

Predicting Battery Health for Energy Storage and Electric Vehicles Systems by Integrating EMD Signal Processing and Machine Learning

Hafiz Sofiuddin, Mohd Abdul Talib Mat Yusoh, Kanendra Naidu and Nur Aina Fatini

Abstract—Accurate prognostics of battery State-of-Health (SOH) and Remaining Useful Life (RUL) are paramount for the operational safety and economic feasibility of sustainable energy systems, yet are frequently hindered by noise-corrupted sensor data. This study introduces and validates a novel hybrid framework that integrates Empirical Mode Decomposition (EMD) as an adaptive signal pre-processing technique with advanced machine learning models to overcome this critical limitation. Utilizing the NASA Ames prognostic dataset with synthetically introduced Gaussian noise to simulate real-world conditions, we demonstrate that EMD-based filtering effectively denoises battery discharge profiles, revealing a more coherent degradation trajectory. A comparative analysis of the resulting hybrid models SVM_EMD, LSTM_EMD, and GRU_EMD conclusively shows that the SVM_EMD model delivers superior performance, consistently achieving the lowest Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), and providing the most accurate RUL predictions across all tested battery units. This research establishes the two-stage SVM_EMD framework as a robust, low-complexity, and highly effective solution for enhancing the reliability and longevity of batteries in real-world applications, underscoring the vital importance of dedicated signal pre-processing in battery prognostics.

Keywords— Batteries, State-of-Health, Energy storage, RMSE, EMD.

I. INTRODUCTION

The global transition towards sustainable energy and transportation systems, a cornerstone of international frameworks like the United Nations' Sustainable Development Goals (SDGs), is fundamentally reliant on advanced energy storage technologies. Electric vehicles (EVs) and battery energy storage systems (BESS) are at the forefront of this paradigm shift, enabling decarbonization and enhancing grid stability. However, the operational viability and economic feasibility of these technologies hinge on the longevity and reliability of their core component: the battery. As batteries degrade through use,

their capacity to store and deliver energy diminishes, posing significant challenges to performance, safety, and system reliability [1, 2, 3]. Consequently, the ability to accurately predict battery health and RUL is paramount, aligning with circular economy principles by maximizing asset lifespan, minimizing electronic waste, and ensuring the sustainable deployment of clean energy solutions [4].

The health of a battery, quantified by metrics such as SOH, reflects its current condition relative to its original specifications, particularly its capacity and internal resistance [5]. Precise SOH estimation and RUL prediction the forecast of when a battery will no longer meet performance thresholds are critical for implementing proactive maintenance strategies, preventing catastrophic failures, and optimizing operational efficiency [6, 7]. Battery degradation is a complex, multifactorial process influenced by a confluence of operational and environmental stressors. These include charge/discharge protocols, depth of discharge (DoD), and operating temperatures, which can accelerate electrochemical decay and structural fatigue within the cell [8, 9, 10, 11]. The inherent nonlinearity of this degradation process, further complicated by fluctuating external conditions [12], renders accurate life prediction a formidable scientific and engineering challenge.

Historically, approaches to RUL prediction have relied on either empirical models or first-principle electrochemical models. Empirical methods, which employ curve-fitting on historical operational data, often lack the precision required for dynamic, real-world applications due to their inability to adapt to varying usage patterns [13, 14]. Conversely, electrochemical models, while providing high-fidelity simulations of internal cell dynamics, are computationally intensive and require extensive, often proprietary, knowledge of the battery's specific chemistry, limiting their practical implementation [15]. These limitations underscore the need for more sophisticated, data-driven techniques that can capture the complex temporal dependencies and stochastic nature of battery degradation.

To address these challenges, this study recognizes battery degradation as a complex time-series problem where the underlying health signals are frequently obscured by measurement noise and operational disturbances. The use of hybrid models, which combine signal processing techniques with machine learning algorithms, has gained traction. For instance, EMD has been successfully used to denoise sensor signals for fault diagnosis in rotating machinery [16]. More recently, the fusion of EMD with deep learning models like

This manuscript is submitted on 25th June 2025, revised on 11th August 2025, accepted on 12th September 2025 and published on 31st October 2025. Hafiz Sofiuddin, Mohd Abdul Talib Mat Yusoh, Kanendra Naidu and Nur Aina Fatini are from Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

*Corresponding author
Email address: mohd.abdul.talib@uitm.edu.my

1985-5389/© 2023 The Authors. Published by UiTM Press. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

LSTM has been explored for RUL prediction of bearings and other mechanical components [17]. However, a comprehensive analysis of EMD as a dedicated pre-processing step for various machine learning models (SVM, LSTM, and GRU) for battery SOH prediction, especially under noisy conditions, remains an area requiring deeper investigation. This research proposes a hybrid methodology to enhance the accuracy of RUL prediction by filling this gap. The primary objectives are twofold: (1) to develop a robust signal pre-processing technique using EMD to effectively denoise SOH data, and (2) to integrate this EMD based filtering with advanced machine learning models specifically Long Short-Term Memory (LSTM), Support Vector Machines (SVM), and Gated Recurrent Units (GRU) to build a more accurate and reliable prediction framework. By improving prognostic accuracy, this work aims to enhance the safety, reliability, and economic value of batteries in critical applications such as EVs and BESS.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive review of existing literature on SOH and RUL estimation techniques and their associated challenges. Section 3 details the proposed methodology, elaborating on the EMD filtering process and its integration with the selected machine learning models. Section 4 presents the experimental setup and a comparative analysis of the results, validating the performance of the proposed approach against established benchmarks. Finally, Section 5 concludes the paper, summarizing the key findings, acknowledging the study's limitations, and proposing directions for future research.

II. BATTERY DATASETS

The datasets used in this study were provided by the NASA Ames Research Center and comprise cyclic aging data for three lithium nickel cobalt aluminium oxide cells, designated as B005, B006, B007 [18]. These 18650-form factor lithium-ion cells were manufactured by Idaho National Laboratory and featured a rated capacity of 2.0 Ah and a nominal voltage of 3.7 V. The aging protocol involved controlled charging and discharging cycles. Charging was performed until the cell voltage reached 4.2 V, followed by constant voltage (CV) charging at 4.2 V until the current tapered below 20 mA. Discharging was conducted with end-of-discharge voltages set at 2.7 V (B005), 2.5 V (B006), and 2.2 V (B007). These variations in cutoff voltages were intended to induce different aging patterns among the cells.

The cycling was continued until each cell's capacity degraded to 70% of its nominal value. Throughout the testing, operational conditions such as temperature were controlled to minimize environmental influences. The datasets include detailed records of voltage, current, capacity, and temperature measurements across cycles, providing high-fidelity data suitable for modelling battery degradation and predicting RUL. Due to their comprehensive structure and experimental rigor, these datasets have been widely adopted in the literature for developing and validating prognostic health management (PHM) models, battery SOH estimation techniques, and

advanced battery management strategies.

