nonlinear activation. The structure of residuals permits the network to develop understanding-based transformations that preserve the original degradation pattern.

Q-TRACNet introduces a learnable phase-modulation stage to capture nonlinear interactions among degradation indicators. The refined representation is projected into a phase-interaction space as

$$y_n = \sin(\Psi \odot p_n) + \cos(\Psi \odot p_n), \quad (15)$$

where Ψ is a learnable feature-wise modulation vector and $\odot$ denotes element-wise coupling. The term quantum-inspired refers only to the use of a phase-like representation and does not imply the simulation of a physical quantum system. Element-wise, the transformation can be written as $y_{n,i} = \sqrt{2} \sin(\Psi_i p_{n,i} + \pi/4)$, providing a smooth and bounded nonlinear mapping with $|y_{n,i}| \leq \sqrt{2}$ and $|\partial y_{n,i} / \partial p_{n,i}| \leq \sqrt{2} |\Psi_i|$. The learnable coefficient Ψ_i controls the sensitivity of each telemetry feature to nonlinear variations, enabling the attention mechanism to distinguish gradual wear, charge–discharge cycling, thermal oscillations, and abrupt stress-induced transitions that may remain indistinguishable under a purely linear projection. Its bounded response also limits the amplification of transient telemetry spikes before attention normalization, thereby improving numerical stability. Thus, the trigonometric transformation serves as a compact degradation-sensitive nonlinear basis rather than representing a literal quantum-physical process.

The proposed modulation differs from conventional sinusoidal encoding and standard attention in both purpose and parameterization. Standard sinusoidal positional encoding assigns fixed frequencies according to the temporal index and therefore identifies the position of an observation, but it does not adapt the encoding frequency to the degradation

behavior of individual telemetry channels. In contrast, Equation (15) applies a learnable feature-wise phase coefficient Ψ_i directly to the refined degradation representation $p_{n,i}$. Standard attention also computes importance directly from a linear projection of $\mathbf{p}_n$ and therefore does not introduce a bounded nonlinear feature interaction before attention scoring. The proposed transformation reshapes the attention-scoring space while Equation (16) continues to weight the original residual representation $\mathbf{p}_n$; consequently, degradation information is not replaced by the trigonometric representation.

To isolate these differences, three controlled alternatives were evaluated while keeping the temporal encoder, residual refinement, prediction head, training schedule, and parameter budget unchanged. Standard attention used $\kappa_n = \text{softmax}(\mathbf{p}_n^\top \mathbf{h})$ without phase transformation. The sinusoidal-encoding variant used $\mathbf{y}_n = \mathbf{p}_n + \mathbf{e}_n$, where the fixed positional vector was defined by $e_{n,2i} = \sin(n/10,000^{2i/f})$ and $e_{n,2i+1} = \cos(n/10,000^{2i/f})$. A fixed-phase variant used $\mathbf{y}_n = \sin(\mathbf{p}_n) + \cos(\mathbf{p}_n)$, corresponding to $\Psi = \mathbf{1}$. These alternatives distinguish the benefit of nonlinear phase mapping from the additional benefit obtained by learning a separate modulation scale for each degradation feature.

Using the established softmax attention-normalization principle [44], the relative importance of each windowed representation is computed as

$$\kappa_n = \frac{\exp(\mathbf{y}_n^\top \mathbf{h})}{\sum_{r=1}^R \exp(\mathbf{y}_r^\top \mathbf{h})}. \quad (16)$$

The attention-adjusted representation is then obtained as

$$\mathbf{q}_n = \kappa_n \mathbf{p}_n, \quad (17)$$

which lets the network prioritize important degradation periods while disregarding unimportant operational times.

The system directs its resulting representation toward a decision layer, which establishes links between learned features and specific class evidence,

$$\mathbf{z}_n = v(\mathbf{L}\mathbf{q}_n + \boldsymbol{\omega}), \quad (18)$$

where $\mathbf{L}$ and $\boldsymbol{\omega}$ are trainable parameters and $v(\cdot)$ denotes a nonlinear mapping. The posterior probability of each health state is computed using

$$\hat{\rho}_n^{(c)} = \frac{\exp(z_{n,c})}{\sum_{d=1}^{|\mathcal{S}|} \exp(z_{n,d})}, \quad c \in \mathcal{S}, \quad (19)$$

along with

$$\mathcal{S} = \{\text{Healthy}, \text{Degrading}, \text{Critical}\}. \quad (20)$$

Following established class-weighted learning for imbalanced data [42], Q-TRACNet employs the following weighted categorical objective:

$$\mathcal{E}_{\text{cls}} = - \sum_{n=1}^R \sum_{c \in \mathcal{S}} \mu_c s_n^{(c)} \log \hat{\rho}_n^{(c)}, \quad (21)$$

where $s_n^{(c)}$ denotes the ground-truth indicator and μ_c is the class-specific weighting factor. The formulation enables health state learning balance while maintaining original data characteristics. Algorithm 1 summarizes the principal processing and optimization steps of Q-TRACNet.

Algorithm 1 Q-TRACNet for Health-State Classification in EV Predictive Maintenance.

Require: Encoded telemetry sequence $\mathcal{U} = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_S\}$, where $\mathbf{g}_s \in \mathbb{R}^f$; temporal window length K ; trainable raw decay parameter a initialized as $a^{(0)} = 0$; trainable parameters $\mathbf{D}, \mathbf{e}, \mathbf{\Psi}, \mathbf{h}, \mathbf{L}, \omega$; nonlinear mappings $\vartheta(\cdot)$ and $v(\cdot)$; class set $\mathcal{S} = \{\text{Healthy}, \text{Degrading}, \text{Critical}\}$; class weights $\{\mu_c\}_{c \in \mathcal{S}}$.

Ensure: Posterior class probabilities $\hat{\rho}_n$ and predicted health-state labels $\hat{c}_n$.

- 1: Compute the number of valid temporal windows: $R \leftarrow S - K + 1$
- 2: Compute the positive decay rate and normalized temporal weights: $\lambda = \text{softplus}(a)$, $\xi_j = \frac{\exp[-\lambda(K-j)/(K-1)]}{\sum_{\ell=1}^K \exp[-\lambda(K-\ell)/(K-1)]}$, $j = 1, \dots, K$
- 3: Initialize phase-feature set $\mathcal{P} \leftarrow []$ and refined-feature set $\mathcal{R} \leftarrow []$
- 4: **for** $n = 1$ to R **do**
- 5: Construct the n th temporal window: $\mathbf{G}_n = \{\mathbf{g}_n, \mathbf{g}_{n+1}, \dots, \mathbf{g}_{n+K-1}\}$
- 6: Perform temporal condensation: $\mathbf{m}_n = \sum_{j=1}^K \xi_j \mathbf{g}_{n+j-1}$
- 7: Apply residual refinement: $\mathbf{p}_n = \vartheta(\mathbf{D}\mathbf{m}_n + \mathbf{e}) + \mathbf{m}_n$
- 8: Compute quantum-inspired phase representation: $\mathbf{y}_n = \sin(\mathbf{\Psi} \odot \mathbf{p}_n) + \cos(\mathbf{\Psi} \odot \mathbf{p}_n)$
- 9: Store $\mathbf{p}_n$ in $\mathcal{R}$ and $\mathbf{y}_n$ in $\mathcal{P}$
- 10: **end for**
- 11: Compute the global attention normalization term: $Z = \sum_{r=1}^R \exp(\mathbf{y}_r^\top \mathbf{h})$
- 12: **for** $n = 1$ to R **do**
- 13: Compute the normalized attention coefficient: $\kappa_n = \frac{\exp(\mathbf{y}_n^\top \mathbf{h})}{Z}$
- 14: Obtain the attention-weighted representation: $\mathbf{q}_n = \kappa_n \mathbf{p}_n$
- 15: Compute decision-layer activations: $\mathbf{z}_n = v(\mathbf{L}\mathbf{q}_n + \omega)$
- 16: Estimate posterior probability for each class $c \in \mathcal{S}$: $\hat{\rho}_n^{(c)} = \frac{\exp(z_{n,c})}{\sum_{d \in \mathcal{S}} \exp(z_{n,d})}$
- 17: Assign the predicted health state: $\hat{c}_n = \arg \max_{c \in \mathcal{S}} \hat{\rho}_n^{(c)}$
- 18: **end for**
- 19: Compute the weighted classification objective: $\mathcal{E}_{\text{cls}} = -\sum_{n=1}^R \sum_{c \in \mathcal{S}} \mu_c s_n^{(c)} \log \hat{\rho}_n^{(c)}$
- 20: Update $a, \mathbf{D}, \mathbf{e}, \mathbf{\Psi}, \mathbf{h}, \mathbf{L}, \omega$ using gradient-based optimization and recompute $\{\xi_j\}_{j=1}^K$ during the subsequent forward pass
- 21: **return** $\{\hat{\rho}_n, \hat{c}_n\}_{n=1}^R$

Each Q-TRACNet component addresses a distinct limitation of EV-fleet prognostics. Maintenance-Aware Temporal Encoding adapts the representation to nonstationary operating conditions and cumulative thermal, charging, and driving stress. Adaptive temporal condensation emphasizes recent degradation evidence while retaining longer-term aging information, thereby reducing transient telemetry noise and sequence-processing cost. Residual refinement preserves the original degradation representation through an identity pathway, improves gradient propagation, and prevents useful health information from being lost during nonlinear transformation. Learnable phase modulation provides a bounded, feature-specific nonlinear mapping that separates gradual wear from abrupt stress-induced changes more effectively than fixed sinusoidal encoding or direct attention scoring. The subsequent attention mechanism prioritizes the most degradation-relevant windows, whereas class-weighted learning limits bias toward the dominant healthy state and improves sensitivity to rare critical conditions. Finally, the hybrid PSO–QGD strategy combines global parameter-space exploration with gradient-guided local refinement, reducing initialization sensitivity and supporting stable convergence. Together, these complementary advantages enable accurate, degradation-aware, and computationally efficient RUL and health-state prediction for heterogeneous EV fleets.

