JTERA (Jurnal Teknologi Rekayasa), Vol. 11, No. 1, Juni 2026, Hal. 117-130 p-ISSN 2548-737X
Terakreditasi "**Peringkat 3**" oleh Kemdiktisaintek, Nomor : 10/C/C3/DT.05.00/2025 e-ISSN 2548-8678
DOI: 10.31544/jtera.v11.i1.2026.117-130

Short-Term Capacity Estimation of 4S $LiFePO_4$ Battery Under Partial Constant Current Cycling Based on Temporal Convolutional Network

Mohammad Rico Ferdian Khoirunizar[1], Denny Irawan[2]
[1,2]Electrical Engineering, Muhammadiyah University, Gresik
ricoferdiank@gmail.com

Abstrak

Baterai lithium iron phosphate ($LiFePO_4$) banyak digunakan dalam sistem penyimpanan energi terbarukan karena kestabilan termal dan umur siklus yang tinggi. Namun, estimasi kapasitas baterai secara akurat tetap menjadi tantangan, terutama pada baterai berskala kecil dengan keterbatasan data. Penelitian ini mengusulkan kerangka estimasi kapasitas jangka pendek berbasis Temporal Convolutional Network (TCN) untuk paket baterai 4S $LiFePO_4$ 12.8V 6Ah pada kondisi siklus arus konstan parsial. Sistem akuisisi data (DAQ) dirancang menggunakan mikrokontroler Raspberry Pi Pico yang terintegrasi dengan sensor ADS1115, INA226, dan DS18B20 untuk memantau tegangan, arus, dan suhu secara real-time. Dataset sebanyak 76 siklus diperoleh dari eksperimen, kemudian 13 data awal yang anomali dibuang karena mencerminkan fenomena stabilisasi elektrokimia, sehingga menghasilkan 63 sampel efektif. Untuk mengatasi keterbatasan ukuran dataset, evaluasi dilakukan menggunakan Time Series Cross-Validation (TSCV) dengan skema walk-forward lima fold sebagai pengganti split train-test statis. Arsitektur TCN yang diusulkan terdiri dari tiga blok residual dengan laju dilation [1, 2, 4], panjang sekuens tiga timestep, dan fungsi loss Huber. Kinerja TCN dibandingkan dengan tiga model baseline, yaitu Naive/Persistence, Moving Average, dan Linear Regression. Evaluasi lintas fold menghasilkan mean RMSE sebesar 0,01618 ± 0,00250 Ah, MAE sebesar 0,01355 ± 0,00202 Ah, dan MAPE sebesar 0,86 ± 0,13%, melampaui seluruh model baseline. Nilai R^2 yang negatif pada semua model dikaitkan dengan variansi kapasitas yang sangat sempit (~0,05 Ah), yang merupakan keterbatasan statistik metrik R^2 pada data deret waktu bervariansi rendah, bukan kegagalan pemodelan. TCN secara konsisten mencapai nilai R^2 tertinggi di antara semua model yang dievaluasi, mengkonfirmasi keunggulan relatifnya dalam mempelajari pola degradasi temporal jangka pendek dari data siklus baterai yang terbatas.

Kata kunci: Lithium Iron Phosphate ($LiFePO_4$), Estimasi Kapasitas, Raspberry Pi Pico, Akuisisi Data, TCN

Abstract

Lithium iron phosphate ($LiFePO_4$) batteries are widely utilized in renewable energy storage systems due to their high thermal stability and long cycle life. However, accurate battery capacity estimation remains a significant challenge, particularly for small-scale batteries with limited datasets. This study proposes a short-term capacity estimation framework based on a Temporal Convolutional Network (TCN) for a 4S $LiFePO_4$ 12.8 V 6 Ah battery pack under partial constant-current cycling conditions. A custom data acquisition (DAQ) system was developed using a Raspberry Pi Pico microcontroller integrated with ADS1115, INA226, and DS18B20 sensors to monitor voltage, current, and temperature in real time. A dataset of 76 cycles was collected, from which 13 anomalous early-cycle samples were removed due to electrochemical stabilization phenomena, yielding 63 effective samples. To address the limited dataset size, a Time Series Cross-Validation (TSCV) scheme with five walk-forward folds was adopted in place of a conventional static train–test split. The proposed TCN architecture comprises three residual blocks with dilation rates of [1, 2, 4], a sequence length of three timesteps, and Huber loss. Model performance was benchmarked against three baselines: Naive/Persistence, Moving Average, and Linear Regression. Cross-validation evaluation yielded a mean RMSE of 0.01618 ± 0.00250 Ah, MAE of 0.01355 ± 0.00202 Ah, and MAPE of 0.86 ± 0.13%, outperforming all baseline models. The negative R^2 values observed across all models are attributed to the extremely narrow variance of the capacity data (~0.05 Ah), a well-known statistical limitation of R^2 for low-variance time series, rather than a modeling failure; RMSE and MAE are therefore adopted as the primary evaluation criteria. The TCN consistently achieved the highest R^2 among all evaluated models, confirming its superior capability in learning short-term temporal degradation characteristics from limited battery cycling data.