III. METHODOLOGY

Fig. 1 illustrates a process for predicting the SoH of a battery using historical data from NASA (B05, B06, and B07) and EMD noise filtering. The process begins with loading historical battery data, followed by extracting relevant features from this dataset. To simulate real-world uncertainties, Gaussian noise is introduced into the extracted data. Subsequently, a Kalman filtering technique is applied to remove noise and enhance data quality before training the prediction model.

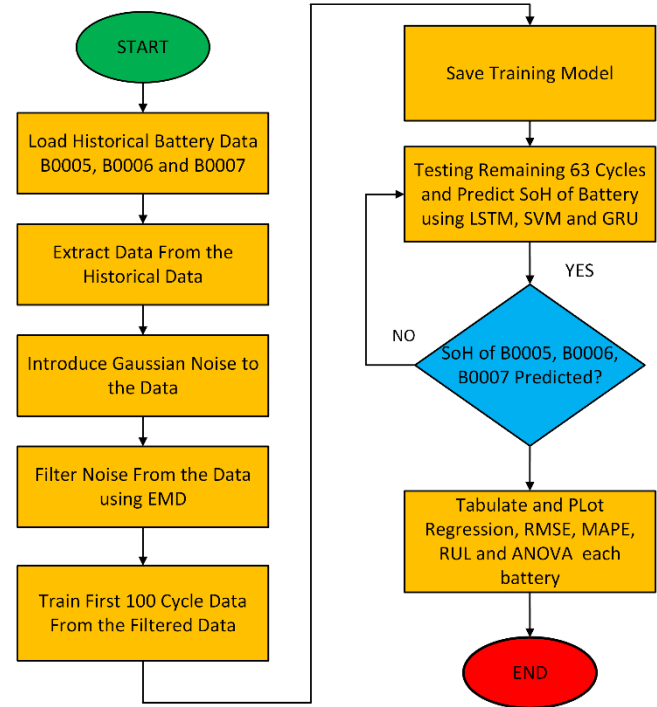


Fig. 1. Flowchart of Battery Prediction

Once the filtered data is prepared, it is used to train a model that predicts the SoH of the battery. A decision node checks whether the SoH of specific battery units (B05, B06, and B07) has been successfully predicted. If the prediction is incomplete, the process iterates; otherwise, it terminates. This structured approach ensures that the predictive model is trained on denoised and reliable data, improving the accuracy of battery health estimation.

A. Research Framework

To simulate real-world measurement uncertainties in battery data, Gaussian noise is introduced to the discharge cycle data. The function to introduce noise into first life of the battery which is first 100 cycle. Noise will be introduced to the measured current and voltage of each discharge cycle by adding a random noise component sampled from a Gaussian distribution with a 0.05 noise level. This process ensures that the dataset accounts for variability typically encountered in practical battery monitoring systems. The Gaussian Noise is computed using equation (1)

$$\mathfrak{x} = x + \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-u)^2}{2\sigma^2}} \quad (1)$$

Where;

- $\mathfrak{x}$ represents the noisy signal,
- x is the original signal,
- u mean of the noise,
- σ^2 variance of the noise.

In real-life battery systems, Gaussian noise represents measurement inaccuracies due to sensor imperfections, temperature fluctuations, and electronic interference. For instance, current and voltage sensors may exhibit drift or random fluctuations due to environmental factors, leading to minor deviations in recorded values. By incorporating Gaussian noise, the robustness of subsequent signal processing and state estimation techniques can be evaluated under realistic operating condition

B. Compute New Capacity

Battery capacity degradation is a key indicator of SoH. However, due to noise and measurement inconsistencies, raw capacity estimates may not accurately reflect the true battery degradation trend. Original battery capacity affected when noise introduced to the original data. By computing capacity from the filtered discharge data, a more reliable SoH estimate is obtained, reducing the impact of sensor errors and fluctuations in the dataset. The new capacity is computed using equation (2)

$$C = -\int I(t)dt \quad (2)$$

where $I(t)$ represents the discharge current as a function of time. In the implementation, the Trapezoidal Rule is used for numerical integration.

C. Empirical Mode Decomposition Filter

To mitigate the effects of noise introduced in the previous step, EMD is applied to the discharge cycle data. EMD is a data-driven, adaptive signal decomposition technique that iteratively separates a signal into intrinsic mode functions (IMFs) and a residual trend. Each voltage and current signal from the discharge cycles is decomposed using EMD into a finite set of IMFs representing oscillatory modes at distinct frequency scales. High-frequency IMFs, typically corresponding to noise, are identified and excluded from the reconstruction. The denoised signal is then obtained by summing the remaining IMFs and the residual component. This filtering operation is expressed in equation (3)

$$x(t) = \sum_{i=1}^N c_i(t) + r_N(t) \quad (3)$$

where $c_i(t)$ are IMFs and $r_N(t)$ is the residual. EMD is beneficial for removing noise from battery degradation signals, making it easier for LSTM to learn long-term dependencies.

Fig. 2. illustrates the EMD process. The EMD algorithm decomposes a nonlinear and non-stationary signal into a finite

set of Intrinsic Mode Functions (IMFs). It iteratively extracts oscillatory components by identifying local extrema, forming upper and lower envelopes via spline interpolation, and computing their mean. This mean is subtracted from the signal in a sifting process, repeated until an IMF is obtained. The residue is then used for further decomposition. This diagram summarizes each step involved in isolating IMFs with high noise from the original signal to compose become filtered signal.

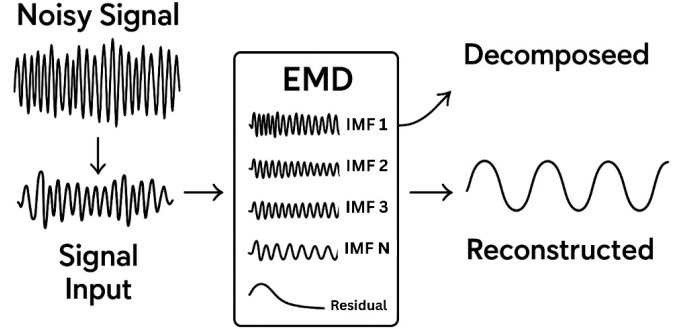


Fig. 2. EMD Filtering flow

EMD filtering is selected over conventional linear filtering approaches and even multiresolution techniques such as wavelet filtering due to its fully adaptive and data-driven nature. Unlike wavelet methods that require predefined basis functions, EMD directly extracts signal components based on the local extrema, making it particularly effective for processing non-linear and non-stationary signals such as battery discharge voltage and current profiles. This adaptability enables precise denoising while retaining critical transient features relevant to capacity estimation and battery health diagnostic.

IV. SoH PREDICTION TECHNIQUES

In this study, 100 initial discharge cycles were used for training, and the remaining 63 cycles served as the prediction dataset. The input features for prediction were the mean discharge voltage and mean discharge current extracted from each cycle. All input features and the target (State of Health, SoH) were normalized using z-score normalization based on the training set statistics and applied consistently to the test set. Model performance was validated using a temporal holdout method, where prediction was performed on remaining 63 cycles.

Z - score normalization equation;

$$Z = \frac{x - u}{\sigma} \quad (4)$$

Where;

- Z is the Z - score,
- x is the predicted value,
- u is the **mean** of the dataset
- σ is the standard deviation of dataset.