3.5. Hybrid PSO–QGD-Based Parameter Tuning

The joint optimization of Q-TRACNet’s architectural and training parameters constitutes a non-convex search problem. Accordingly, standard Particle Swarm Optimization (PSO) is used for global exploration following the established formulation in [45]. To

avoid repeating well-known PSO theory, only the parameterization, objective function, and update employed in the proposed hybrid optimizer are presented.

Let the complete set of tunable parameters be represented as

$$\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_Z\}, \quad (22)$$

where Λ contains the architectural hyperparameters and trainable coefficients of Q-TRACNet. The tuning objective is defined as

$$\mathcal{J}(\Lambda) = \mathcal{L}_{\text{cls}}(\Lambda) + \tau \mathcal{S}(\Lambda), \quad (23)$$

where $\mathcal{L}_{\text{cls}}$ is the weighted classification loss, $\mathcal{S}(\cdot)$ is a stability regularizer, and τ controls the regularization contribution.

A population of N bounded candidate solutions is initialized for global exploration. Following the standard personal-best and global-best PSO principles [45], the update used in the proposed hybrid optimizer is

$$\Lambda_{t+1}^{(m)} = \Lambda_t^{(m)} + \rho_1 \mathbf{u}_m \odot (\Lambda_{\text{best}}^{(m)} - \Lambda_t^{(m)}) + \rho_2 \mathbf{v}_m \odot (\Lambda^* - \Lambda_t^{(m)}), \quad (24)$$

where $\Lambda_{\text{best}}^{(m)}$ and Λ^* denote the personal-best and global-best solutions, respectively; $\mathbf{u}_m$ and $\mathbf{v}_m$ are random coefficient vectors; and ρ_1 and ρ_2 control the corresponding search contributions.

After global PSO exploration, local refinement is performed through the proposed Quantum-Guided Descent stage, motivated by hybrid gradient-based optimization [46]. Probabilistic scoring provides a structured mechanism for ranking competing alternatives [47]; accordingly, normalized relevance scores are assigned to the candidate solutions as

$$\chi_m = \frac{\exp(-\mathcal{J}(\Lambda^{(m)}))}{\sum_{n=1}^N \exp(-\mathcal{J}(\Lambda^{(n)}))}. \quad (25)$$

The score system manages the parameter update by using scores to perform a weighted gradient update, which fine-tunes locations throughout the local space.

$$\Lambda^{(m)} \leftarrow \Lambda^{(m)} - \zeta \chi_m \nabla_{\Lambda} \mathcal{J}(\Lambda^{(m)}), \quad (26)$$

the rate that updates the cost of the element ζ acts as a growth rate during adaptive refinement. The mechanism that uses probabilistic weighting enables multiple strong candidates to work together during the refinement phase while the system decreases its response to both irregular gradient patterns and initial setup changes.

The hybrid PSO-QGD procedure operates through two distinct phases, which involve searching for solutions and refining the results until the system reaches its stopping points which occur when either loss values stabilize or the predetermined iteration limit is reached. The element

$$\Lambda^{\circ} = \arg \min_{\Lambda^{(m)}} \mathcal{J}(\Lambda^{(m)}), \quad (27)$$

will serve as the basis for Q-TRACNet training and evaluation process. The proposed framework uses population-based global search along with quantum-inspired probabilistic descent to create an optimization system which enables exploratory activities while achieving exact convergence.

The system increases training stability while enhancing generalization abilities, which makes it ideal for predictive maintenance systems in real-world Internet of Things environments that encounter operational data showing random and changing patterns.

3.6. Performance Evaluation

The predictive maintenance framework uses regression performance metrics to evaluate its performance because the Remaining Useful Life (RUL), Time-to-Failure (TTF), and multi-horizon State-of-Health (SOH) trajectories represent continuous-valued data. The model performance evaluation uses Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2), which collectively measure how much predictions deviate from actual values and how sensitive the model is to major prediction errors, and how well the model matches actual degradation patterns. For the imbalanced health-state classification task, Macro-F1 is reported because it assigns equal importance to each class and provides a more informative assessment than overall accuracy when class frequencies differ substantially [48]. To evaluate prediction reliability during the most maintenance-critical stage, we introduce the Critical Degradation Awareness Index (CDAI). Let $\mathcal{C}$ denote the subset of samples whose true RUL values fall below the critical threshold τ . CDAI is defined as

$$\text{CDAI} = 1 - \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \frac{|r_i - \hat{r}_i|}{r_i + \epsilon}, \quad (28)$$

where r_i and $\hat{r}_i$ denote the actual and predicted RUL values, respectively, $|\mathcal{C}|$ is the number of critical-stage samples, and ϵ is a scale-adjusted stabilizing constant. To avoid dependence on dataset-specific RUL units, the critical threshold is expressed as: $\tau_\gamma = \gamma r_{\max}$, where $r_{\max}$ is the maximum RUL in the corresponding dataset and γ is a normalized end-of-life ratio. The primary experiments use $\gamma = 0.20$, selected using the validation partition only, while sensitivity is evaluated over $\gamma \in \{0.10, 0.15, 0.20, 0.25, 0.30\}$. Higher CDAI values indicate more reliable prediction within the selected critical degradation region. This normalized formulation permits comparable threshold interpretation across EV, turbofan, and bearing prognostics datasets with different lifecycle scales.

4. Simulation Results and Discussion

The section offers an in-depth assessment of the Q-TRACNet framework, which serves multi-horizon electric vehicle predictive maintenance. The experiments were conducted on the EV-HLM-RUL dataset using a chronological training, validation, and testing split that maintains chronological order to enable authentic degradation prediction. This study simulated the implementation in Python through the PyTorch deep learning framework, which ran on a workstation that featured an NVIDIA GeForce RTX 4060 GPU, 64 GB RAM, and a multi-core Intel processor. The model was trained through adaptive momentum-based optimization, which applied cosine learning rate scheduling and early stopping to control overfitting. The study used MAE and RMSE and R^2 and MAPE together with the Critical Degradation Awareness Index (CDAI) to assess performance. We conducted experiments on three standard prognostics benchmarks to validate their findings on the original EV dataset and assess the ability to generalize. Moreover, it also provides a comprehensive analysis of predictive accuracy, convergence behavior, ablation results, computational efficiency, and statistical robustness.

The multi-horizon EV predictive maintenance study used the experimental environment and model configuration shown in Table 3 as its testing base. The setup reflects a realistic, resource-aware training pipeline, which uses long-term, high-resolution telemetry data and maintains evaluation time consistency throughout the entire testing process. The configuration helps Remaining Useful Life prediction by using horizon-specific learning

methods, stability-oriented optimization techniques, and degradation-sensitive loss design, which enable accurate predictions for short-, medium-, and long-term forecasting while maintaining reproducibility and enabling researchers to compare results with previous predictive maintenance models.

Table 3. Simulation environment and model configuration for multi-horizon EV predictive maintenance.

Aspect	Specification/Selected Value
Computing Platform	NVIDIA GeForce RTX 4060 GPU, 64 GB RAM, Intel-based multi-core CPU
Software Stack	Python 3.10, PyTorch 2.1, CUDA 12.0, NumPy 1.26, SciPy 1.11
Deployment Setting	Centralized training with time-aware evaluation (chronological split)
Data Span and Resolution	Multi-year EV telemetry, sampled at 15-min intervals
Input Representation	96-dimensional multivariate time-series vectors
Temporal Window Size	48 time steps (12-h historical context)
Prediction Horizons	Short (H1), Medium (H7), Long (H30)
Learning Rate	6×10^{-4} with cosine decay schedule
Mini-Batch Size	72
Optimization Strategy	Adaptive momentum-based optimizer with gradient normalization
Regularization	Dropout rate 0.2; temporal smoothness penalty
Training Epochs	120 (early stopping after 18 epochs without validation improvement)
Loss Formulation	Horizon-weighted regression loss with degradation emphasis
Training Time per Epoch	1.84 s
Inference Latency	14.2 ms per sample
Serialized Memory Footprint	402 MB
Evaluation Criteria	MAE, RMSE, MAPE, R^2 , CDAI

The worldwide demonstration of Q-TRACNet shows how different operational and health assessment metrics affect Remaining Useful Life prediction through the research results shown in Figure 3. The ranking shows that battery-based degradation signals offer the most predictive value while environmental and usage-related factors deliver additional contextual information. The proposed model establishes the existence of physical dependencies because it learned genuine physical connections instead of false relationships. The Q-TRACNet system presents a novel approach because it maintains interpretability across complex architectural designs, which allows users to conduct transparent maintenance evaluations and trust operations in actual electric vehicle fleet deployments.