Keywords: *Lithium Iron Phosphate ($LiFePO_4$), Capacity Estimation, Raspberry Pi Pico, Data Acquisition, TCN*

I. INTRODUCTION

Lithium iron phosphate ($LiFePO_4$) batteries have gained significant attention in recent years as one of the most promising energy storage technologies for electric vehicles (EV), stationary energy storage systems, and renewable energy integration. Compared to other lithium-ion battery chemistries, $LiFePO_4$ batteries offer superior thermal stability, enhanced safety, lower toxicity, and a longer cycle life, making them highly suitable for large-scale deployment [1]. However, the performance of $LiFePO_4$ batteries inevitably degrades over repeated charge–discharge cycles due to internal electrochemical reactions, structural changes in electrode materials, and electrolyte decomposition [2]. This degradation process directly affects system reliability and safety, particularly in applications where consistent power delivery is critical.

Data-driven approaches have gained increasing attention due to their capability to capture the complex and nonlinear relationship between battery capacity degradation patterns and operating conditions [3]. Since the degradation behavior of batteries cannot be fully represented by conventional physics-based models, deep learning methods have emerged as a promising solution for modeling and predicting battery capacity degradation more accurately.

Battery capacity is the most direct and physically meaningful indicator of battery aging, as it quantitatively reflects the amount of charge a battery can store and deliver over its operational lifetime. Accurate capacity estimation enables early detection of battery degradation, supports preventive maintenance scheduling, and reduces the risk of unexpected system failures. Furthermore, capacity serves as the fundamental basis for deriving the State of Health (SOH) and estimating the Remaining Useful Life (RUL), making it a critical parameter for Battery Management Systems (BMS) to ensure system reliability and user safety [4].

The shift toward data-driven analysis has also transformed the role of the BMS from merely monitoring voltage and temperature into performing real-time battery health prediction [5]. With the increasing availability of operational data collected by BMS platforms, the opportunity to develop more precise battery capacity estimation models has significantly expanded. Modern BMS trends have increasingly integrated artificial intelligence-based approaches, including Neural Networks, Linear Regression, Gaussian Processes, Random Forests, and Support Vector Machines, to improve the accuracy of battery condition monitoring, capacity estimation, and real-time lifetime prediction [6].

Neural networks also provide high flexibility for integration with other techniques to improve prediction performance. In transfer learning frame-works, LSTM models can be pre-trained on a source domain and fine-tuned on a target domain to increase prediction accuracy for target batteries [7]. This approach is especially useful when datasets are limited or battery cell variability is high. Moreover, deep neural networks can be trained using full charge–discharge cycle data to capture long-term degradation trends and capacity recovery patterns [8]. These advantages make neural networks a strong approach for battery capacity estimation, as they can model the nonlinear and complex degradation behavior more effectively than conventional statistical methods [9], [10], [11].

The application of short-term capacity estimation methods specifically for $LiFePO_4$ batteries under partial constant current cycling conditions remains relatively limited. Although $LiFePO_4$ batteries exhibit a stable voltage plateau, they can experience significant capacity degradation under high-load conditions [12]. This behavior increases the complexity of capacity estimation, as degradation characteristics cannot always be clearly identified from voltage signals alone. Directly estimating capacity from time-series cycling data therefore offers a more straightforward and reliable approach compared to indirect methods that rely on intermediate parameters such as internal resistance or differential voltage analysis [13].

Considering that most existing battery performance prediction studies are primarily based on neural networks and hybrid learning approaches, this study employs a quantitative experimental methodology designed to build a $LiFePO_4$ data acquisition system capable of generating accurate and reliable experimental datasets from 32700-sized cells 4S $LiFePO_4$ 12.8V 6Ah battery pack under partial constant current charge-discharge operating conditions. Based on the acquired dataset, a basic TCN architectture is subsequently developed to perform short-term capacity estimation by learning the temporal degradation characteristics of the battery throughout its operational cycle. The methodology consists of four main stages: System Architecture and Design, Data Acquisition System Hardware Configuration, Capacity Estimation, and Temporal Convolutional Network (TCN) Algorithm. Each

stage was carefully structured to ensure data integrity, experimental consistency, and reliable temporal feature extraction for capacity estimation of $LiFePO_4$ battery applications for renewable energy-based energy storage systems (ESS). The importance of accurate battery capacity estimation becomes increasingly significant as renewable energy systems continue to expand, since battery storage serves as a critical component for maintaining power continuity, reliability, and energy management in renewable energy installations such as solar and wind power generation systems [14], [15].