Furthermore, the performance of various machine learning models for battery parameter prediction (such as capacity degradation or remaining useful life (RUL)) was evaluated using statistical metrics such as Root Mean Square Error

(RMSE), Mean Absolute Percentage Error (MAPE), and Regression Equation analysis. These evaluation techniques serve as critical tools for quantitatively assessing the accuracy and reliability of the predictive models used.

➤ RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Where;

- y_i is the actual value,
- $\hat{y}_i$ is the predicted value,
- n is the number of data points.

RMSE is sensitive to large errors, as it squares the residuals before averaging. This makes it a suitable metric to penalize large deviations from the true battery health values. In the context of battery degradation prediction, a lower RMSE implies that the model is making predictions closer to the actual performance, especially in critical degradation stages. RMSE is particularly useful when large prediction errors are undesirable, as in EVs applications where battery health is safety critical.

➤ MAPE

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

Where;

- y_i is the actual value,
- $\hat{y}_i$ is the predicted value,
- n is the number of data points.

MAPE provides the average error as a percentage, making it easy to interpret. It is particularly useful in battery prediction scenarios because it allows comparison across datasets with different scales or capacities. A lower MAPE value indicates better model accuracy in predicting the battery behaviour over time. However, MAPE can be unstable when actual values are near zero, which must be considered during implementation.

➤ Regression Equation

$$Y_i = f(X_i, \beta) + e_i \quad (7)$$

Where;

- Y_i is the dependent variable,
- f is the function,
- X_i is the independent variable,
- β is the y-intercept.
- e_i is the error term or residual

The regression line ideally should have a slope $a \approx 1$ and intercept $b \approx 0$. These indicate that the model accurately captures the trend of the battery's actual performance. In battery health prediction, a high regression value supports the model's ability to generalize across unseen data, while poor regression alignment can indicate bias or underfitting. Data charts which are typically black and white, but sometimes include color.

A. SoH Estimation Using Support Vector Machine (SVM)

SVM is a supervised learning algorithm widely used for classification and regression tasks. In the context of battery health prognosis, SVM regression is particularly effective for

modelling nonlinear degradation patterns by projecting input features into a higher-dimensional feature space using a kernel function. This allows the model to capture complex relationships between extracted features (such as voltage time intervals) and the SoH of a battery.

After obtaining the denoised discharge cycle data, features such as the Time Interval for End-of-Discharge Voltage Drop (TIEDVD) and Time Interval for End-of-Charge Voltage Drop (TIECVD) are extracted.

$$TIEDVD = t_{3.6V} - t_{3.8V} \quad (8)$$

$$TIECVD = t_{4.2V} - t_{3.5V} \quad (9)$$

Where:

- $t_{3.6V}$ is the time when discharge voltage first reaches 3.6V,
- $t_{3.8V}$ is the time when discharge voltage first reaches 3.8V,
- $t_{4.2V}$ is the time when charge voltage first reaches 4.2V,
- $t_{3.5V}$ is the time when charge voltage first reaches 3.5V.

These features, along with the filtered capacity data, serve as input predictors for training an SVM regression model. The model is trained using 100 cycles of first-life battery data, and tested on the remaining 63 cycles, simulating a realistic future prediction scenario. The regression function learned by SVM is represented by:

$$f(x) = \omega^T \phi(x) + b \quad (10)$$

Where;

- x is the input feature vector,
- $\phi(x)$ is a transformation function,
- ω is the weight vector,
- b is the bias term.

In this study, the Gaussian Radial Basis Function (RBF) kernel was used, which is well-suited for nonlinear regression tasks such as SoH prediction. The Gaussian kernel enables the SVM model to learn complex, non-linear relationships between the features and the target output.

The Gaussian RBF kernel is defined as:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

Where;

- x_i, x_j are feature vectors from two different cycles,
- $\|x_i - x_j\|^2$ is the squared Euclidean distance between them,
- σ (or KernelScale) controls the kernel's spread.

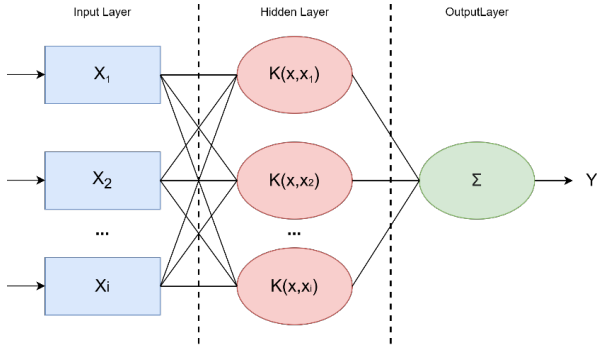


Fig. 3. Structure of the SVM Network

Fig. 3. shows the architecture of the SVM used for regression-based SoH estimation. The SVM model receives a feature vector comprising filtered capacity values and voltage time intervals. These features are mapped into a higher-dimensional space through a kernel function, enabling the model to capture nonlinear degradation trends in battery performance. The output is a scalar value representing the predicted SoH. The training process involves tuning hyperparameters such as the kernel scale, box constraint, and epsilon to optimize model generalization. This structure provides a balance between complexity and accuracy, allowing the model to learn degradation behaviours even with limited training data.

B. SoH Estimation Using Long Short-Term Memory (LSTM)

In battery prediction, LSTMs are widely used for forecasting the RUL and SOH based on historical sensor data, such as voltage, current, and temperature. The model processes past battery performance data in a sequential manner, capturing long-term dependencies to predict future degradation trends. During training, the LSTM learns complex temporal relationships by minimizing a loss function, typically Mean Squared Error (MSE), using backpropagation through time (BPTT). Once trained, the model can accurately estimate battery aging and failure points, helping optimize energy management systems and predictive maintenance strategies. Variants like bidirectional LSTMs and attention mechanisms further enhance predictive accuracy by considering dependencies in both past and future time steps.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (12.1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (12.2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (12.3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (12.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12.5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (12.6)$$

Where;

- f_t, i_t, o_t are forget, input, and output gates.
- C_t is the cell state.
- h_t is the hidden state.
- W_f, W_i, W_c, W_o and b_f, b_i, b_c, b_o are weights and biases.

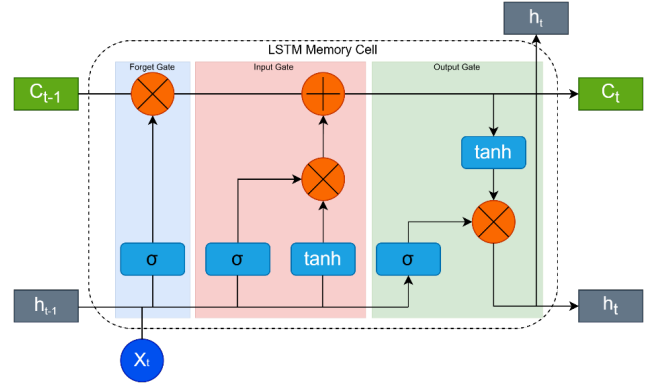


Fig. 4. Structure of the LSTM Network

Fig. 4. presents the LSTM network used for sequential modelling of battery data. The LSTM processes time-series inputs voltage, current, and temperature capturing both short-term fluctuations and long-term degradation patterns. The internal architecture consists of input, forget, and output gates that regulate the information flow, and a memory cell that retains relevant features across cycles. During training, the network minimizes prediction error using BPTT. This memory-enhanced structure is particularly effective for applications where battery degradation evolves gradually over long operational periods.

