Current-Aware Transformer for Cross Life Stage Battery Aging Prediction in Second Life Applications

Yuou Chen, Guangyuan Ma, Xuan Zhang*
Tsinghua Shenzhen International Graduate School
Tsinghua University
Shenzhen, China

***Abstract*—Accurate prediction of battery aging under transitions in operating condition is essential for the efficient second life utilization of retired power batteries, as it enables reliable performance assessment and value recovery while ensuring safe deployment. However, this task is challenging due to the operating condition switch and limited early life data available for second life batteries. A cross life stage aging prediction framework for battery second life applications is proposed in this research. Through SHapley Additive exPlanations (SHAP) and Pearson correlation analysis, we optimize feature selection and develop a current-aware cross-attention Transformer (CACA-T) with current modulation for condition-adaptive prediction. Current-specific Gaussian Process Regression (GPR) models are employed for trajectory synthesis. Experimental results demonstrate that using only early data at the first life stage, our method accurately predicts aging trajectories with mean absolute percentage error (MAPE) around 1% under operating condition transitions in second life scenarios, enabling reliable technical support for retired battery valuation.**

***Index Terms*--Battery aging, Cross life stage prediction, Cross-attention Transformer, Second life batteries**

I. Introduction

The sustainable utilization of power batteries is vital to advancing a green circular economy amid the global transition toward cleaner energy systems. Cross life stage aging prediction that considers transitions in operating condition provides a scientific basis for second life repurposing and application matching, enabling value extension from first life to second life [1].

In pursuit of this goal, existing studies have conducted systematic investigations into battery lifetime prediction and cross-domain modeling. Research on degradation forecast primarily focuses on characterizing the evolution of battery performance. Existing approaches extract health indicators from voltage, current, and capacity to predict remaining useful life (RUL) and state of health (SOH) using temporal modeling and physics-informed constraints[2]. Early prediction studies estimate full-life trajectories from limited initial cycling data using transfer learning, domain adaptation, and uncertainty quantification[3]. Reference [4] presents a mechanism–data fusion model for ultra-early prediction of full-life battery capacity and energy trajectories. A joint multiple-output Gaussian process and prompt-learning framework was proposed in [3] for accurate early prediction of Li-ion battery capacity trajectories. Through an attention-based multisource sequence-to-sequence model, the study in [5] integrates historical state and future load information to achieve accurate and robust long-term SOH prediction. In cross-domain health assessment, research mainly aims to improve model predictive capability across different battery types and operating conditions, with some research leveraging deep feature extraction and cross-domain mapping to predict aging trends under diverse temperatures, currents, and charge-discharge protocols. To predict battery aging across types, the study in [6] employs voltage-capacity curve reconstruction with transfer learning and a Global Attention seq-to-seq model. The work in [7] develops an end-to-end cross-domain SOH estimation method, enabling robust battery health prediction across varying operating conditions.

Previous works have shown limitations in cross life stage aging prediction under operating condition transitions in second life applications. Most early prediction models focus on the current operating regime and neglect the diverse conditions in second life use. Moreover, existing cross-domain studies mainly improve generalization across different batteries, rather than modeling first-to-second life transitions within the same battery. Therefore, cross life stage aging prediction remains largely unexplored.

In this paper, we propose a cross life stage aging prediction framework for battery second life applications. Battery features are decomposed into intrinsic aging and external condition components, and feature selection is guided by SHapley Additive exPlanations (SHAP) analysis and correlation metrics. A current-aware cross-attention Transformer (CACA-T) with a current modulation module is developed for iterative operational feature prediction, enabling adaptation to varying current stresses. Finally, current-specific Gaussian Process Regression (GPR) models map predicted

This work was supported by Key Scientific Research Support Project of Shanxi Energy Internet Research Institute (No. SXEI2023A002) and Meituan Scholar Program-International Collaboration Project (No. 202209A). Corresponding Author: Xuan Zhang.

979-8-3315-8138-1

features to SOH for reliable estimation under different currents. Remarkably, the framework is trained solely on the early data in the first life stage yet generalizes effectively to diverse second life operating conditions.

The rest of the paper is structured as follows. Section II details the proposed methodology. Section III presents experimental results and performance analysis. Section IV concludes the study with future research directions.