II. RESEARCH METHOD

A. System Architecture and Design

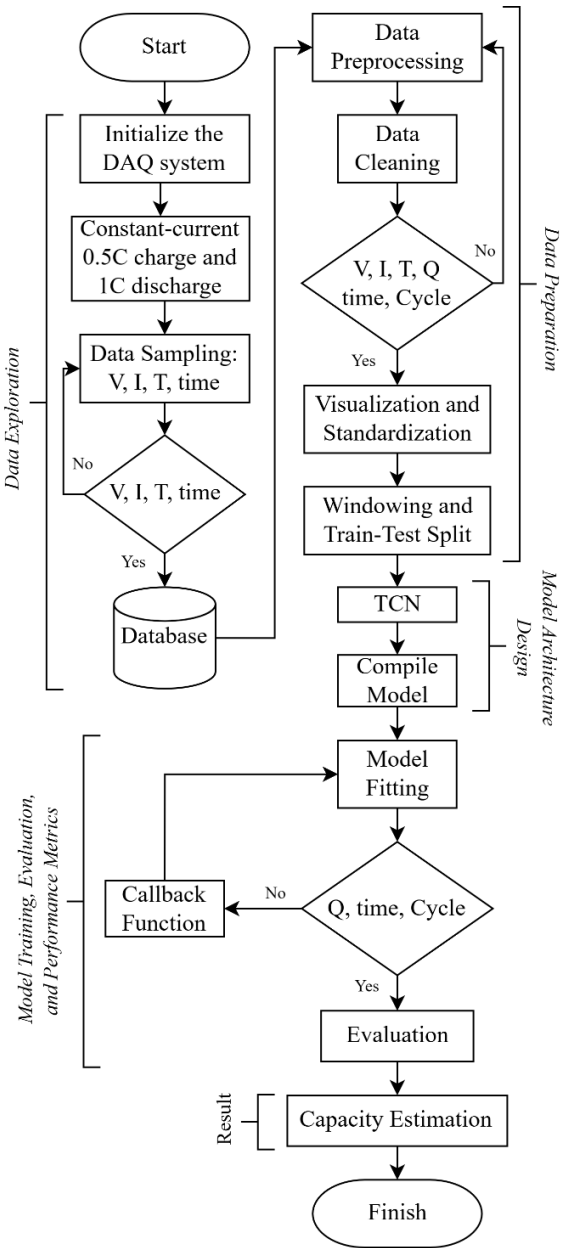

Figure 1 Flowchart of the experiment

The proposed research methodology consists of five main stages: data exploration, data preparation, model architecture design, model training–evaluation–performance metrics, and result analysis, as illustrated in Figure 1. In the data exploration stage, the DAQ system was initialized to acquire battery parameters including voltage (V), current (I), temperature (T), and time during partial constant-current 0.5C charging and 1C discharging processes of the $LiFePO_4$ battery. The collected data were stored in a database for further processing and degradation analysis.

In the data preparation stage, the raw datasets underwent preprocessing and data cleaning to remove noise and inconsistent measurements. The processed variables consisting of voltage (V), current (A), temperature (T), capacity (Q), time, and cycle data were then visualized and standardized to improve model learning performance. Furthermore, a windowing technique and train–test split were applied to generate sequential input data suitable for time-series capacity estimation.

The model architecture design stage focused on the development of a TCN to learn temporal degradation characteristics of the battery. Subsequently, the model was compiled and trained using the prepared dataset, while callback functions were employed to improve convergence and reduce overfitting during the training process. In the evaluation stage, the trained model performance was assessed using prediction error metrics to validate the reliability of the proposed framework. Finally, the developed TCN model was utilized for short-term capacity estimation of the $LiFePO_4$ battery under constant-current cycling conditions.

B. Data Acquisition System Hardware Configuration

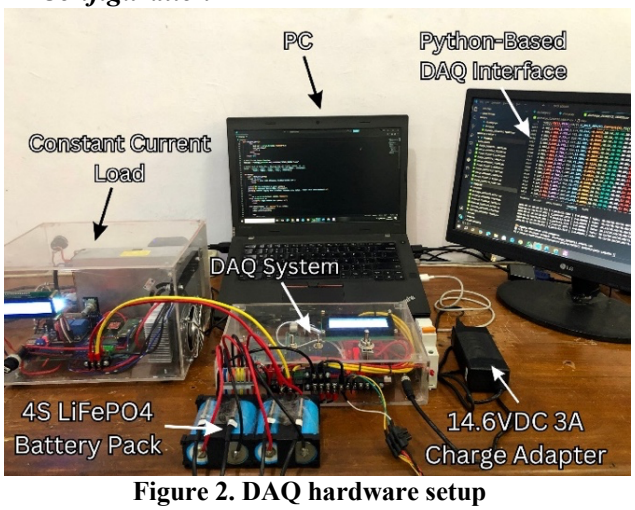

Figure 2. DAQ hardware setup

The data acquisition system (DAQ) hardware configuration was designed to perform real-time