C. SoH Estimation Using Gated Recurrent Units (GRU)

In battery prediction tasks, GRUs are commonly employed to forecast the RUL and SOH based on historical sensor data, including voltage, current, and temperature. GRUs process battery performance data in a sequential manner, capturing temporal dependencies without the need for a separate cell state, as used in LSTMs. Instead, GRUs rely on an update gate and reset gate to regulate the flow of information. During training, the GRU model learns complex temporal patterns by minimizing a loss function typically Mean Squared Error (MSE) through BPTT. Once trained, it can effectively estimate battery aging and predict failure events, contributing to optimized energy management systems and predictive maintenance. Enhancements such as bidirectional GRUs and attention mechanisms further improve performance by allowing the model to consider both past and future dependencies.

The GRU operates according to the following equations:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (13.1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (13.2)$$

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \cdot h_{t-1}, x_t] + b_h) \quad (13.3)$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \quad (13.4)$$

Where;

- z_t, r_t are the update and reset gates.
- $\tilde{h}_t$ is the cell state.
- h_t is the hidden state.
- W_z, W_r, W_h and b_z, b_r, b_h are the model weights and biases.

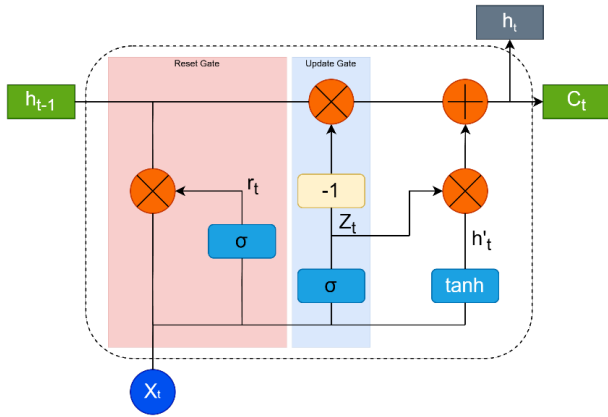


Fig. 5. Structure of the GRU Network

Fig. 5. illustrates the architecture of the GRU model, an alternative to LSTM with reduced complexity. Unlike LSTM, GRU does not maintain a separate memory cell; instead, it utilizes update and reset gates to manage the internal hidden state. This architecture enables the model to learn temporal dependencies with fewer parameters and computational cost. GRUs are particularly well-suited for real-time implementations where efficiency is critical, yet the model still needs to track degradation over time with reasonable precision.

V. RESULTS AND DISCUSSION

Fig. 6(a) shows the original discharge cycle of battery B0005 from the NASA dataset, representing the baseline voltage response during a standard discharge event. The curve captures the characteristic voltage drop over time as the battery depletes its stored energy. This signal serves as the unaltered reference for analysing the impact of signal processing and machine learning algorithms applied in subsequent stages of the study.

Fig. 6(b) presents the original discharge cycle for battery B0006, recorded under similar conditions. The voltage profile demonstrates the natural degradation behaviour associated with aging lithium-ion cells. Like battery B0005, this signal remains unprocessed and free of synthetic noise, preserving the integrity of the raw dataset used for performance benchmarking.

Fig. 6(c) depicts the original discharge cycle of battery B0007. This curve, along with those of B0005 and B0006, contributes to establishing a representative dataset of real-world battery degradation without any artificial disturbances. These original signals form the foundation for evaluating the effectiveness of subsequent noise filtering and state-of-health prediction techniques applied across different battery units.

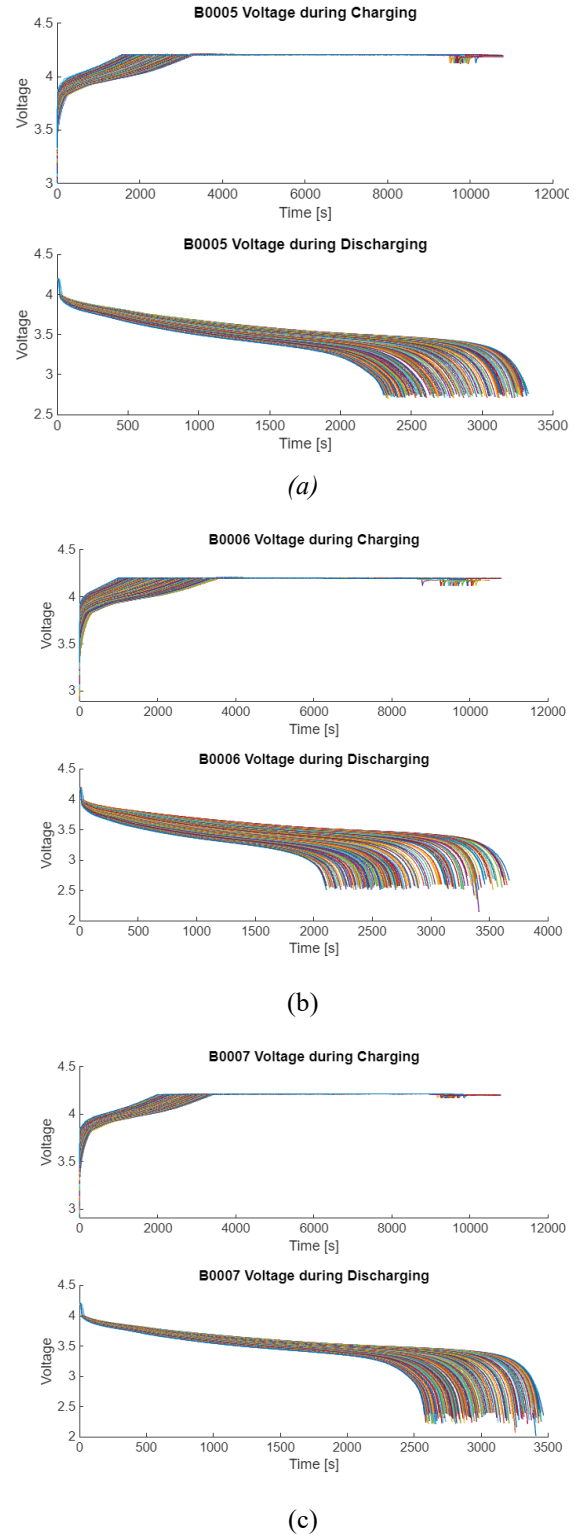


Fig. 6. a) Original B05 battery discharge cycle from NASA b) Original B06 battery discharge cycle from NASA c) Original B07 battery discharge cycle from NASA dataset