Table 4 compares Equation (9) with mutual-information and SHAP-based filtering under an equal retained-feature budget. SHAP-based filtering provides a marginal improvement of 0.1 in MAE and RMSE, while the variance-based method remains within 1.03% in MAE and 0.68% in RMSE of the SHAP result. Mutual-information filtering produces similarly close performance. Inspection of the selected subsets further showed that the principal battery, thermal, inverter, vibration, and charging-stress precursors ranked as failure-relevant by both SHAP and mutual information were also retained by the variance-based filter. Thus, the conservative variance threshold did not remove the principal failure-critical signals, while providing a target-independent and computationally lightweight preprocessing stage.

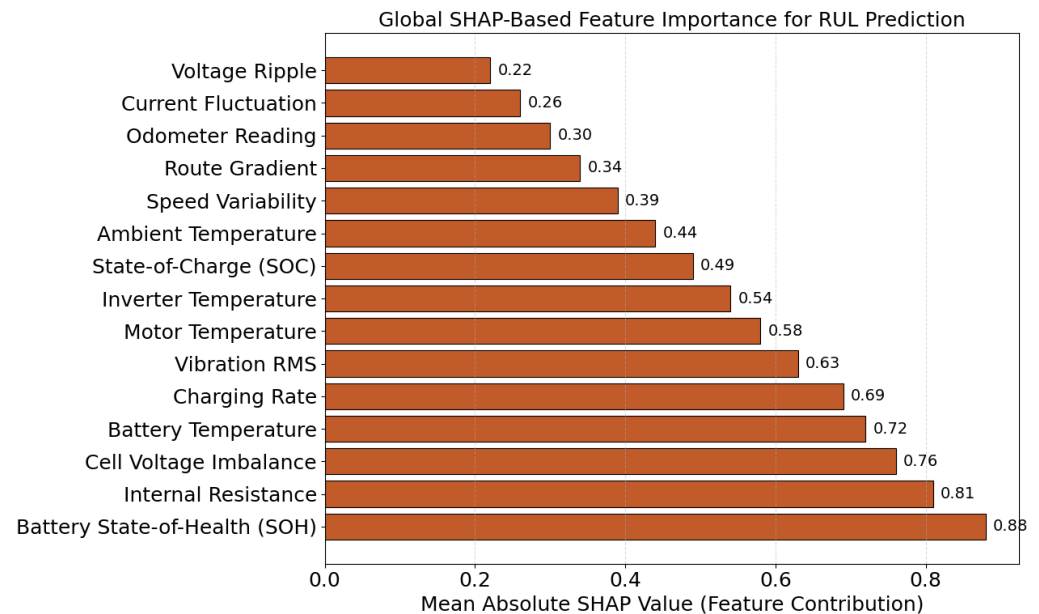


Figure 3. Global SHAP-based feature importance for RUL prediction using Q-TRACNet.

Table 4. Comparison of variance-based, mutual-information, and SHAP-based feature filtering on EV-HLM-RUL.

Feature-Selection Method	Features	MAE ↓	RMSE ↓	R^2 ↑	CDAI ↑
No feature filtering	96	10.3	15.4	0.976	0.897
Variance ratio, Equation (9)	82	9.8	14.7	0.979	0.910
Mutual-information filtering	82	9.9	14.8	0.978	0.908
SHAP-based filtering	82	9.7	14.6	0.980	0.912

Note: ↓ indicates that lower values are better, whereas ↑ indicates that higher values are better.

The coupled convergence behavior of Q-TRACNet demonstrates its dual capability to stabilize both optimization and generalization performance throughout extended training periods according to Figure 4. The results of the architectural coupling and regularization method show their effectiveness through the synchronized decline of training and testing loss, which narrows the accuracy gap after the point of convergence. Q-TRACNet establishes its innovative approach through balanced learning, which maintains stability and helps the model preserve its RUL regression performance under real-world operational situations, which typically lead to RUL regression model degradation.

Figure 5 shows how well the predicted Remaining Useful Life paths match the actual paths under real-world operational conditions. The model successfully maintains its predictive accuracy through all stages of degradation according to the lifecycle assessment, which shows the model's tracking performance and the detailed examination of the end-of-life period. Q-TRACNet provides stable and continuous failure threshold estimation despite its operation close to these limits, while traditional methods show prediction problems at these points. The new framework developed in this research provides a method to track long-term degradation patterns, which maintains its ability to detect important late-stage changes needed for maintenance decision-making.

Table 5 shows how Q-TRACNet performs multi-task regression with its native and derived prognostic targets. The framework shows stable performance across all three domains of RUL and TTF and both SOH estimation time frames because it maintains cross-target generalization throughout its testing on the EV-HLM-RUL dataset. The proposed method achieves better accuracy and higher explained variance than existing benchmark datasets

because it can handle different patterns of degradation. The results demonstrate that the system supports multi-objective predictive maintenance modeling by processing multiple prognostic labels instead of using a single label.

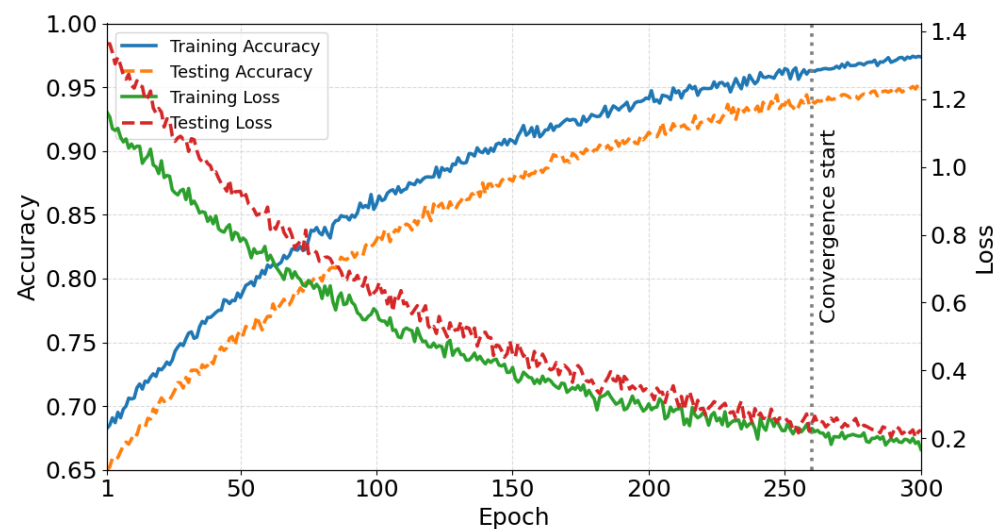


Figure 4. Training convergence behavior of Q-TRACNet showing accuracy and loss evolution over 300 epochs.

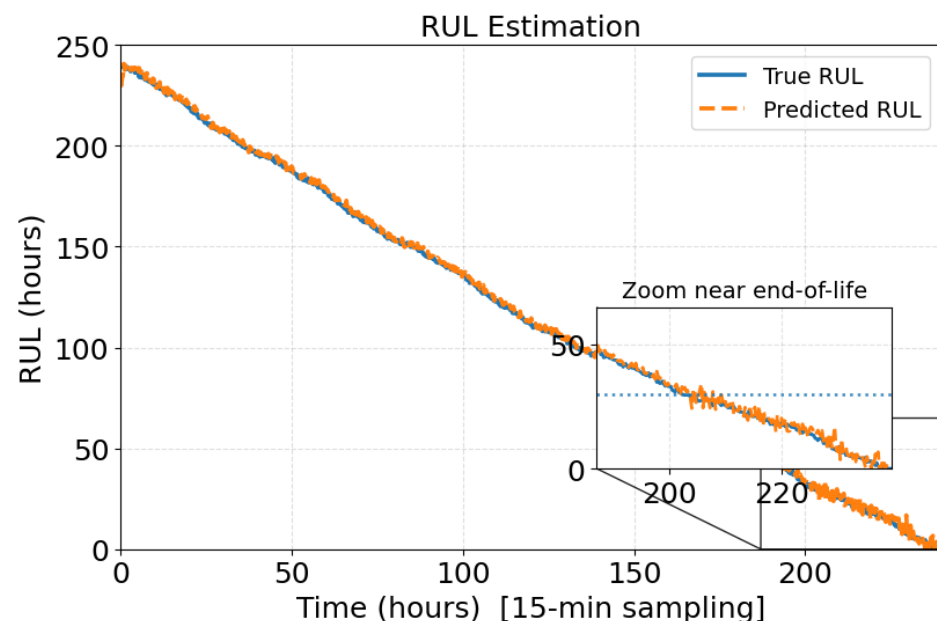


Figure 5. Remaining Useful Life (RUL) estimation with end-of-life zoomed analysis using Q-TRACNet.