monitoring and recording of electrical and thermal parameters during the charge–discharge cycling experiments of the 4S $LiFePO_4$ battery pack. As illustrated in Figure 2, the experimental setup consists of a 4S $LiFePO_4$ battery pack, a constant-current electronic load, a 14.6VDC 3A charging adapter, a custom DAQ system, and a Python-based monitoring interface operated on a PC. The DAQ system was developed using a Raspberry Pi Pico as the primary microcontroller for sensor interfacing and serial communication. Voltage measurements of each battery cell node of the 4S configuration, including overall pack voltage, were acquired using the ADS1115 analog-to-digital converter, while charging and discharging currents were monitored using the INA226 current sensor. In addition, DS18B20 temperature sensors were installed to monitor the thermal behavior of each battery cell during continuous cycling operations. The acquired sensor data were transmitted serially to the PC and processed using a Python-based DAQ interface for real-time visualization, monitoring, and experimental data logging. The recorded datasets were subsequently stored in structured comma-separated values (.csv) format for further preprocessing and machine learning analysis.

C. Capacity

Considering that current, voltage and internal resistance are used as the input of data acquisition is complicated and local regeneration is ignored, which leads to low prediction accuracy, this paper directly takes capacity as the prediction target. Battery capacity directly reflects the amount of charge a battery can store and deliver during each charge–discharge cycle, making it the most physically meaningful indicator of battery aging. Capacity can be computed straightforwardly from the time integration of the measured current during a discharge cycle, as expressed in the following equation:

$$Q_t = \int_{t_0}^{t_{end}} I(t)\, dt \qquad (1)$$

In (1), Q_t represents the discharged capacity at cycle t, $I(t)$ is the instantaneous discharge current, and t_0 and t_{end} denote the start and end times of the discharge process, respectively.

In this study, capacity Q_t(Ah) extracted from each complete discharge cycle is used as the primary prediction target. The short-term capacity estimation problem is formulated as a sequence-to-one prediction task: given a window of n consecutive historical capacity measurements $\{Q_{t-n+1}, Q_{t-n+2}, \dots, Q_t\}$, the model estimates the capacity of the next cycle $\hat{Q}_{t+1}$. This formulation allows the TCN model to capture the local degradation trend from recent cycling history and project the capacity trajectory one step ahead.

D. Temporal Convolutional Network (TCN) Algorithm

Temporal Convolutional Network (TCN) is a deep learning architecture designed to process sequential or time-series data using one-dimensional causal convolution operations. Unlike conventional recurrent architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), TCN applies convolutional operations along the temporal axis, enabling parallel computation and more efficient learning of long-term temporal dependencies. Due to its capability in capturing degradation trends from sequential battery data, TCN has recently gained significant attention in battery State of Health (SOH) monitoring and Remaining Useful Life (RUL) prediction applications [16].

In this study, the TCN algorithm was employed to learn the temporal degradation characteristics of the $LiFePO_4$ battery under partial charge–discharge cycling experiments. The input sequence consisted of battery capacity degradation data continuously collected from the experimental setup. Capacity (Q) was selected as the primary prediction feature because it directly represents battery aging characteristics and SOH degradation. Meanwhile, voltage, current, and temperature measurements were utilized as supporting parameters to validate the physical degradation characteristics of the battery system.

TCN utilizes causal convolution to ensure that the output at time $t$only depends on present and previous inputs, thereby preventing information leakage from future data. The causal convolution operation can be expressed as:

$$y(t) = \sum_{i=0}^{k-1} f(i)x(t-i) \qquad (2)$$

where $x(t)$ denotes the input sequence, $f(i)$ represents the convolution filter, and k is the kernel size.

To capture long-term temporal dependencies in battery degradation sequences, TCN further employs dilated convolution. By introducing a dilation factor, the receptive field can be expanded without significantly increasing computational complexity. The dilated convolution operation is formulated as:

$$\mathrm{F(s)} = \sum_{\mathrm{i=0}}^{\mathrm{k-1}} \mathrm{f(i)x_{s-d\cdot i}} \qquad (3)$$

where d denotes the dilation factor that controls the spacing between convolution elements.

In addition, residual connections are incorporated into the TCN architecture to improve gradient propagation and stabilize the training process of deep neural networks. The residual mapping can be expressed as:

$$O = \text{Activation}\big(x + F(x)\big) \qquad (4)$$

where x represents the input feature map and $F(x)$ denotes the learned residual function.

In the implementation of this research, the TCN model was trained using sequential battery capacity data obtained from the cycling experiments of a 4S $LiFePO_4$ battery pack. The model learns temporal degradation patterns from the capacity sequence to perform short-term capacity estimation. Compared with recurrent neural network approaches, TCN offers several advantages, including stable gradient propagation, lower computational complexity, faster parallel training, and improved capability in modeling long-term battery capacity degradation behavior.