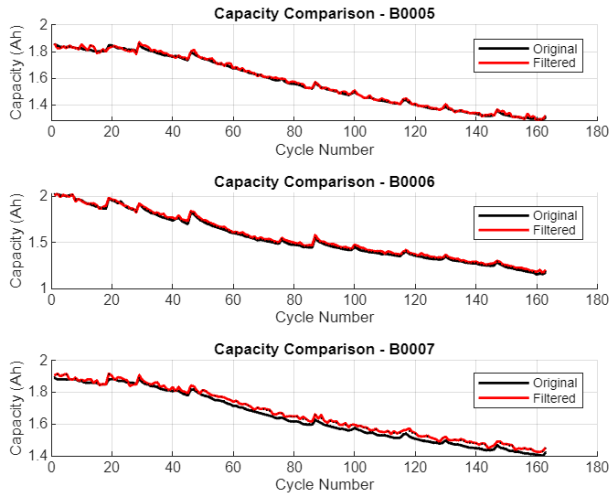
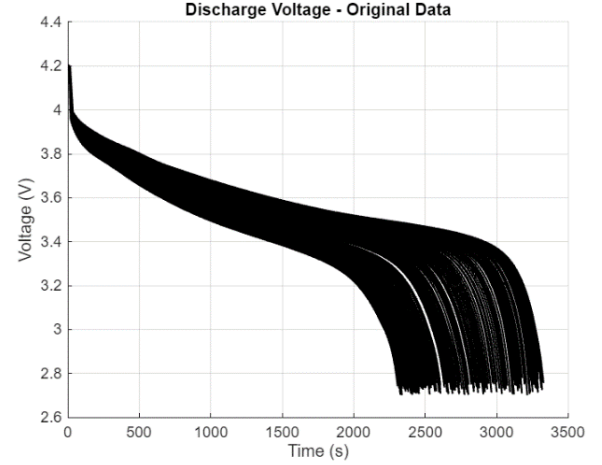


Fig. 7. Original and filtered capacity

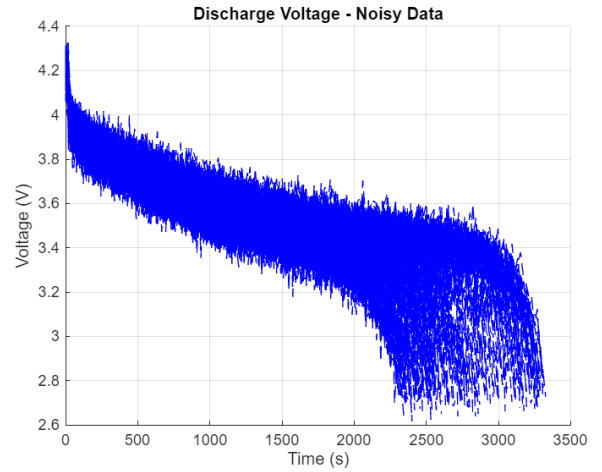
Fig. 7 shows the capacity profiles before and after applying EMD filtering. The unfiltered trajectory, derived from noisy voltage and current data, exhibits irregular fluctuations that are inconsistent with typical battery degradation trends. These deviations stem from synthetic noise introduced during pre-processing and underscore the limitations of relying on raw data for accurate SoH estimation. In contrast, the EMD-filtered profile demonstrates a smoother and more coherent degradation pattern, aligning more closely with expected aging behaviour. This denoised capacity signal is subsequently used as a key input feature in machine learning-based SoH prediction models.

Fig. 8(a) reaffirms the original discharge cycle of battery B0005 from the NASA dataset, providing a baseline signal prior to any noise intervention. The purpose of this repetition is to validate the consistency of the data across multiple analysis stages. Fig. 8(b) shows the same B0005 discharge cycle after the addition of Gaussian noise. The introduced distortion mimics real-life conditions such as sensor drift and temperature-induced signal perturbation, providing a realistic test case for filtering and prediction algorithms. Fig. 8(c) demonstrates the denoised version of the B0005 discharge cycle. The signal, processed using EMD filtering, shows a significant reduction in noise while maintaining the essential discharge characteristics. This confirms the effectiveness of EMD filtering in preparing the data for robust feature extraction.

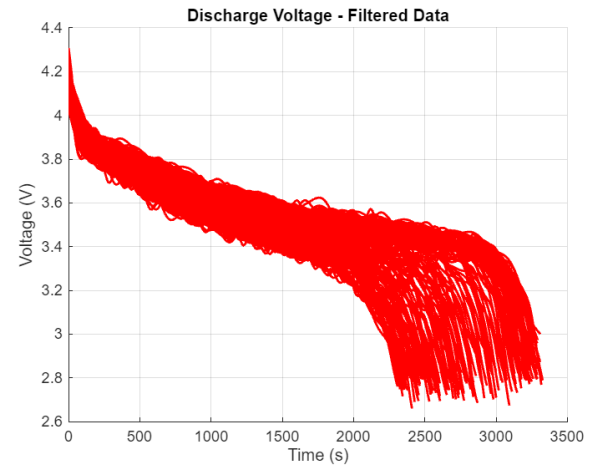
The effectiveness of EMD filtering lies in its adaptive, data-driven approach. Unlike conventional filters that use a fixed basis function, EMD decomposes the signal into a series of Intrinsic Mode Functions (IMFs) based on the local characteristics of the data itself. This allows it to separate the high-frequency measurement noise (captured in the initial IMFs) from the underlying, slow-moving battery degradation trend (captured in later IMFs and the final residual). By reconstructing the signal without the noise-related IMFs, we obtain a cleaner, more stable representation of the battery's health, which serves as a superior input for the subsequent machine learning models.



(a)



(b)



(c)

Fig. 8.: a) Original B05 battery discharge cycle from NASA b) Battery discharge cycle after introduction of noise in B05 battery c) Discharge cycle of B05 battery after denoising

Fig. 9. presents a comparative analysis of predicted and actual SoH trajectories for batteries B0005, B0006, and B0007 using three hybrid models: LSTM_EMD, SVM_EMD, and GRU_EMD. In Fig. 9(a), the LSTM_EMD model demonstrates strong predictive performance, with predicted SoH curves closely tracking the actual degradation patterns across all three batteries. Fig. 9(b) shows similar results for the SVM_EMD model, where the predicted curves align well with the true SoH profiles, indicating effective modelling of degradation dynamics. In Fig. 9(c), the GRU_EMD model also exhibits high fidelity in capturing the temporal evolution of battery aging. Although minor deviations appear during later degradation stages, particularly in more abrupt capacity declines, the overall trend alignment confirms the model's robustness and generalization across battery units. Collectively, these results highlight the effectiveness of EMD-based pre-processing in enhancing the predictive accuracy of data-driven SoH estimation models.

Fig. 10(a) displays a regression plot for battery B0005, comparing predicted versus actual SoH values. Most data points lie close to the identity line, reflecting high predictive accuracy and low deviation in the SoH estimates. Fig. 10(b) presents the regression plot for battery B0006. While the general correlation between predicted and actual SoH remains strong, a slight increase in variance is observed compared to B0005, suggesting slightly reduced model stability for this cell. Fig. 10(c) shows the regression analysis for battery B0007. The distribution of points remains closely aligned with the identity line, indicating that the model effectively generalizes across varying battery conditions.