Figure 6 demonstrates how Q-TRACNet maintains its performance across longer prediction periods, which scientists studied through performance evaluations of stability, reliability, and uncertainty measurement. The system demonstrates stable predictive capabilities because its stability metrics decrease at an expected rate from short-term predictions to extended forecasting periods. Q-TRACNet demonstrates a gradual performance decline that matches the increasing operational uncertainty, while traditional RUL models experience immediate reliability loss with longer prediction periods. The model demonstrates its original design by enabling multi-horizon maintenance planning under electric vehicle operating conditions while maintaining system reliability.

Table 5. Multi-task regression performance of Q-TRACNet across native and derived prognostic targets.

Dataset	Prediction Target	MAE ↓	RMSE ↓	R^2 ↑
EV-HLM-RUL (Proposed)	Remaining Useful Life (RUL)	9.8	14.7	0.979
	Time-to-Failure (TTF)	10.1	15.0	0.976
	SOH (Short-Horizon, 24–48 h)	0.031	0.046	0.983
	SOH (Long-Horizon, 7–14 days)	0.044	0.062	0.971
NASA CMAPSS	Remaining Useful Life (RUL)	13.2	18.8	0.908
	Derived SOH Index	0.057	0.081	0.914
PHM 2012	Remaining Useful Life (RUL)	12.5	18.0	0.914
	Derived SOH Index	0.054	0.078	0.919
XJTU-SY	Time-to-Failure (TTF)	14.7	20.6	0.901
	Derived RUL	14.3	20.1	0.905

Note: ↓ indicates that lower values are better, whereas ↑ indicates that higher values are better. For the benchmark datasets (NASA CMAPSS, PHM 2012, and XJTU-SY), State-of-Health (SOH), Remaining Useful Life (RUL), and Time-to-Failure (TTF) are derived from run-to-failure trajectories and normalized degradation indices following the established prognostics literature.

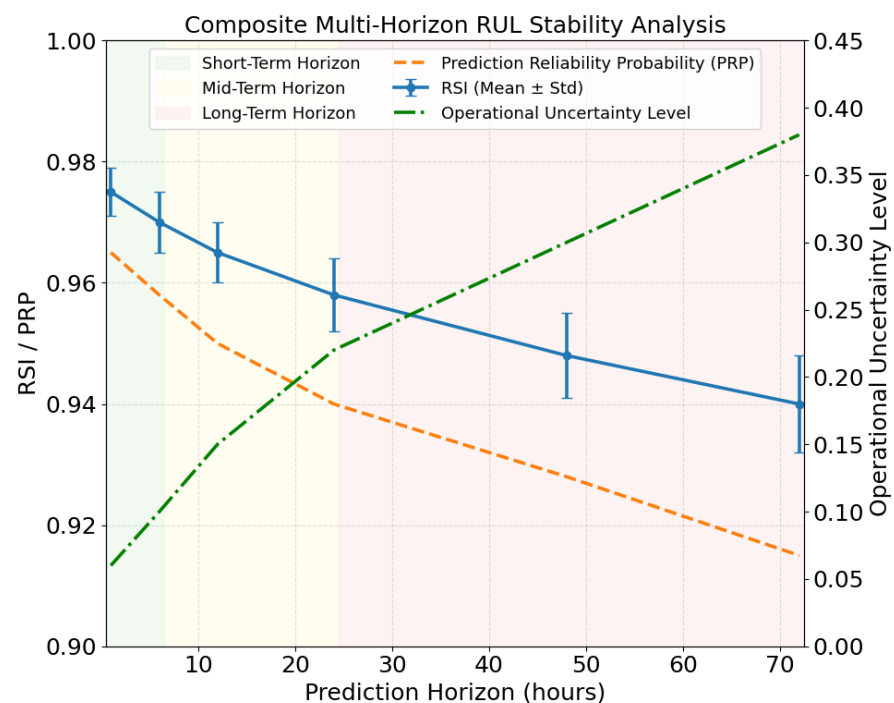
**Figure 6.** Composite multi-horizon stability analysis of RUL prediction using Q-TRACNet.

Table 6 quantitatively confirms the advantages of the principal Q-TRACNet components. Relative to the model without Maintenance-Aware Temporal Encoding, the complete framework reduces MAE by 7.55% and RMSE by 5.77%, while increasing CDAI from 0.88 to 0.91, confirming the benefit of incorporating operating-stress and recency information. Temporal condensation provides MAE and RMSE reductions of 5.77% and 4.55%, respectively, by compressing the historical window into a degradation-oriented representation. Residual refinement reduces MAE by 3.92% and RMSE by 3.29%, demonstrating the advantage of preserving the original degradation signal while learning nonlinear corrections. Compared with standard attention, learnable phase modulation reduces MAE by 6.67% and RMSE by 6.37%, while improving CDAI from 0.88 to 0.91. The fixed-phase and fixed-sinusoidal variants remain below the complete model, showing that both nonlinear phase mapping and feature-wise adaptation are beneficial. Finally, the full framework reduces RMSE by 8.70% relative to the backbone-only configuration and improves CDAI from 0.87 to 0.91, confirming that the components provide complementary rather than redundant advantages.

Table 6. Ablation and controlled attention comparison for Q-TRACNet on EV-HLM-RUL.

Variant	Representation Used for Attention Scoring	MAE ↓	RMSE ↓	R ² ↑	CDAI ↑
Full Q-TRACNet	Learnable feature-wise phase modulation	9.8	14.7	0.979	0.91
Fixed-phase mapping	$\sin(\mathbf{p}_n) + \cos(\mathbf{p}_n)$	10.0	15.0	0.978	0.90
Fixed sinusoidal positional encoding	$\mathbf{p}_n + \mathbf{e}_n$	10.2	15.3	0.977	0.89
Standard attention	Direct score $\mathbf{p}_n^\top \mathbf{h}$	10.5	15.7	0.975	0.88
Without MTE	Learnable phase modulation	10.6	15.6	0.974	0.88
Without temporal condensation	Learnable phase modulation	10.4	15.4	0.975	0.89
Without residual refinement	Learnable phase modulation	10.2	15.2	0.976	0.89
Backbone only	No condensation, residual refinement, or phase modulation	10.9	16.1	0.972	0.87

Note: ↓ indicates that lower values are better, whereas ↑ indicates that higher values are better. MTE denotes Maintenance-Aware Temporal Encoding.

Table 7 evaluates the computational suitability of Q-TRACNet using training time per epoch, inference latency, serialized memory footprint, RMSE, and CDAI. The proposed framework requires 1.84 s per training epoch, 14.2 ms for inference, and 402 MB of serialized model memory. Relative to the strongest competing baseline for each computational index, Q-TRACNet reduces training time by 2.13%, inference latency by 7.79%, and memory footprint by 1.95%. It also reduces RMSE by 20.11% relative to DM-DC while increasing CDAI by 10.98%. Across all evaluated baselines, the RMSE reduction ranges from 20.11% to 40.96%, whereas the CDAI improvement ranges from 10.98% to 33.82%. These results indicate that Q-TRACNet provides substantial predictive improvements while maintaining lower execution time, latency, and memory requirements than the competing architectures. The 14.2 ms inference latency and 402 MB memory footprint support its suitability for deployment on GPU-enabled fleet gateways, vehicle-management servers, and edge-computing platforms, although further compression would be required for highly constrained embedded controllers.

Table 7. Computational efficiency, memory footprint, and predictive performance comparison of Q-TRACNet and competing methods.

Model	Train Cost/Epoch (s)	Inference Latency (ms)	Serialized Memory Footprint (MB)	RMSE ↓	CDAI ↑
TE [13]	2.14	18.9	468	24.9	0.68
PFFN [14]	1.96	16.7	432	23.6	0.70
DS-ViT [15]	2.08	17.5	445	22.4	0.72
cGAN-VR [18]	2.21	19.3	486	23.1	0.69
KP-PGML [19]	1.88	15.9	410	21.3	0.74
HCMDT [23]	2.03	16.8	438	20.6	0.76
LS-SA-T [24]	1.91	15.4	421	20.1	0.77
HS-STT [26]	2.17	18.1	462	19.4	0.79
ET [27]	2.06	16.6	439	18.8	0.81
CART-Net [28]	1.99	16.2	428	19.0	0.80
AttMoE [29]	2.04	17.0	444	19.7	0.78

Table 7. Cont.

Model	Train Cost/Epoch (s)	Inference Latency (ms)	Serialized Memory Footprint (MB)	RMSE ↓	CDAI ↑
DM-DC [34]	1.92	15.8	417	18.4	0.82
Q-TRACNet (Proposed)	1.84	14.2	402	14.7	0.91
Improvement over Best Baseline (%)	2.13	7.79	1.95	20.11	10.98

Note: ↓ indicates that lower values are better, whereas ↑ indicates that higher values are better.

Table 8 evaluates the robustness of Q-TRACNet under realistic operational stresses encountered in IoT-enabled electric vehicle fleets. The evaluation of performance proceeds through three specific assessment methods, which examine non-IID fleet heterogeneity, missing telemetry data, noisy telemetry data, and intermittent edge device dropout. The prediction accuracy declines at a slow rate because distribution diversity increases while data quality decreases, which demonstrates that the system can generalize beyond its training data. The regression metrics MAE, RMSE, and R^2 together with the degradation-aware CDAI demonstrate stable performance under extreme testing conditions, while Macro-F1 ensures accurate health-state classification. The low p-values further confirm the presence of statistical significance. The results show that Q-TRACNet achieves reliable predictive maintenance results across various noisy, partial, and diverse real-world conditions.