III. Results and Discussion

A. Accuration and Calibration Sensors

A calibration validation procedure was conducted to verify the measurement accuracy of the DAQ system sensors. The ADS1115 voltage ADC module demonstrated high precision in monitoring individual cell voltages, with a maximum deviation of 0.02 V across the tested operating range. Meanwhile, the INA226 current sensor exhibited a maximum per-centage error of 0.6%. In addition, the DS18B20 temperature sensor showed a slightly higher deviation of up to 1.14% relative to the reference thermometer. Detailed measurement accuracy results for each monitored parameter are presented in Table 1.

Table 1. Result of Accuration and calibration sensor

Parameter	Sensor	Reference	Error (%)
Voltage	ADS1115	Voltmeter	0.33%
Current	INA226	Clamp Meter	0.6%
Temperature	DS18B20	Thermometer	1.4%

B. Battery Cycling Experiment

The battery cycling experiment, illustrated in Figure 3, was performed on a 4S $LiFePO_4$ 12.8V 6Ah battery pack assembled from 32700-sized cells under a partial constant current charge–discharge protocol. Each complete cycle consisted of three sequential phases. In the charging phase, the battery pack was charged at a constant current rate of 0.5C (3A) using a 14.6VDC 3A charging adapter until the pack voltage reached the upper cutoff threshold,

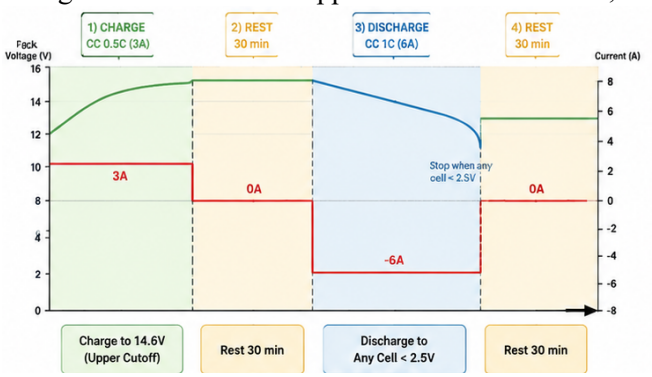

Figure 3. Battery cycling experiment

followed by a 30-minute rest period to allow the battery to stabilize and for internal electrochemical gradients to dissipate.

The discharge phase was subsequently initiated at a constant current rate of 1C (6A) using a programmable electronic load, and was terminated as soon as any individual cell voltage within the 4S configuration fell below 2.5V, thereby protecting the cells from deep discharge damage. Upon completion of each discharge phase, a second 30-minute rest period was applied before initiating the next cycle. This protocol was repeated continuously to accumulate sufficient cycling data for the capacity degradation analysis.

C. *Data Extraction*

Figure 4 illustrates the individual cell voltage discharge profiles of all four series-connected $LiFePO_4$ cells during Cycle 1, plotted as a function of discharge time. The profiles reveal a characteristic discharge behavior comprising three distinct phases: an initial rapid voltage drop associated with ohmic polarization at the onset of discharge, followed by an extended flat plateau region corresponding to the two-phase lithium intercalation reaction intrinsic to $LiFePO_4$ electrochemistry, and a terminal steep voltage decline as the cell approaches full discharge. A critical observation is that Cell 4 (V4) was the first to reach the 2.5 V cutoff threshold at approximately 14.8 minutes, triggering the immediate termination of the discharge load for the entire pack. Following load removal, all cells exhibited a voltage recovery response attributable to the relaxation of internal electrochemical gradients, with V1 and V2 recovering to approximately 3.2 V while V4 recovered to approximately 2.9 V was indicating that V4 exhibited the highest degree of electrochemical polarization and the most advanced state of discharge among the four cells. This cell-level imbalance, which was consistently observed across multiple cycles, underscores the critical role of individual cell

monitoring in multi-cell series configurations for accurate capacity and SOH assessment.

Figure 4. The cell voltage discharge profile

Figure 5 presents the temperature evolution heatmap of the battery pack across all discharge cycles, where each row represents one cycle and the color intensity encodes the average cell temperature computed from the four DS18B20 sensor readings.

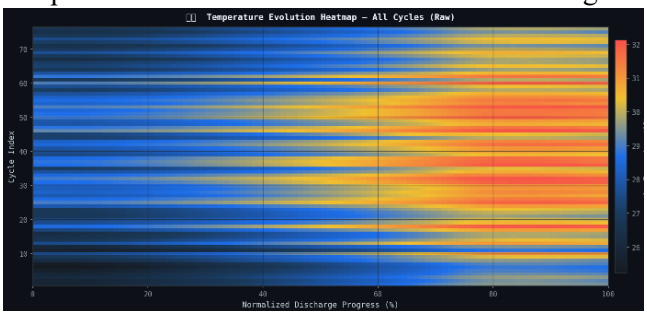

Figure 5. The temperature evolution heatmap

The temperature uniformly increases from approximately 25–26°C at the onset of discharge (0% progress) to a maximum of 31–32°C toward the end of discharge (100% progress) across all cycles, consistent with Joule heating generated by the internal resistance of the cells under the applied 1C constant-current discharge rate. The heatmap reveals a generally stable thermal pattern across the entire 76-cycle dataset, with no evidence of progressive thermal accumulation or anomalous hotspot development. Notably, several cycles particularly those in the mid-cycle range exhibit slightly elevated temperature profiles relative to adjacent cycles, which may be correlated with the localized capacity fluctuations observed in Figure 3.2 and potentially attributable to transient variations in internal resistance during those cycles. The consistent thermal behavior across cycles confirms that the DAQ thermal monitoring system operated reliably throughout the experiment and that no thermal runaway risk was present under the applied cycling conditions.