II. METHODS

A. Problem Formulation

This work investigates cross life stage SOH prediction for batteries operating condition transitions. To effectively capture the gradual degradation patterns of batteries while reducing computational complexity, the original cycling data is discretized such that every 10 cycles constitute one time step. The operational data sequence of a battery in its first and second life can be denoted as:

$$D_1 = \left\{ \left(m_t, c_t \right) \right\}_{t=1}^{T_1}, \tag{1}$$

$$D_2 = \left\{ \left(m_t, c_t \right) \right\}_{t=1}^{T_2}, \tag{2}$$

where m_t represents the primary feature vector capturing intrinsic aging characteristics, c_t denotes the conditional feature vector reflecting external operating conditions. T_1 and T_2 represent the total number of time steps in the first and second life stages, respectively. t is the current time step.

The primary features comprise time features F0–F3 measured at 5.0, 5.5, 6.0, and 6.5 V during the discharge stage, incremental capacity (IC) curve features F4–F6 corresponding to the IC value at 6.0 V, the IC curve peak, and the voltage at the IC peak, and discharge capacity features F7–F10 measured at 5.0, 5.5, 6.0, and 6.5 V. The conditional features include the current and cycle index representing instantaneous load and operational progress, temperature-related features F11–F15 including maximum discharge temperature, time of maximum discharge temperature, the initial temperature rise rate, end of discharge temperature, and cumulative heat generation, and current variation features F16–F17 representing the maximum current change rate and the corresponding time.

Early stage operational data can be used to extrapolate the operational data in both the later first life and the second life. The cross life stage SOH prediction problem can then be formalized as learning a mapping function F such that:

$$\hat{y}_t = F\left(m_{e:T_e}, c_{e:T_e} \right), t = e, \ldots, T_2, \tag{3}$$

where $\hat{y}_t$ denotes the predicted SOH at time step t, extending from first life to second life operation. e is the prediction start.

The prediction performance is evaluated using the mean absolute percentage error (MAPE), defined as:

$$\mathrm{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{4}$$

B. Feature Selection Analysis

Here, we consider feature contribution, correlation, and prediction accuracy in this analysis. Feature contribution is quantified through SHAP analysis, while Pearson correlation coefficients assess statistical associations with SOH, enabling the identification of the most informative inputs for reliable prediction [8].

SHAP is derived from the Shapley value in game theory, assigning each feature an importance score that reflects its marginal contribution to the model output. For a given prediction model f and input $z = [m_t, c_t]$, the SHAP value of the i-th feature is:

$$\varphi_i\left(f, z \right) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\left(|N| - |S| - 1 \right)!}{|N|!} \left[f_{S \cup \{i\}}\left(z_{S \cup \{i\}} \right) - f_S\left(z_S \right) \right], \tag{5}$$

where N denotes the set of all features, S is any subset not containing feature i, and f_S is the model prediction using only subset S.

The global importance of feature i is measured by the mean absolute SHAP value over all samples:

$$SHAP_{global}\left(i \right) = \frac{1}{T_1} \sum_{t=1}^{T_1} \left| \varphi_i\left(f, z^{(t)} \right) \right|. \tag{6}$$

To further assess the statistical relationship between features and SOH, the Pearson correlation coefficient is computed for each feature time series $Z_i = \left\{ z_{i,t} \right\}_{t=1}^{T_1}$ and the SOH sequence $Y = \left\{ y_t \right\}_{t=1}^{T1}$:

$$\rho\left(Z_i, Y \right) = \frac{\sum_{t=1}^{T}\left(z_{i,t} - \bar{Z}_i \right)\left(y_t - \bar{Y} \right)}{\sqrt{\sum_{t=1}^{T_1}\left(z_{i,t} - \bar{Z}_i \right)^2 \sum_{t=1}^{T_1}\left(y_t - \bar{Y} \right)^2}}, \tag{7}$$

where $\bar{Z}_i$ and $\bar{Y}$ denote their respective means.

Considering SHAP values, Pearson correlation, and the iterative prediction performance of features, the feature structure is optimized to enhance both prediction accuracy and robustness.