Table 8. Robustness of Q-TRACNet under non-IID fleet heterogeneity, missing/noisy EV telemetry, and edge device dropout in IoT-enabled electric vehicle predictive maintenance (mean ± std.).

Stress Axis	Setting	Macro-F1 (%)	MAE	RMSE	R^2	CDAI	p-Value
A. Non-IID Fleet Heterogeneity (Dirichlet vehicle distribution parameter α ; lower α = stronger diversity)							
Non-IID (α)	$\alpha = 1.0$ (mild)	98.7 ± 0.13	9.6 ± 0.3	14.5 ± 0.4	0.980 ± 0.002	0.92 ± 0.01	0.041
Non-IID (α)	$\alpha = 0.5$ (moderate)	98.3 ± 0.15	9.9 ± 0.4	14.9 ± 0.5	0.979 ± 0.003	0.91 ± 0.01	0.028
Non-IID (α)	$\alpha = 0.3$ (high)	97.9 ± 0.18	10.3 ± 0.5	15.4 ± 0.6	0.977 ± 0.004	0.90 ± 0.02	0.016
Non-IID (α)	$\alpha = 0.1$ (severe)	97.3 ± 0.21	10.9 ± 0.6	16.1 ± 0.7	0.974 ± 0.005	0.88 ± 0.02	0.009
B. Missing Telemetry and Sensor Noise (EV onboard IoT streams)							
Missing rate	0%	98.8 ± 0.12	9.5 ± 0.3	14.4 ± 0.4	0.981 ± 0.002	0.92 ± 0.01	0.043
Missing rate	10%	98.1 ± 0.16	10.1 ± 0.4	15.1 ± 0.5	0.978 ± 0.003	0.90 ± 0.02	0.024
Missing rate	20%	97.4 ± 0.20	10.8 ± 0.6	15.9 ± 0.7	0.975 ± 0.004	0.88 ± 0.02	0.011
Noise level	0%	98.8 ± 0.12	9.5 ± 0.3	14.4 ± 0.4	0.981 ± 0.002	0.92 ± 0.01	0.043
Noise level	10%	97.9 ± 0.17	10.2 ± 0.5	15.2 ± 0.6	0.977 ± 0.003	0.90 ± 0.02	0.019
Noise level	20%	97.0 ± 0.22	11.1 ± 0.7	16.3 ± 0.8	0.973 ± 0.005	0.87 ± 0.03	0.007
C. Edge Device Dropout (intermittent vehicle connectivity during distributed data collection)							
Dropout	0%	98.8 ± 0.12	9.5 ± 0.3	14.4 ± 0.4	0.981 ± 0.002	0.92 ± 0.01	0.043
Dropout	10%	98.4 ± 0.14	9.7 ± 0.3	14.7 ± 0.4	0.980 ± 0.002	0.92 ± 0.01	0.036
Dropout	20%	98.0 ± 0.17	10.1 ± 0.4	15.2 ± 0.5	0.978 ± 0.003	0.90 ± 0.02	0.021
Dropout	30%	97.5 ± 0.20	10.6 ± 0.6	15.8 ± 0.7	0.976 ± 0.004	0.89 ± 0.02	0.012

The maintenance timeline in Figure 7 shows how Remaining Useful Life progresses through time together with maintenance activities and scheduled re-forecasting periods. The figure demonstrates how Q-TRACNet continuously updates RUL estimates as new telemetry becomes available, while remaining stable across normal operation, warning periods, and service windows. The timeline demonstrates how the new framework helps electric vehicle fleets practice maintenance planning through its three visual elements, which show degradation patterns, critical thresholds, and update times.

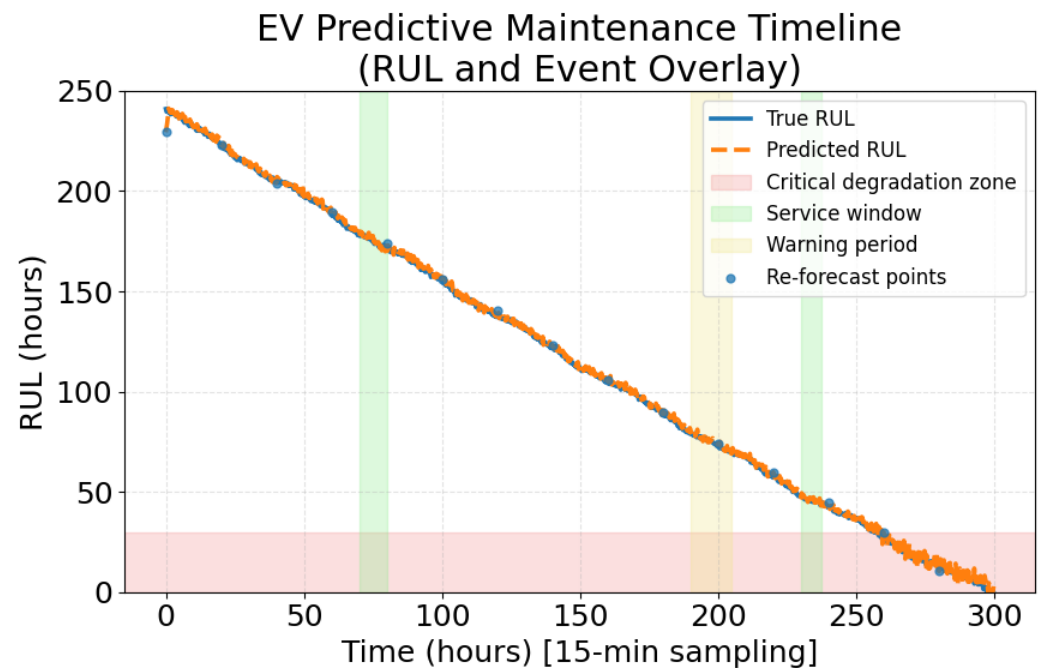


Figure 7. EV predictive maintenance timeline with RUL evolution, event overlays, and re-forecast points using Q-TRACNet.

Figure 7 illustrates the time-dependent evolution of the true and predicted RUL values together with the critical degradation zone, service windows, warning period, and periodic re-forecast points. The close agreement between the true and predicted trajectories demonstrates the ability of Q-TRACNet to continuously update RUL estimates while supporting timely maintenance planning as the vehicle approaches critical degradation.

Figure 8 presents the component-level training dynamics and contribution stability of Q-TRACNet across the early, middle, and late training stages. The progressively decreasing feature-space divergence of the temporal encoding, residual refinement, and attentive modulation components indicates stable convergence during training. Meanwhile, the relatively consistent contribution weights demonstrate that the three components maintain balanced and complementary roles throughout the optimization process.

The multi-horizon RUL forecasting behavior of Figure 9 demonstrates its main function by showing the complete operational period, which includes future target points, model output predictions, and uncertainty intervals. The shaded interval highlights confidence variation under changing operating conditions, while marked excursion and maintenance episodes emphasize clinically relevant degradation irregularities and partial recovery patterns. The zoomed inset demonstrates that Q-TRACNet maintains continuous tracking during localized degradation transitions without generating any oscillation artifacts. The proposed framework introduces a novel approach that provides horizon-consistent prognosis with calibrated uncertainty, supporting maintenance planning in nonstationary fleet operations.

Table 9 shows that different regression-based prognostic models achieve different levels of statistical robustness and significance. The Q-TRACNet system delivers its highest statistical reliability through its testing results, which show Friedman and Diebold–Mariano test values below 10^{-4} evidence of its consistent performance lead across multiple test sessions. The measurement of Kendall's τ shows better rank-order matching with actual degradation paths, while the NRMSE value, which is the smallest, demonstrates that actual errors are distributed less than expected. The Cliff's δ value shows a practical effect size which is much greater than baseline methods. The results show that performance

improvements result from actual effects, which researchers can measure through statistical tests and which have real-world value.