D. Data Preprocessing

An initial inspection of the raw discharge capacity dataset, as illustrated in Figure 5, identified 13 data points in the early cycling stage that exhibited anomalous behavior, characterized by irregular capacity fluctuations that did not represent the steady-state degradation trend of the battery. These anomalies were consistently observed during the initial cycles and are believed to be associated with early electrochemical stabilization phenomena, including the formation and stabilization of the solid electrolyte interphase (SEI) layer, lithium-ion redistribution, electrode wetting, and internal equilibration processes commonly occurring in newly cycled $LiFePO_4$ cells. Such transient behavior has been widely reported in battery degradation studies and may introduce non-representative temporal patterns into data-driven prediction models.

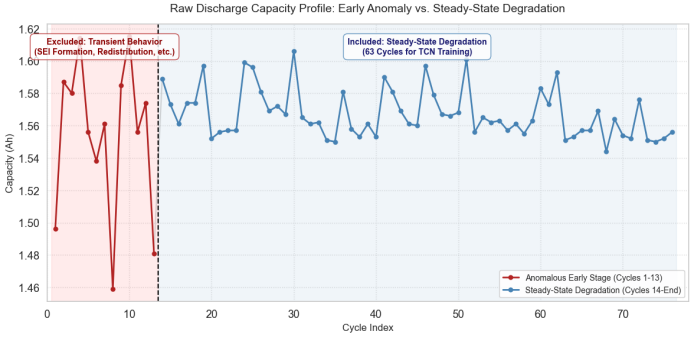

Figure 6. Raw discharge capacity profile

Therefore, the first 13 cycles were excluded from the dataset to ensure that the Temporal Convolutional

Network (TCN) model learned the stable long-term degradation characteristics of the battery. The final dataset used for model training and evaluation consisted of 63 cycles representing the steady-state degradation region.

Table 2. Selected input features

Feature	Unit	Description	Role
Cycle	-	Cycle index number	Temporal index
Time	s	Cumulative discharge time per cycle	Temporal duration
Capacity	Ah	Discharged capacity per cycle	Primary prediction target

The selected input features, as summarized in Table 2, were chosen based on their relevance to the temporal degradation behavior of the $LiFePO_4$ battery during constant-current cycling. The *Cycle* feature represents the chronological progression of battery aging, while *Time* reflects the discharge duration and energy utilization characteristics in each cycle. Meanwhile, *Capacity (Q)* serves as the primary degradation indicator and prediction target, enabling the TCN model to learn historical degradation patterns through temporal sequence modeling. These features provide sufficient degradation-related information while maintaining a simple and computationally efficient input structure.

The cleaned dataset consisting of 63 capacity samples was partitioned and evaluated using Time Series Cross-Validation (TSCV) with TimeSeriesSplit (n_splits = 5) in order to preserve the temporal integrity of the battery degradation sequence and provide a more robust performance estimate compared to a single static train-test split. This walk-forward validation scheme ensures that the model is always trained exclusively on past observations and evaluated on strictly future unseen data, thereby preventing information leakage (data leakage) from the testing subset into the training process. In each fold k, the training and testing indices were assigned chronologically such that:

$$(Train)_k = x_1, x_2, \dots, x_{(n_k)} \quad (5)$$

$$(Test)_k = x_{(n_k+1)}, \dots, x_{(n_k+t_k)} \quad (6)$$

where n_k and t_k denote the number of training and testing samples in fold k, respectively, with n_k expanding with each successive fold.

Within each fold, all capacity values in the training and testing subsets were normalized to the range of [0, 1] using the Min-Max Scaling transformation to ensure numerical compatibility with the activation functions and optimization convergence within the TCN architecture. Critically, the normalization parameters namely the minimum value Q_{min} and maximum value Q_{max} were calculated exclusively from the training subset of each fold and subsequently applied consistently to the corresponding testing subset in order to prevent statistical information contamination from the test data into the training process. The normalization transformation is defined as:

$$Q_i' = \frac{Q_i - Q_{\min}^{\text{train}}}{Q_{\max}^{\text{train}} - Q_{\min}^{\text{train}}} \quad (7)$$

where Q_i represents the discharge capacity value at the i-th cycle in Ah, Q_i' denotes the normalized capacity value, and $Q_{\min}$ and $Q_{\max}$ represent the minimum and maximum capacity values obtained exclusively from the training subset, respectively. After the predictions were generated by the model, the normalized outputs were inverse-transformed back into the original physical unit (Ah) using the following inverse transformation:

$$\widehat{Q_i} = \widehat{Q_i'} \cdot \left(Q_{\max}^{\text{train}} - Q_{\min}^{\text{train}}\right) + Q_{\min}^{\text{train}} \quad (8)$$