Fig. 11(a) shows the regression plot for battery B0005 using the LSTM_EMD model. Increased scatter around the diagonal line highlights the model's reduced predictive precision. Fig. 11(b) presents regression results for battery B0006. The dispersion of data points indicates less consistent performance and a tendency toward higher prediction variance. Fig. 11(c) shows the regression plot for battery B0007. The data points exhibit considerable deviation from the identity line, particularly at low SoH values, suggesting the model struggles with capturing nonlinear degradation behaviour.

Fig. 12(a) displays the regression analysis for battery B0005 using the GRU_EMD model. The prediction points are generally well-aligned with the diagonal, supporting moderate accuracy. Fig. 12(b) shows the regression plot for battery B0006. The increased spread of points suggests a less precise relationship between input features and predicted SoH. Fig. 12(c) presents regression results for battery B0007. The model shows improved predictive stability compared to LSTM, although deviations from the identity line persist.

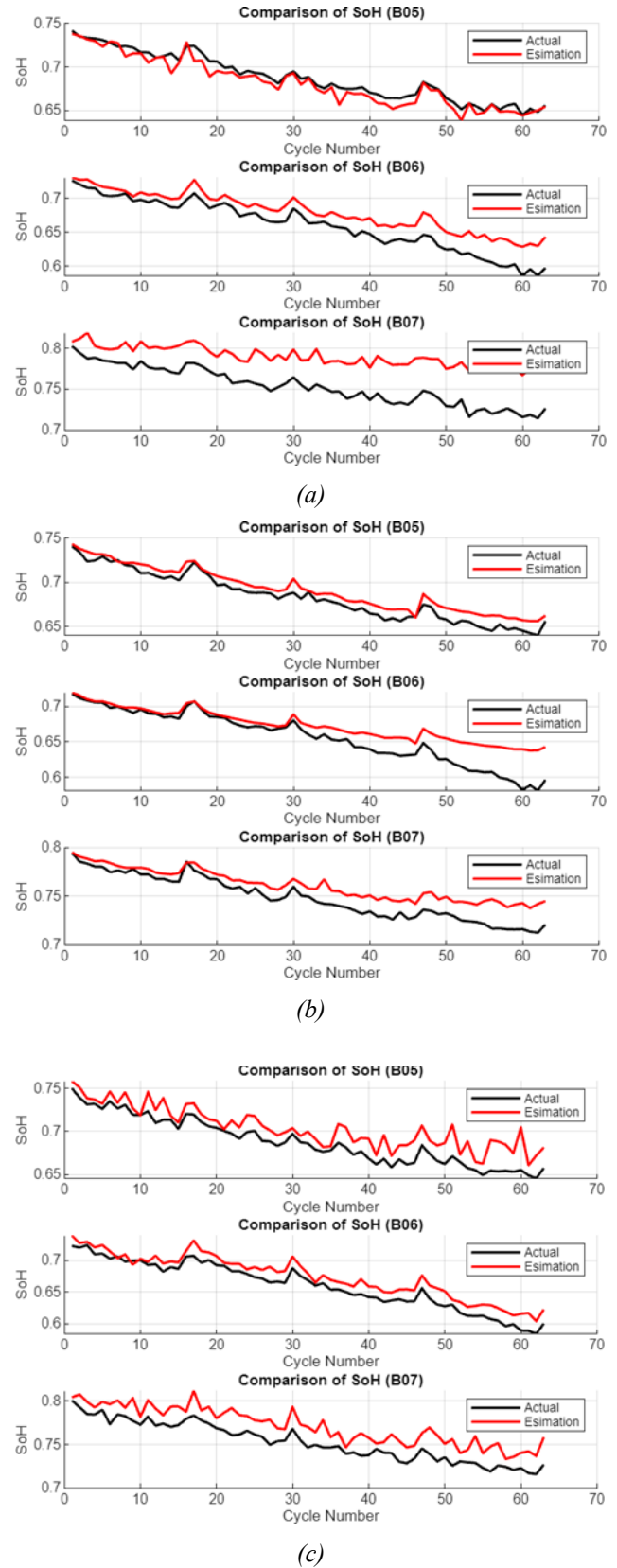
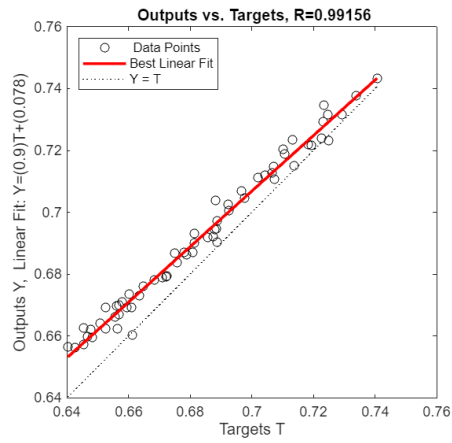
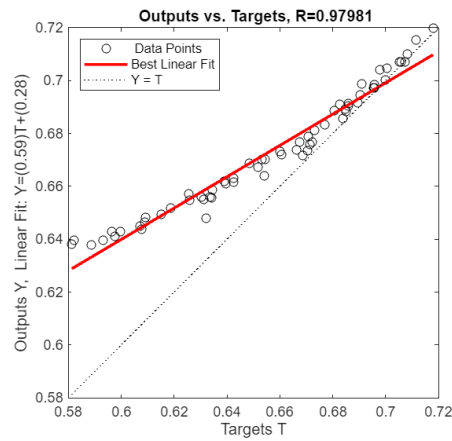


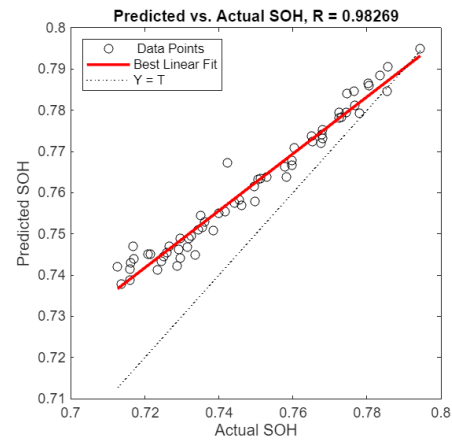
Fig. 9. SoH Comparison a) LSTM_EMD, b) SVM_EMD, c) GRU_EMD



(a)

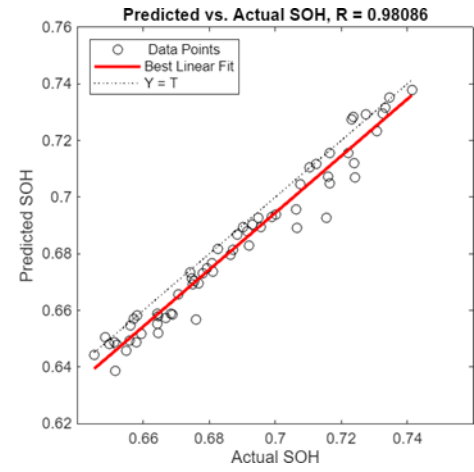


(b)