C. Current-Aware Cross-Attention Transformer

CACA-T incorporates a current modulation mechanism that explicitly embeds operating current into the attention computation, enabling adaptive feature interactions under varying load conditions [9]. The network input consists of main features X_{main} and conditional features X_{cond}. The first dimension of conditional features is the current I_t. Inputs are projected to a hidden dimension d_h:

$$H_{main} = X_{main} W_m + b_m, \tag{8}$$

$$H_{cond} = X_{cond} W_c + b_c, \tag{9}$$

The operating current scalar is extracted from conditional features:

$$I = X_{cond}[:,:,0]. \tag{10}$$

It is passed through a small fully connected network to generate modulation parameters for query and key vectors. The modulation is split into query scale S_q, query shift T_q, key scale S_k, and key shift T_k, applied feature wise modulation to main and conditional features:

$$Q = H_{main} \odot (1+S_q)+T_q, \tag{11}$$

$$K = H_{cond} \odot (1+S_k)+T_k, \tag{12}$$

$$V = H_{cond}. \tag{13}$$

Cross-attention is computed as:

$$Attention(Q,K,V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_h}}\right)V. \tag{14}$$

D. Current-Specific Gaussian Process Regression Module

By constructing a GPR model library for different current conditions and combining a trend correction mechanism, the method ensures cross life stage predictions are both reasonable and reliable. d_g features are selected as GPR inputs [2]:

$$X_{gpr} = [m_{i_1}, m_{i_2}, ..., m_{i_{dg}}]. \tag{15}$$

The target variable output is the corresponding SOH:

$$y_{target} = [y_e, y_2, ..., y_{T_2}]. \tag{16}$$

For different current conditions, a GPR model set is built:

$$M = \{GPR_c | c \in C\}, \tag{17}$$

where c is the current value in the current set C.

Each GPR is a Gaussian process:

$$f_c(x) \sim \psi(m_c(x), k_c(x,x')), \tag{18}$$

where $m_c(x)$ is the mean function and $k_c(x,x')$ is the covariance function.

Hyperparameters θ_c are optimized:

$$\theta_c^* = \arg\max_{\theta_c} \log p(y_c | X_c, \theta_c). \tag{19}$$

For test battery b_{test} at switching point s:

$$\hat{y}_{pre,t} = GPR_{c_{pre}}(x_t), t \in [e,s], \tag{20}$$

$$\hat{y}_{post,t} = GPR_{c_{post}}(x_t), t \in [s+1, T_2], \tag{21}$$

$$\hat{y}_{full} = [\hat{y}_{pre,1}, ..., \hat{y}_{pre,s}, \hat{y}_{post,s+1}, ..., \hat{y}_{post,T_2}], \tag{22}$$

where $\hat{y}_{pre,t}$ is the predicted SOH before current switch and $\hat{y}_{post,t}$ is the predicted SOH after current switch at time t.

TABLE I. CURRENT SWITCH LOADING PROFILES

Module ID	2	12	24	32	33	53
1st Current (A)	16.0	16.0	16.0	16.0	16.0	14.3
2nd Current (A)	12.9	9.3	9.3	14.3	12.9	16.0

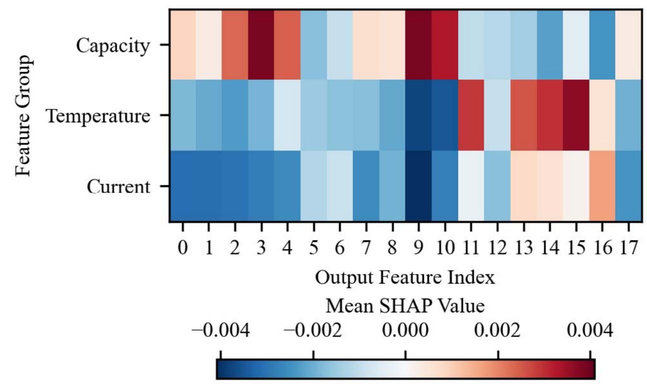

Figure 1. The three feature groups with the highest SHAP values.

III. EXPERIMENTS

A. Data Description

Two 18650 cells with 2.5 Ah rated capacity were assembled in series to form a battery module in the experiment. Six battery modules were tested in the research [10]. Charging profiles and resting times were defined as follows: charging at a constant current of 3 A, with a rest phase of 10 minutes before each charge and discharge phase. The first group contains eight batteries subjected to a single loading profile at 9.3 A, 12.9 A, 14.3 A, and 16.0 A. The second group contains six modules subjected to different loading types throughout their lifetimes, with the loading profile switching when SOH reached 0.86. Details of the loading profiles are shown in Table I. These second life battery packs were operated at less aggressive current levels compared to the loading conditions in their respective first lives, except for Module 53.