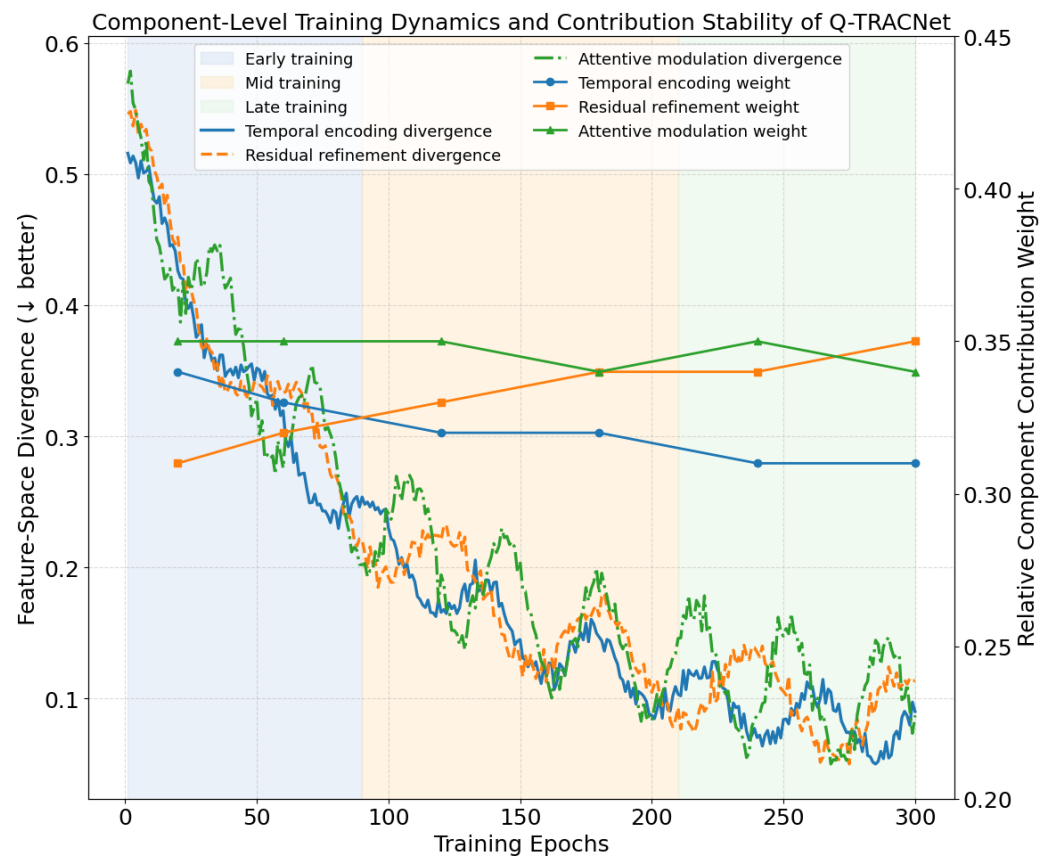


Figure 8. Component-level training dynamics and contribution stability of Q-TRACNet.

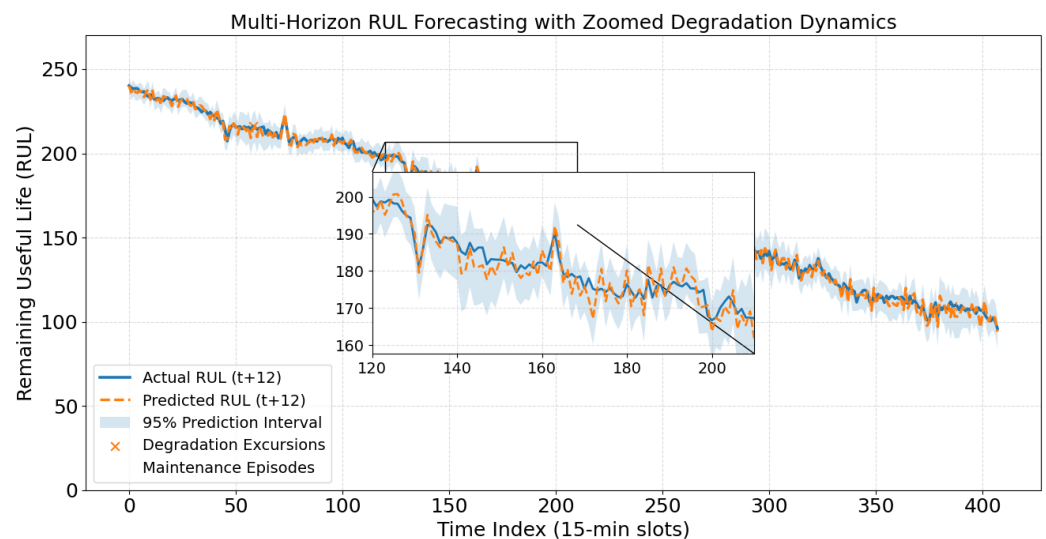
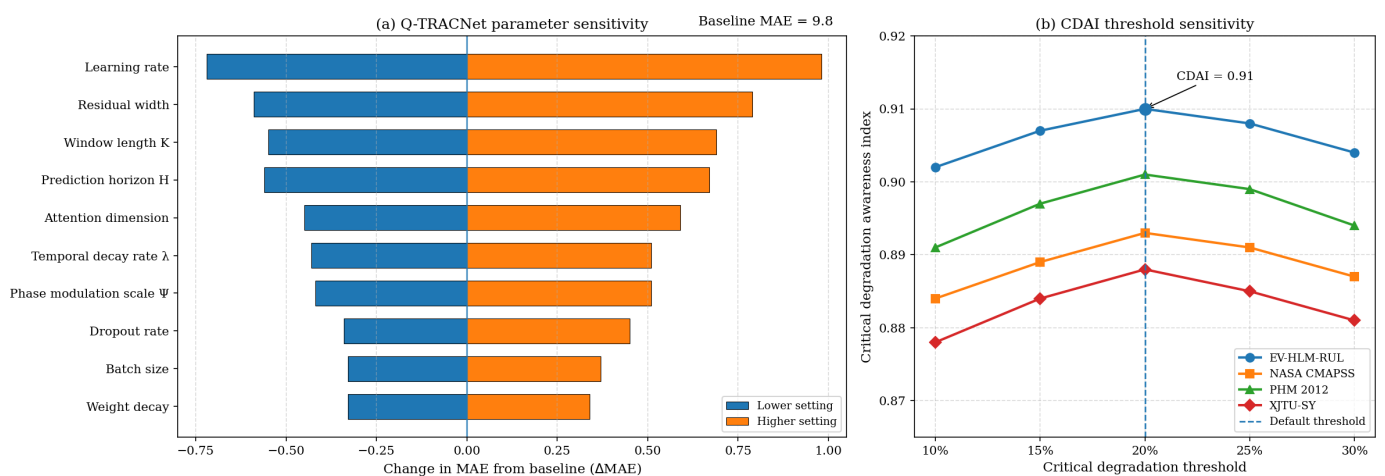


Figure 9. Multi-horizon RUL forecasting with prediction intervals and zoomed degradation dynamics using Q-TRACNet.

Table 9. Statistical robustness and significance analysis of regression-based predictive maintenance models.

Method	Friedman Test (p)	Diebold–Mariano (p)	Kendall τ	NRMSE	Cliff’s δ
TE [13]	4.2×10^{-3}	5.1×10^{-3}	0.61	0.184	0.42
PFFN [14]	3.8×10^{-3}	4.6×10^{-3}	0.64	0.176	0.46
DS-ViT [15]	3.3×10^{-3}	4.1×10^{-3}	0.67	0.168	0.49
cGAN-VR [18]	3.9×10^{-3}	4.9×10^{-3}	0.63	0.172	0.45
KP-PGML [19]	2.7×10^{-3}	3.5×10^{-3}	0.69	0.159	0.54
HCMDT [23]	2.2×10^{-3}	3.1×10^{-3}	0.72	0.151	0.58
LS-SA-T [24]	2.0×10^{-3}	2.8×10^{-3}	0.74	0.147	0.61
HS-STT [26]	1.6×10^{-3}	2.3×10^{-3}	0.77	0.139	0.65
ET [27]	1.3×10^{-3}	2.0×10^{-3}	0.79	0.134	0.69
CART-Net [28]	1.5×10^{-3}	2.1×10^{-3}	0.78	0.136	0.67
AttMoE [29]	1.8×10^{-3}	2.5×10^{-3}	0.75	0.142	0.63
DM-DC [34]	1.1×10^{-3}	1.8×10^{-3}	0.81	0.128	0.72
Q-TRACNet (Proposed)	$<10^{-4}$	$<10^{-4}$	0.89	0.102	0.91

Figure 10 presents complementary sensitivity analyses of the proposed framework and evaluation metric. Panel (a) shows that the learning rate, residual width, temporal window, and prediction horizon produce the largest changes in MAE, whereas dropout, batch size, and weight decay have comparatively smaller effects. This behavior indicates that Q-TRACNet is primarily influenced by its temporal representation and learning dynamics while remaining less sensitive to moderate changes in regularization settings. Panel (b) evaluates CDAI at normalized critical degradation thresholds of 10%, 15%, 20%, 25%, and 30%. Across the four datasets, the absolute CDAI variation remains limited to 0.008–0.010, while the default 20% threshold yields an EV-HLM-RUL CDAI of 0.91. The stable cross-threshold trends demonstrate that CDAI is not dependent on a narrowly selected end-of-life boundary and can provide consistent degradation-aware evaluation across datasets with different lifecycle scales.

**Figure 10.** Sensitivity analysis of Q-TRACNet and CDAI: (a) variation in RUL prediction error under different architectural and optimization settings and (b) robustness of CDAI across critical degradation thresholds from 10% to 30%.

The sensitivity of RUL prediction to the behavioral-index definitions was further examined by jointly scaling the acceleration, braking, speeding, and jerk thresholds by factors of 0.90, 1.00, and 1.10, representing relaxed, nominal, and strict event criteria, respectively. The resulting MAE values were 10.0, 9.8, and 9.9, while the corresponding RMSE values were 14.9, 14.7, and 14.8. The associated R^2 values were 0.978, 0.979, and 0.978, respectively. Relative to the nominal configuration, the maximum variations were 2.04% in MAE and 1.36% in RMSE. These limited changes indicate that moderate variation in behavioral-event thresholds does not materially affect RUL prediction accuracy, confirming that the model is not dependent on a single threshold configuration.