To enable the TCN architecture to learn temporal dependencies within the battery capacity degradation trajectory, the normalized dataset within each fold was transformed into input–output pairs using a sliding-window approach with a window length of $L = 3$ cycles. Each input sample consisted of L consecutive historical capacity values, while the target output corresponded to the capacity value at the subsequent cycle. Formally, for the normalized dataset $\{Q_1', Q_2', \dots, Q_N'\}$, the i-th input–output pair is defined as:

$$x_i = [Q_i', Q_{i+1}', \dots, Q_{i+L-1}'] \in \mathbb{R}^{\mathbb{L}} \quad (9)$$

$$\boldsymbol{y_i} = Q_{i+L}' \quad (10)$$

with $i = 1, 2, \dots, N_{subset} - L$. The total number of sequence samples generated from each subset is given by:

$$n_{\text{seq}} = N_{\text{subset}} - L \quad (11)$$

The number of training and testing sequence samples per fold therefore varies across folds, as N_{train} expands with each successive fold following the walk-forward scheme. Performance metrics including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2) were computed on the inverse-transformed predictions (in Ah) for each fold individually. The final reported performance is expressed as the mean ± standard deviation across all five folds.

E. TCN Architecture

The proposed Temporal Convolutional Network (TCN) architecture was designed to model the sequential degradation behavior of the $LiFePO_4$ battery using causal and dilated one-dimensional convolutions. The input layer receives sequential capacity data with a window length of five timesteps ($L = 3$) and one input feature. The network consists of three residual TCN blocks with 8 convolution filters and a kernel size of 3. Dilated convolution rates of $d = 1$, $d = 2$ and $d = 4$ were employed to expand the temporal receptive field while maintaining computational efficiency. The resulting receptive field covers four timesteps, enabling the model to capture short-term temporal dependencies in the degradation trajectory.

Each residual block contains two causal Conv1D layers followed by Layer Normalization, ReLU activation, and Spatial Dropout 1D regularization. Causal padding ensures that the prediction at a given timestep depends only on present and past information, thereby preventing future information leakage during temporal learning. Residual skip connections were also incorporated to improve gradient propagation and stabilize the training process in deep temporal feature extraction.

F. Training Model

The TCN model was trained independently within each fold of the Time Series Cross-Validation (TSCV) procedure using the Adam optimizer with an initial learning rate of $\eta = 5 \times 10^{-4}$ and Huber loss as the objective function. The use of Huber loss was motivated by its robustness to outliers in capacity measurements compared with the conventional mean squared error (MSE), as it behaves quadratically for small prediction errors and linearly for large deviations. The maximum number of training epochs was set to 300 with a mini-batch size of 4, and the training data were processed without shuffling ($shuffle = False$) in order to preserve the temporal ordering of the battery degradation sequence.

Two adaptive callback mechanisms were employed during training to improve convergence and mitigate overfitting. First, Early Stopping monitored the validation loss with a patience value of 30 epochs and restored the model weights corresponding to the lowest validation loss. Second, ReduceLROnPlateau automatically reduced the learning rate by a factor of 0.5 whenever the validation loss failed to improve for 15 consecutive epochs, with a minimum learning rate threshold of 10^{-6}. Within each fold, 20% of the training sequences were allocated as an internal validation set whenever the number of training samples was sufficient ($n_{train} \geq 6$) to enable reliable monitoring of the model generalization performance during training.

The entire training process was conducted independently for each of the five folds of the TSCV framework, resulting in five separately trained model instances. This fold-wise retraining strategy ensures that each model is evaluated exclusively on unseen future observations, thereby eliminating information leakage from the testing period into the training process and providing a realistic assessment of forecasting performance.

To contextualize the predictive performance of the proposed TCN model, three baseline models were implemented and evaluated under the identical TSCV framework. The first baseline was the Naive (Persistence) model, which predicts the capacity at the next cycle as equal to the most recently observed capacity value within the input sequence. The second baseline was the Moving Average (MA) model, which estimates the next capacity value as the arithmetic mean of the $L = 3$ input timesteps. The third baseline was a Linear Regression (LR) model, which was trained using the flattened input sequences from each training fold and subsequently applied to the corresponding testing fold.

For consistency and fairness, all baseline models were trained and evaluated using the same fold partitions as the TCN model. Predictions generated by the baseline models were first obtained in the normalized domain and then inverse-transformed to the original capacity scale (Ah) prior to the calculation of evaluation metrics. The complete hyperparameter configuration adopted for the TCN model is summarized in Table 3.