B. Feature Selection Results

Based on the analysis of feature contribution, correlation, and prediction accuracy, we propose a hypothesis for optimized feature selection and validate it through comparative experiments.

By SHAP analysis, we evaluated the importance of the input feature groups to the iterative prediction. The results reveal a significant polarization in the contribution of five feature groups to operational data prediction. Features in the capacity, temperature, and current groups show high absolute SHAP values, whereas features in the time and IC groups exhibit low values. Some features have SHAP values close to zero, indicating minimal contribution to model predictions and suggesting their potential removal in subsequent feature screening. However, it cannot be ruled out that features with higher contributions may pose a risk of steering the model away from accurate predictions. Fig. 1 visualizes the three feature groups with the highest SHAP values.

979-8-3315-8138-1

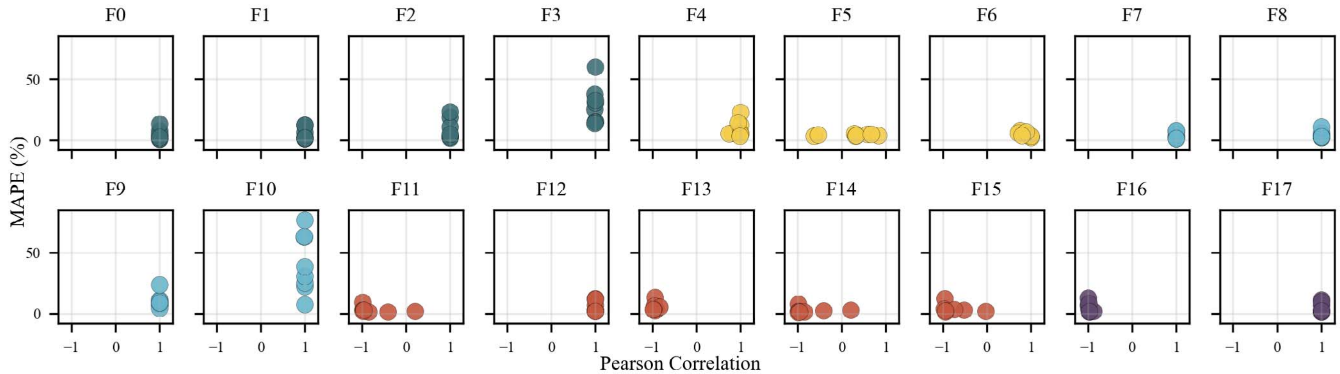

Figure 2. The distribution of prediction accuracy and Pearson correlation coefficients for different features.

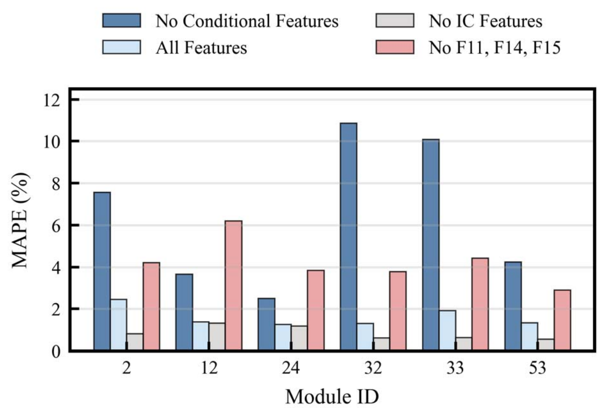

Figure 3. The prediction performance under different feature configurations.

Fig. 2 illustrates the distribution of prediction accuracy and Pearson correlation coefficients for different features. It is observed that although the absolute correlation coefficients of temperature features deviate somewhat from the ideal value of 1, their SHAP values remain high. This suggests that while these features may not contribute significantly to establishing a direct mapping to SOH, their presence considerably influences the prediction of feature outcomes in CACA-T. Among the primary features characterizing aging, both time and capacity features demonstrate high correlation with SOH and high prediction accuracy, making them possibly suitable as key inputs for the GPR mapping module. In contrast, IC features show more scattered correlations, reflecting unstable mapping capability to SOH and a comparatively lower contribution to the model's predictions. It is supposed that these features can be considered for removal to reduce computational cost and improve efficiency.