The experimental results show that Q-TRACNet achieves accurate and reliable forecasting results, which perform well across different degradation patterns. The framework maintains low regression error, strong rank consistency, and superior degradation awareness near end-of-life thresholds, as confirmed by CDAI and statistical significance analysis. The results of ablation studies demonstrate that temporal encoding, residual refinement, and quantum-inspired attentive modulation work together to enhance performance instead of producing separate architectural advantages. The model demonstrates improved predictive accuracy through the computational efficiency tests, which keep both inference latency and memory requirements at their original levels, making it ready for actual IoT-powered EV fleet operations. The predictive maintenance framework, which we developed, shows robust performance and scalable capabilities together with practical value for operational use.

5. Conclusions

This study proposed Q-TRACNet as a degradation-aware predictive maintenance framework for estimating Remaining Useful Life, Time-to-Failure, State-of-Health, and health conditions in IoT-enabled electric vehicle fleets. The methodology integrates causal maintenance-aware preprocessing, stress-sensitive temporal encoding, adaptive temporal condensation, residual refinement, learnable phase modulation, attention-based degradation prioritization, and class-weighted learning, while the hybrid Particle Swarm Optimization–Quantum-Guided Descent strategy combines global parameter exploration with stable local refinement. The Critical Degradation Awareness Index was also introduced to assess prediction reliability specifically near end-of-life conditions, where maintenance decisions are most critical. On the EV-HLM-RUL dataset, Q-TRACNet achieved an MAE of 9.8, an RMSE of 14.7, an R^2 of 0.979, and a CDAI of 0.91. Compared with the strongest competing baseline, the proposed method reduced RMSE by 20.11% and increased CDAI by 10.98%, while also achieving an NRMSE of 0.102 and a Kendall correlation of 0.89 with statistical significance of $p < 10^{-4}$. Cross-dataset evaluation on NASA CMAPSS, PHM 2012, and XJTU-SY demonstrated stable generalization across RUL, TTF, and short- and long-horizon SOH prediction tasks. The ablation results further showed that the complete framework reduced RMSE by 8.70% relative to the backbone-only configuration, confirming that maintenance-aware encoding, temporal condensation, residual refinement, and learnable phase attention provide complementary benefits. Sensitivity and robustness experiments also demonstrated stable performance under threshold variation, non-IID fleet heterogeneity, missing telemetry, noise, and device dropout. From a computational perspective, Q-TRACNet required 1.84 s of training time per epoch, 14.2 ms of inference latency per sample, and a 402 MB serialized memory footprint, corresponding to reductions of 2.13%, 7.79%, and 1.95%, respectively, relative to the strongest competing computational baselines. These findings demonstrate that Q-TRACNet achieves substantial improvements in predictive accuracy and critical-stage reliability without imposing additional computational burden, supporting its practical use for maintenance prioritization, fleet rotation, workshop scheduling, inspection planning, and

resource allocation through GPU-enabled fleet gateways, edge-computing platforms, and vehicle-management servers.

Future work will investigate federated and privacy-preserving learning across geographically distributed vehicle fleets. Adaptive uncertainty calibration and online updating will also be explored to address long-term operational drift. In addition, pruning, quantization, and knowledge distillation will be considered to support deployment on more resource-constrained embedded devices.

Author Contributions: Conceptualization, M.A.; methodology, M.A., J.A. and I.A.; software, M.A.; validation, M.A., J.A. and I.A.; formal analysis, M.A., J.A. and I.A.; investigation, M.A.; resources, M.A.; data curation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A., J.A. and I.A.; visualization, M.A.; supervision, M.A.; project administration, M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2026/01/41358).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study is publicly available at Zenodo <https://doi.org/10.5281/zenodo.18730983>.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
CART-Net	Causal Attention Remaining-Cycle Network
CDAI	Critical Degradation Awareness Index
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
cGAN-VR	Conditional Generative Adversarial Network for Voltage Reconstruction
CNN	Convolutional Neural Network
DM-DC	Denoising Model with Dilated Convolution
DS-ViT-ESA	Dual-Stream Vision Transformer with Efficient Self-Attention
ET	Exponential Transformer
EV	Electric Vehicle
EV-HLM-RUL	Electric Vehicle High-Level Maintenance Remaining Useful Life
GAN	Generative Adversarial Network
GPU	Graphics Processing Unit
HCMDT-Net	Hierarchical Convolution Multi-Channel Dense Transformer Network
HS-STT	Hierarchical Stochastic Spatial-Temporal Transformer
IoT	Internet of Things
KP-PGML	Knee-Point Physics-Guided Machine Learning
LS-SA-T	LogSparse Self-Attention Transformer
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MB	Megabyte
MTE	Maintenance-Aware Temporal Encoding
NASA	National Aeronautics and Space Administration
NRMSE	Normalized Root Mean Squared Error
PFFN	Parallel Feature Fusion Network

PHM	Prognostics and Health Management
PSO	Particle Swarm Optimization
PyTorch	Python Torch
Q-TRACNet	Quantum-Temporal Residual Attentive Classification Network
Q-Att. Mod.	Quantum-Attentive Modulation
QGD	Quantum-Guided Descent
RMSE	Root Mean Squared Error
RPM	Revolutions Per Minute
RUL	Remaining Useful Life
SOH	State of Health
TE	Transformer Ensemble
Temp. Cond.	Temporal Condensation
TS	Time-Series
TTF	Time-to-Failure
XJTU-SY	Xi'an Jiaotong University-SY Bearing Dataset

References

1. Della Libera, L.; Andreoli, J.; Dalle Pezze, D.; Ravanelli, M.; Susto, G.A. Bayesian deep learning for remaining useful life estimation via Stein variational gradient descent. *IEEE Trans. Autom. Sci. Eng.* **2025**, *23*, 600–611.
2. Wu, J.; Du, J.; Yang, C.; Yang, C.; Zhang, M. Lithium-ion battery remaining useful life prediction based on CEEMDAN decomposition and frequency-division learning. *J. Inf. Intell.* **2025**, *in press*. [[CrossRef](#)]
3. Aldossary, M. Multi-layer fog-cloud architecture for optimizing the placement of IoT applications in smart cities. *Comput. Mater. Contin.* **2023**, *75*, 633–652. [[CrossRef](#)]
4. Iftikhar, M.; Shoaib, M.; Altaf, A.; Iqbal, F.; Villar, S.G.; Dzul López, L.A.; Ashraf, I. A deep learning approach to optimize remaining useful life prediction for Li-ion batteries. *Sci. Rep.* **2024**, *14*, 25838. [[CrossRef](#)] [[PubMed](#)]
5. Ekberg, J.; Sridharan, N.V.; Karim, R.; Atta, K.T. State estimation and remaining useful life prediction for lithium-ion batteries. *Cell Rep. Phys. Sci.* **2025**, *6*, 102854. [[CrossRef](#)]
6. Lopez-Salazar, C.; Ekwaro-Osire, S.; Dabetwar, S.; Alemayehu, F. A comprehensive framework for estimating the remaining useful life of Li-ion batteries under limited data conditions with no temporal identifier. *Reliab. Eng. Syst. Saf.* **2025**, *253*, 110517. [[CrossRef](#)]
7. Aldossary, M. Enhancing urban electric vehicle (EV) fleet management efficiency in smart cities: A predictive hybrid deep learning framework. *Smart Cities* **2024**, *7*, 3678–3704. [[CrossRef](#)]
8. Zhao, B.; Zhang, W.; Zhang, Y.; Zhang, C.; Zhang, C.; Zhang, J. Research on the remaining useful life prediction method for lithium-ion batteries by fusion of feature engineering and deep learning. *Appl. Energy* **2024**, *358*, 122325. [[CrossRef](#)]
9. Zhou, K.; Zhang, Z. Remaining useful life prediction of lithium-ion batteries based on data denoising and improved transformer. *J. Energy Storage* **2024**, *100*, 113749. [[CrossRef](#)]
10. Li, W.; Chen, J.; Chen, S.; Li, P.; Zhang, B.; Wang, M.; Yun, J. A comprehensive review of artificial intelligence-based algorithms for predicting the remaining useful life of equipment. *Sensors* **2025**, *25*, 4481. [[CrossRef](#)] [[PubMed](#)]
11. Hu, J.; Fu, P.; Wei, Z.; Huang, Y.; Early, J.; Fly, A.; Zhang, Y. Early prediction of lithium-ion battery degradation with a generative pre-trained transformer. *Nat. Commun.* **2025**, *17*, 126. [[CrossRef](#)] [[PubMed](#)]
12. Chen, S.; Fang, X. Remaining useful life prediction of lithium-ion batteries via spatial attention TLSTM and dilated CNN with evolutionary optimization. *Sci. Rep.* **2025**, *15*, 32340. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, J.; Huang, J.; Zhang, T.; He, E.; Wang, S.; Yao, L. Early Remaining Useful Life Prediction of Lithium-Ion Batteries Based on a Hybrid Machine Learning Method with Time Series Augmentation. *Sensors* **2026**, *26*, 1238. [[CrossRef](#)] [[PubMed](#)]
14. Dong, Z.; Yang, M.; Wang, J.; Wang, H.; Lai, C.S.; Ji, X. PFFN: A Parallel Feature Fusion Network for Remaining Useful Life Early Prediction of Lithium-Ion Battery. *IEEE Trans. Transp. Electr.* **2024**, *11*, 2696–2706.
15. Liu, Y.; Ahmed, M.; Feng, J.; Mao, Z.; Chen, Z. Deep Learning-Powered Lifetime Prediction for Lithium-Ion Batteries Based on Small Amounts of Charging Cycles. *IEEE Trans. Transp. Electr.* **2024**, *11*, 3078–3090.
16. Aldossary, M. Q-MobiGraphNet: Quantum-inspired multimodal IoT and UAV data fusion for coastal vulnerability and solar farm resilience. *Mathematics* **2025**, *13*, 3051. [[CrossRef](#)]
17. Luo, Y.; Zhu, M.; Chen, T.; Zheng, Z. Remaining useful life prediction for stratospheric airships based on a channel and temporal attention network. *Commun. Nonlinear Sci. Numer. Simul.* **2025**, *143*, 108634. [[CrossRef](#)]
18. Liu, X.; Gao, Z.; Tian, J.; Wei, Z.; Fang, C.; Wang, P. State of Health Estimation for Lithium-Ion Batteries Using Voltage Curve Reconstruction by Conditional Generative Adversarial Network. *IEEE Trans. Transp. Electr.* **2024**, *10*, 10557–10567. [[CrossRef](#)]