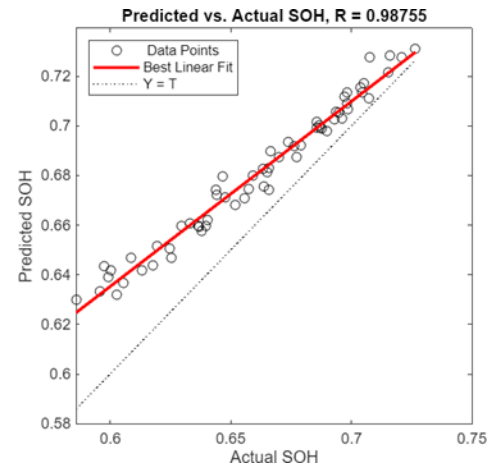


(c)

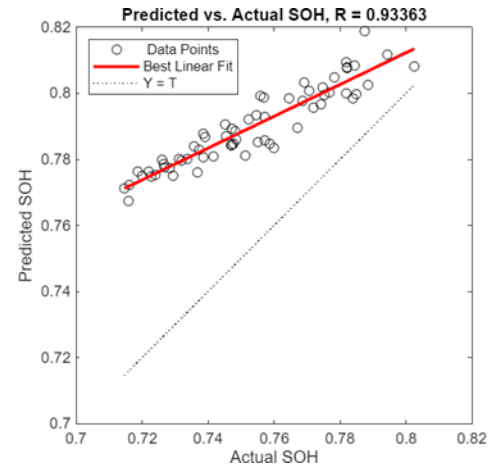
Fig. 10. Regression Analysis SVM_EMD a) B05 b) B06 c) B07



(a)



(b)



(c)

Fig. 11. Regression Analysis LSTM_EMD a) B05 b) B06 c) B07

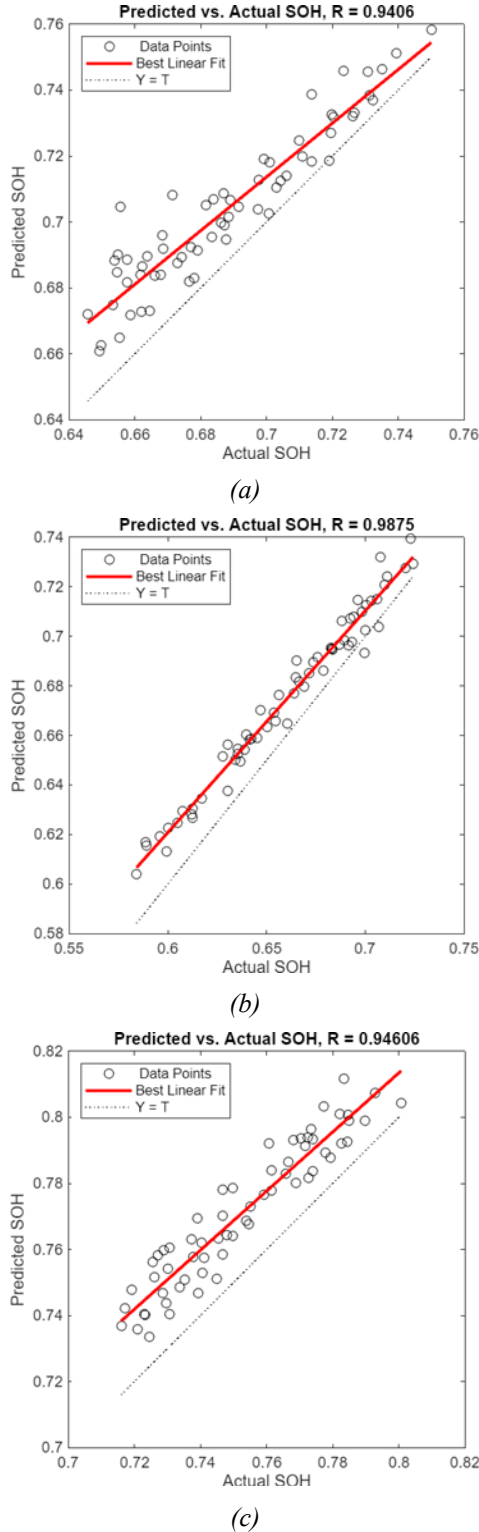


Fig. 12: Regression Analysis GRU_EMD a) B05 b) B06 c) B07

TABLE I. RMSE, MAPE, RUL AND REGRESSION MACHINE LEARNING_EMD

SVM_EMD					
BattName	RMS E	MAP E (%)	Tru e RU L	Predicted RUL	Regression
"B0005"	0.0095	1.2806	44	40	0.99156
"B0006"	0.0232	2.7429	58	56	0.97981
"B0007"	0.0152	1.7926	63	63	0.98269
LSTM_EMD					
BattName	RMS E	MAP E (%)	Tru e RU L	Predicted RUL	Regression
"B0005"	0.0077	0.9245	43	44	0.97737
"B0006"	0.0288	3.8553	55	42	0.98162
"B0007"	0.0283	3.4790	63	63	0.95767
GRU_EMD					
BattName	RMS E	MAP E (%)	Tru e RU L	Predicted RUL	Regression
"B0005"	0.0263	3.5607	43	31	0.97810
"B0006"	0.0236	3.1371	54	52	0.98399
"B0007"	0.0456	5.7281	63	63	0.94311

Table 1 presents a comparative evaluation of three hybrid models SVM_EMD, LSTM_EMD, and GRU_EMD based on Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), RUL prediction accuracy, and regression scores for batteries B0005, B0006, and B0007. These metrics collectively assess the model performance in predicting the SoH and estimating battery life span.

For battery B0005, the LSTM_EMD model achieves the lowest RMSE (0.0077) and MAPE (0.9245 %), with a predicted RUL of 44 cycles, closely matching the true RUL of 43. In comparison, the SVM_EMD model also performs well with slightly higher RMSE (0.0095) and MAPE (1.2806%) and a higher regression score of 0.99156. The GRU_EMD model exhibits the highest error and underpredicts RUL significantly at 31 cycles.

For battery B0006, the SVM_EMD model again shows strong performance with low RMSE (0.0232) and MAPE (3.8553%). Its predicted RUL (56) is very close to the actual value (58). The GRU_EMD model provides a slightly lower RMSE than SVM_EMD but a less accurate RUL prediction.

For battery B0007, all models accurately predict the RUL at 63 cycles. However, the SVM_EMD model yields the lowest error metrics (RMSE: 0.0152; MAPE: 1.7926%) and the highest regression score (0.98269). The GRU_EMD model exhibits the highest error values, indicating weaker performance.

Among the three models, SVM_EMD consistently demonstrates superior or highly competitive performance across all battery units and metrics. This result is significant because SVM is often considered a less complex model than deep learning architectures like LSTM and GRU.