According to the above analysis, we propose removing IC features due to their low contribution and unstable correlation, while retaining time features, which exhibit relatively low contribution but high prediction accuracy and correlation. The primary features F0, F1, F7, and F8 demonstrate high prediction accuracy and strong correlation, making them well-suited for SOH mapping. Fig. 3 compares the prediction performance under different feature configurations. The results show that incorporating operating condition features significantly enhances prediction accuracy under condition transitions. Temperature features, which exhibit high contribution but relatively low correlation, positively influence model predictions, consistent with our initial hypothesis. Removing IC features improves model performance, albeit marginally, confirming that features with low SHAP values have limited impact on model predictions.

C. Comparison Results

Fig. 4 presents the SOH evolution curves of the batteries under cross life stage condition switching scenarios. Since GPR generally delivers slightly better predictive performance than Random Forest (RF) and both methods exhibit highly similar trends, only the GPR-based prediction curves are plotted for clarity. This indicates that the overall prediction performance is primarily determined by the effectiveness of the feature prediction stage. The comparison therefore focuses on evaluating the cross life stage feature prediction capability of CACA-T, Transformer, Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The complete error metrics for all eight methods are reported in Table II.

The results show that LSTM and GRU achieve good prediction accuracy in the first life stage but exhibit varying degrees of lag and slope deviation after the condition switch. Although both models possess strong temporal modeling ability, they lack modules that explicitly account for current variations and merely treat current as a regular input feature, making them insufficiently sensitive to long-sequence, cross-condition dynamics. In particular, for module 24, LSTM predicts accurately in the first life stage but deviates significantly afterward, indicating a lack of adaptive response to the current switch. Transformer incorporates current as a conditional input but still shows noticeable deviations near the switching point and struggles to capture the abrupt change in degradation rate caused by the current shift. In contrast, the proposed CACA-T closely follows the true SOH trajectory across the remaining lifetime, achieving smooth transitions at the switching point without jumps and quickly adapting the degradation slope to reflect the new operating conditions.

It is concluded that CACA-T not only adapts to abrupt operating condition transitions but also maintains robustness over long prediction horizons. These advantages collectively

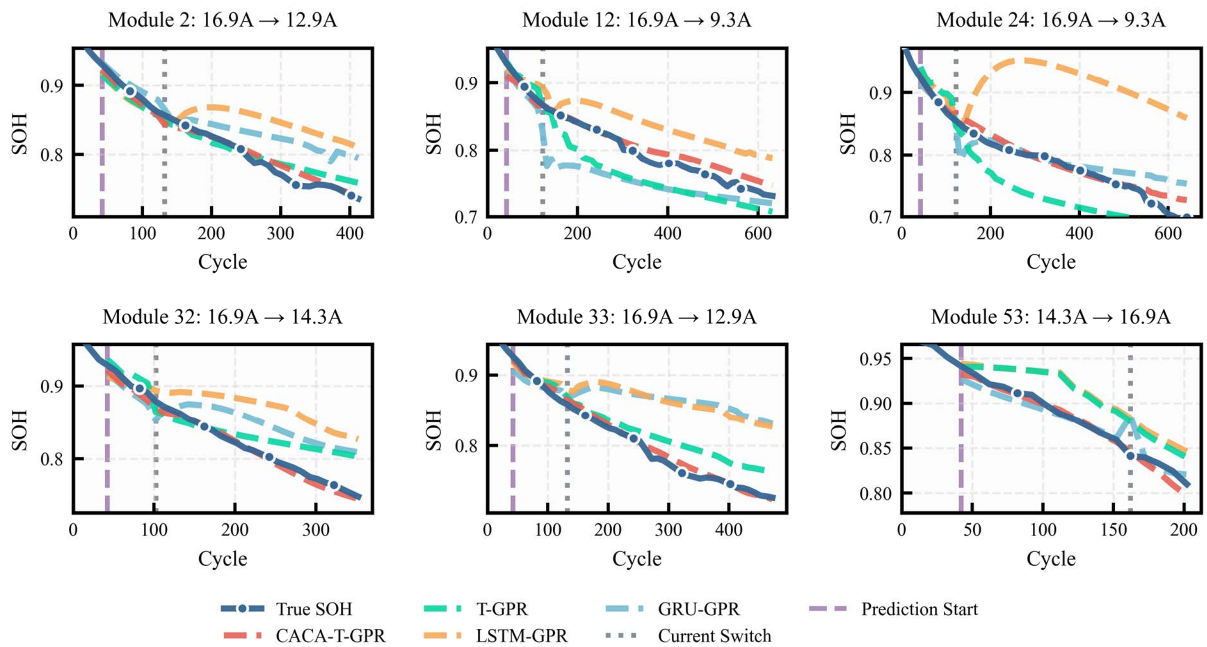

Figure 4. Comparison results using different prediction methods.