19. Jia, X.; Zhang, C.; Li, Y.; Zou, C.; Wang, L.Y.; Cai, X. Knee-Point-Conscious Battery Aging Trajectory Prediction Based on Physics-Guided Machine Learning. *IEEE Trans. Transp. Electr.* **2023**, *10*, 1056–1069.
20. Al-Rahamneh, A.; Izco, I.; Serrano-Hernandez, A.; Faulin, J. Machine Learning-Based State-of-Health Estimation of Battery Management Systems Using Experimental and Simulation Data. *Mathematics* **2025**, *13*, 2247. [\[CrossRef\]](#)
21. Wei, M.; Yuan, S.; Chen, L.; Liu, Y.; Hu, J. State of Health Prediction of Pure Electric Vehicle Batteries Based on Patch Wavelet Transformer. *Mathematics* **2025**, *13*, 2099. [\[CrossRef\]](#)
22. Han, J.; Dong, Y.; Wang, W. Combined Framework for State of Charge Estimation of Lithium-Ion Batteries: Optimized LSTM Network Integrated with IAOA and AUKF. *Mathematics* **2025**, *13*, 2590. [\[CrossRef\]](#)
23. Zhao, Y.; Guo, H.; Guo, X. Lithium-Ion Battery State-of-Health Estimation Method Based on Hierarchical Convolution and Multi-Channel Interactive Dense Transformer Network. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 2534714. [\[CrossRef\]](#)
24. Ma, Y.; Li, J.; Hu, Y.; Chen, H. A Battery Prognostics and Health Management Technique Based on Knee Critical Interval and Linear-Complexity Self-Attention Transformer in Electric Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 10216–10230. [\[CrossRef\]](#)
25. Bai, J.; Zhu, W.; Liu, S.; Ye, C.; Zheng, P.; Wang, X. A Temporal Convolutional Network–Bidirectional Long Short-Term Memory prediction model for temporal faults in industrial equipment. *Appl. Sci.* **2025**, *15*, 1702. [\[CrossRef\]](#)
26. Shi, D.; Zhao, J.; Wang, Z.; Zhao, H.; Wang, J.; Lian, Y.; Burke, A.F. Spatial-Temporal Self-Attention Transformer Networks for Battery State of Charge Estimation. *Electronics* **2023**, *12*, 2598. [\[CrossRef\]](#)
27. Wang, C.; Bao, Z.; Lin, H.; He, Z.; Gao, M. An Exponential Transformer for Learning Interpretable Temporal Information in Remaining Useful Life Prediction of Lithium-Ion Battery. *IEEE Trans. Transp. Electr.* **2025**, *11*, 7945–7956. [\[CrossRef\]](#)
28. Zhang, Y.; Du, M.; Miao, Q. CART-Net: A Causal Adaptive Residual Time Network for Remaining Useful Life Prediction of Aero-Engines Under Varying Operating Conditions. *IEEE Internet Things J.* **2025**, *12*, 50291–50303. [\[CrossRef\]](#)
29. Zhao, L.; Song, S.; Wang, P.; Wang, C.; Wang, J.; Guo, M. An MLP-Mixer and Mixture-of-Expert Model for Remaining Useful Life Prediction of Lithium-Ion Batteries. *Front. Comput. Sci.* **2024**, *18*, 185329. [\[CrossRef\]](#)
30. Aldossary, M.; Alharbi, H.A.; Ayub, N. Exploring multi-task learning for forecasting energy-cost resource allocation in IoT-cloud systems. *Comput. Mater. Contin.* **2024**, *79*, 4603–4620. [\[CrossRef\]](#)
31. Niu, S.; Liu, W.; Liu, C.; Lyu, J.; Yang, C.; Chau, K.T. Wireless Power Transfer for IoT Bolts in Railways: Overcoming Coupling Challenges due to Conductive and Ferromagnetic Rails. *IEEE Trans. Ind. Electron.* **2026**, *73*, 5409–5419. [\[CrossRef\]](#)
32. Gao, J.; Wang, S.; Chen, R.; Zhang, C.; Zio, E.; Zhang, Y.; Lu, Y. ZSTT: A zero-shot time-series prediction model based on Transformer and its application to hydraulic actuator performance degradation prediction. *IEEE/ASME Trans. Mechatron.* **2025**, *31*, 2946–2957. [\[CrossRef\]](#)
33. Guo, D.; Zhang, Z.; Liu, J.; Zhang, J.; Lin, Y. Multi-horizon flight trajectory prediction enabled by time-frequency wavelet transform. *Nat. Commun.* **2025**, *17*, 633. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Liu, Z.; Du, X.; Shi, Y. Application of Multi-Modal Temporal Neural Network Based on Enhanced Sparrow Optimization in Lithium Battery Life Prediction. *Sci. Rep.* **2024**, *14*, 27476. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Guo, R.; Yi, J.; Luo, X. An efficient classification and error correction knowledge distillation framework for remaining useful life prediction of bearings in air turbine starter. *Neurocomputing* **2025**, *657*, 131611. [\[CrossRef\]](#)
36. Fraunhofer Institute for Transportation and Infrastructure Systems. *EV-IoT Fleet Predictive Maintenance Dataset, Version v1*; Zenodo: Geneva, Switzerland, 2025. [\[CrossRef\]](#)
37. Xue, B.; Yu, Y.; Fu, H.; Zhu, J.; Xu, Z. Advanced Prognostics Techniques for Machinery Under Time-Varying Operating Conditions. *Appl. Intell.* **2025**, *55*, 1066. [\[CrossRef\]](#)
38. Aldossary, M.; Alharbi, H.A.; Ayub, N. Optimizing electric vehicle (EV) charging with integrated renewable energy sources: A cloud-based forecasting approach for eco-sustainability. *Mathematics* **2024**, *12*, 2627. [\[CrossRef\]](#)
39. Alghamdi, E.A.; Alkinani, M.H.; Almazroi, A.A.; Alqarni, M.; Aldhahri, E.A.; Ayub, N. REST Network: An Ensemble Deep Learning Approach for EV Charging Load Forecasting in Artificial Port Supply Chains. *IEEE Access* **2025**, *13*, 133128–133144. [\[CrossRef\]](#)
40. Ren, Y.; Zhang, L.; Shi, P. Adaptive cooperative intra–inter domain adversarial network with stage division for RUL prediction of ship electric propulsion system. *IEEE Trans. Ind. Electron.* **2026**, *73*, 12477–12489. [\[CrossRef\]](#)
41. Ding, J.; Liu, Z.; Zheng, G.; Jin, H.; Kong, L. CondTSF: One-Line Plugin of Dataset Condensation for Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 128227–128259. [\[CrossRef\]](#)
42. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [\[CrossRef\]](#)
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
44. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.

45. Chauhan, D.; Suganthan, P.N. Learning Strategies for Particle Swarm Optimizer: A Critical Review and Performance Analysis. *Swarm Evol. Comput.* **2025**, *98*, 102048. [[CrossRef](#)]
46. Arjunan, T.; Arjunan, G.; Kumar, N.J. Optimizing the Quantum Circuit of Quantum K-Nearest Neighbors Using Hybrid Gradient Descent and Golden Eagle Optimization Algorithm. In Proceedings of the 2025 International Conference on Computing Technologies & Data Communication (ICCTDC), Hassan, India, 4–5 July 2025; pp. 1–7.
47. Hanselle, J.; Heid, S.; Fürnkranz, J.; Hüllermeier, E. Probabilistic Scoring Lists for Interpretable Machine Learning. *Mach. Learn.* **2025**, *114*, 55. [[CrossRef](#)]
48. de la Cruz Huayanay, A.; Bazán, J.L.; Russo, C.M. Performance of Evaluation Metrics for Classification in Imbalanced Data. *Comput. Stat.* **2025**, *40*, 1447–1473. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.