The outperformance of SVM_EMD can be attributed to several factors: (1) data Efficiency and Robustness: The training dataset consists of 100 cycles, which is relatively small for training complex deep learning models from scratch. SVMs, particularly with the Radial Basis Function (RBF) kernel, are highly effective at finding non-linear relationships even with limited data. They are less prone to overfitting on smaller datasets compared to LSTMs and GRUs, which have a much larger number of trainable parameters, (2) problem Formulation: After EMD filtering and feature extraction (e.g., mean discharge voltage), the prediction task can be viewed as a non-linear regression problem mapping a feature vector to an SoH value for each cycle. SVMs are exceptionally well-suited for this type of mapping. While battery degradation is inherently a temporal process, the recurrent nature of LSTMs and GRUs may be "overkill" if the extracted features already encapsulate the necessary temporal information for a single-step prediction, and (3) Simplicity and Generalization: The relative simplicity of the SVM architecture contributes to its strong generalization capabilities. LSTMs and GRUs are designed to capture long-range dependencies in sequential data. However, in this noisy, filtered dataset, they might struggle to distinguish between the true, subtle degradation pattern and residual noise artifacts, potentially leading to less stable predictions. The SVM, by focusing on finding an optimal hyperplane in a high-dimensional feature space, provides a more robust solution that is less sensitive to minor fluctuations in the input sequence.

Therefore, the combination of EMD's powerful denoising and the SVM's efficient and robust regression capability makes the SVM_EMD framework the most reliable method for accurate battery SoH and RUL prediction in this study.

VI. CONCLUSION

Based on the observed performance gains, it is recommended that future battery health prediction frameworks incorporate a dedicated noise filtering stage such as EMD filtering prior to regression modelling. This approach demonstrably enhances prediction accuracy by reducing the impact of signal noise, a common issue in real-world battery monitoring systems. Specifically, for applications in electric vehicles, grid storage,

or portable electronics, where sensor data is often noisy and incomplete, the two-stage SVM_EMD framework offers a low-complexity yet highly effective solution. Additionally, further research should investigate the integration of adaptive or nonlinear filtering techniques, such as Extended Kalman Filters (EKF) or Particle Filters, to handle non-Gaussian noise and more complex degradation behaviours. Exploring real-time deployment scenarios and evaluating the model's robustness across different battery chemistries and usage profiles would also help validate its generalizability. Finally, combining the SVM_EMD model with lightweight edge-computing solutions could support real-time health monitoring in embedded systems, extending the method's utility in industrial and consumer-grade applications.

Overall, the integration of signal processing and machine learning models offers a robust approach to battery prognostics. Future work should focus on incorporating larger and more diverse datasets, exploring attention mechanisms, and validating across different battery chemistries to enhance model generalizability and deployment feasibility.

REFERENCES

- [1] A. Thelen, X. Huan, N. Paulson, S. Onori, Z. Hu, and C. Hu, "Probabilistic machine learning for battery health diagnostics and prognostics—review and perspectives," 2024.
- [2] H. Yang, J. Hong, F. Liang, and X. Xu, "Machine learning-based state of health prediction for battery systems in real-world electric vehicles," *Journal of Energy Storage*, vol. 66, p. 107426, Aug 2023.
- [3] Wei, Y., "Prediction of State of Health of Lithium-Ion Battery Using Health Index Informed Attention Model," vol. 23, p. 2587, 2023.
- [4] Al-Meer, M.H., "A Deep Learning Method for the Health State Prediction of Lithium-Ion Batteries Based on LUT-Memory and Quantization," *World Electr. Veh. J.*, vol. 15, p. 38, 2024.
- [5] Y. Ge, F. Zhang, and Y. Ren, "Lithium Ion Battery Health Prediction via Variable Mode Decomposition and Deep Learning Network With Self-Attention Mechanism," *Frontiers in Energy Research*, vol. 10, 2022.
- [6] Wu, J., Fang, L., Meng, J., Lin, M., & Dong, G., "Optimized Multi-Source Fusion Based State of Health Estimation for Lithium-Ion Battery in Fast Charge Applications," *IEEE Transactions on Energy Conversion*, vol. 37, pp. 1489-1498, 2022.
- [7] Jafari, S., & Byun, Y., "Optimizing Battery RUL Prediction of Lithium-Ion Batteries Based on Harris Hawk Optimization Approach Using Random Forest and LightGBM," *IEEE Access*, vol. 11, pp. 87034-87046, 2023.
- [8] Palacin, M., "Understanding ageing in Li-ion batteries: a chemical issue," *Chemical Society reviews*, vol. 13, pp. 4924-4933, 2018.
- [9] Li, T., Thelen, A., Karan, N., Liu, J., Tischer, C., & Hu, C., "Degradation Diagnostics of Lithium-Ion Batteries Cycled Under Varying Conditions," 2023.
- [10] Shen, W., Wang, N., Zhang, J., Wang, F., & Zhang, G., "Heat Generation and Degradation Mechanism of Lithium-Ion Batteries during High-Temperature Aging," *ACS Omega*, vol. 7, pp. 44733 - 44742, 2022.
- [11] Hasan, M., Pourmousavi, S., Bai, F., & Saha, T., "The impact of temperature on battery degradation for large-scale BESS in PV plant," *Australasian Universities Power Engineering Conference (AUPEC)*, pp. 1-6, 2017.
- [12] Wesley Chang, Clement Bommier, Thomas Fair, Justin Yeung, Shripad Patil and Daniel Steingart, "Understanding Adverse Effects of Temperature Shifts on Li-Ion Batteries: An Operando Acoustic Study," *Journal of The Electrochemical Society*, vol. 167, 2020.
- [13] Kong, J., Yang, F., Zhang, X., Pan, E., Peng, Z., & Wang, D., "Voltage-temperature health feature extraction to improve prognostics and health management of lithium-ion batteries," *Energy*, vol. 223, p. 120114, 2021.
- [14] Lui, Y., Li, M., Downey, A., Shen, S., Nemani, V., Ye, H., VanElzen, C., Jain, G., Hu, S., Laflamme, S., & Hu, C., "Physics-based prognostics

- of implantable-grade lithium-ion battery for remaining useful life prediction," *Journal of Power Sources*, 2021.
- [15] Voskuilen, T., Moffat, H., Schroeder, B., & Roberts, S., "Multi-fidelity electrochemical modeling of thermally activated battery cells," *Journal of Power Sources*, vol. 488, p. 229469, 2021.
 - [16] Lv, Y., Yang, D., Yuan, R., Yang, K., & Zhong, H., "A novel multivariate signal processing-based fault diagnosis approach of rotating machinery under various operating conditions," *Measurement Science and Technology*, vol. 33, no. <https://doi.org/10.1088/1361-6501/ac60d5>, 2022.
 - [17] M. Ma and Z. Mao, "Deep-Convolution-Based LSTM Network for Remaining Useful Life Prediction," *IEEE Transactions on Industrial Informatics*, vol. 17, no. <https://doi.org/10.1109/TII.2020.2991796>, pp. 1658-1667, 2021.
 - [18] B. Saha, K. Goebel, S. Poll, J. Christophersen, "Prognostics methods for battery health monitoring using a Bayesian framework," *IEEE Trans. Instrum. Meas.*, vol. 58, no. <https://doi.org/10.1109/TIM.2008.2005965>, p. 291–296, 2008.