TABLE II. MAPE (%) UNDER DIFFERENT PREDICTION METHODS

Module ID	2	12	24	32	33	53
CACA-T-GPR	0.81	1.31	1.18	0.61	0.64	0.56
T-GPR	1.51	3.89	5.54	2.62	2.76	3.08
LSTM-GPR	5.88	5.10	15.8	6.21	8.42	3.35
GRU-GPR	3.84	4.70	2.35	4.14	8.59	0.97
CACA-T-RF	0.95	1.41	1.36	0.58	0.59	0.55
T-RF	1.43	3.38	4.23	2.80	2.18	3.02
LSTM-RF	4.15	3.38	12.2	5.76	6.21	3.43
GRU-RF	3.14	3.63	3.19	3.96	6.27	0.72

highlight its generalization capability for second life batteries where operating conditions differ significantly from those in the first life.

IV. Conclusion

This paper proposes a cross life stage aging prediction framework for second life batteries. Feature importance is quantified using SHAP and Pearson analysis to select an optimal feature set. A CACA-T model with current modulation is developed for condition-adaptive prediction, and current-specific GPR models are used for trajectory synthesis. The method accurately predicts aging curves under operating-condition transitions and achieves reliable trajectory prediction using only early first-life data. Future work will enhance adaptability to complex condition shifts, improve cross-chemistry generalization, and optimize computational efficiency for real-time engineering applications.

References

[1] S. Azizighalehsari, P. Venugopal, D. Pratap Singh, T. Batista Soeiro, and G. Rietveld, "Empowering Electric Vehicles Batteries: A Comprehensive Look at the Application and Challenges of Second-Life Batteries," Batteries, vol. 10, no. 5, 2024, doi: 10.3390/batteries10050161.

[2] Y. He, W. Bai, L. Wang, H. Wu, and M. Ding, "SOH estimation for lithium-ion batteries: An improved GPR optimization method based on the developed feature extraction," Journal of Energy Storage, vol. 83, 2024, doi: 10.1016/j.est.2024.110678.

[3] Z. Fei, F. Yang, K.-L. Tsui, L. Li, and Z. Zhang, "Early prediction of battery lifetime via a machine learning based framework," Energy, vol. 225, 2021, doi: 10.1016/j.energy.2021.120205.

[4] B. Cui et al., "Ultra-early prediction of lithium-ion battery performance using mechanism and data-driven fusion model," Applied Energy, vol. 353, 2024, doi: 10.1016/j.apenergy.2023.122080.

[5] C. Qian, B. Xu, Q. Xia, Y. Ren, B. Sun, and Z. Wang, "SOH prediction for Lithium-Ion batteries by using historical state and future load information with an AM-seq2seq model," Applied Energy, vol. 336, 2023, doi: 10.1016/j.apenergy.2023.120793.

[6] C. Zhang, J. Du, S. Li, L. Zhang, and W. Zhang, "Cross-domain machine transfer learning for capacity aging trajectory prediction of lithium-ion batteries," Journal of Power Sources, vol. 624, 2024, doi: 10.1016/j.jpowsour.2024.235534.

[7] J. Tian et al., "A Transfer Gated Recurrent Unit Model for Battery Health Monitoring Under Cross-Domain Conditions," IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1-12, 2025, doi: 10.1109/tim.2025.3548227.

[8] L. Antwarg, R. M. Miller, B. Shapira, and L. Rokach, "Explaining anomalies detected by autoencoders using Shapley Additive Explanations," Expert Systems with Applications, vol. 186, 2021, doi: 10.1016/j.eswa.2021.115736.

[9] H. Lin, X. Cheng, X. Wu, and D. Shen, "CAT: Cross Attention in Vision Transformer," presented at the 2022 IEEE International Conference on Multimedia and Expo (ICME), 2022.

[10] K. Fricke, R. Nascimento, M. Corbetta, C. Kulkarni, and F. Viana, "Accelerated Life Testing Dataset for Lithium-Ion Batteries with Constant and Variable Loading Conditions," International Journal of Prognostics and Health Management, vol. 14, no. 2, 2023, doi: 10.36001/ijphm.2023.v14i2.3587.

979-8-3315-8138-1