Table 3. Hyperparameter configuration of TCN model

Hyperparameter	Value
Residual block numbers	3
Filters per block	8
Kernel size	3
Dilation rates	[1, 2, 4]
Batch size	4
Epochs	1000
Leraning rate	5×10^{-4}
Optimizer	Adam

G. Estimation and Evaluation Indicators

The predictive performance of the proposed TCN model was quantitatively evaluated using four regression metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The mathematical formulations of these evaluation metrics are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (12)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (13)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (15)$$

The predictive performance of the TCN model and the three baseline models as Naive/Persistence, Moving Average (MA), and Linear Regression (LR) was evaluated across five folds of Time Series Cross-Validation. The mean and standard deviation of each metric across all folds are presented in Table 4.

As illustrated in Figure 7, the TCN model achieved the lowest mean RMSE of 0.01618 ± 0.00250 Ah, outperforming the Naive model (RMSE = 0.02014 Ah), the MA model (RMSE = 0.01739 Ah), and the LR model (RMSE = 0.01625 Ah). Similarly, the TCN yielded the lowest MAE of 0.01355 ± 0.00202 Ah and the lowest MAPE of 0.86 ± 0.13%, indicating that its predictions deviated from the actual capacity values by less than 1% on average. These results confirm that the TCN model possesses a superior ability to generalize the one-step-ahead capacity degradation trajectory compared to all three baseline models.

Table 2. Comparison of Evaluation Metrics Time Series Cross-Validation (Mean ± Std)

Model	RMSE (Ah)	MAE (Ah)	MAPE (%)	R^2
Naïve	0.02014 ± 0.00361	0.01443 ± 0.00282	0.92 ± 0.18	-1.0656 ± 0.3640
MA	0.01739 ± 0.00413	0.01467 ± 0.00381	0.93 ± 0.24	-0.4753 ± 0.1804
LR	0.01625 ± 0.00201	0.01385 ± 0.00217	0.88 ± 0.14	-0.6242 ± 1.1634
TCN	**0.01618 ± 0.00250**	**0.01355 ± 0.00202**	**0.86 ± 0.13**	**-0.3757 ± 0.3634**

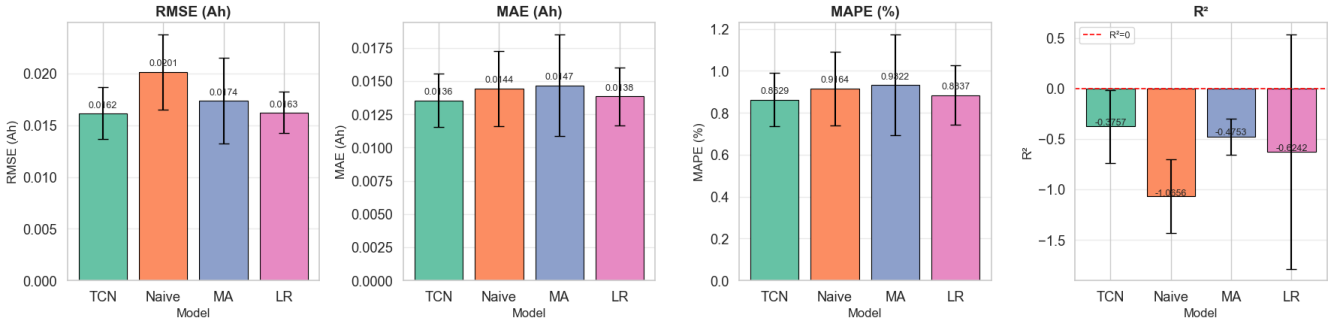

Figure 7. Bar chart RMSE, MAE, MAPE, R^2 all models

Regarding the R^2 metric, all models produced negative values across the five folds, with TCN achieving the least negative mean R^2 of −0.3757 ± 0.3634. This outcome is primarily attributable to two factors: (1) the extremely narrow variance of the discharge capacity data (≈ 0.05 Ah), which causes the denominator SS_{tot} in Eq. (15) to be very small, making R^2 highly sensitive to even minor prediction errors; and (2) the multi-fold concatenation effect, whereby predictions aggregated across folds with independently fitted scalers introduce distributional shifts that artificially suppress the global R^2 value. Consequently, RMSE and MAE are adopted as the principal evaluation criteria in this study, with R^2 serving only as a relative comparative indicator between models.

Figure 8 presents the scatter plots of predicted versus actual discharge capacity values aggregated across all five cross-validation folds for each model. In an ideal predictor, all data points would lie exactly on the diagonal reference line (y = x), shown as the red dashed line. The TCN model (top-left panel) exhibits the tightest clustering of data points around the ideal diagonal, with a positive-slope regression line that most closely approximates the y = x reference among the four models (global RMSE = 0.01637 Ah, R^2 = −0.0882). This indicates that the TCN predictions are positively correlated with the actual capacity values and exhibit relatively low systematic bias.

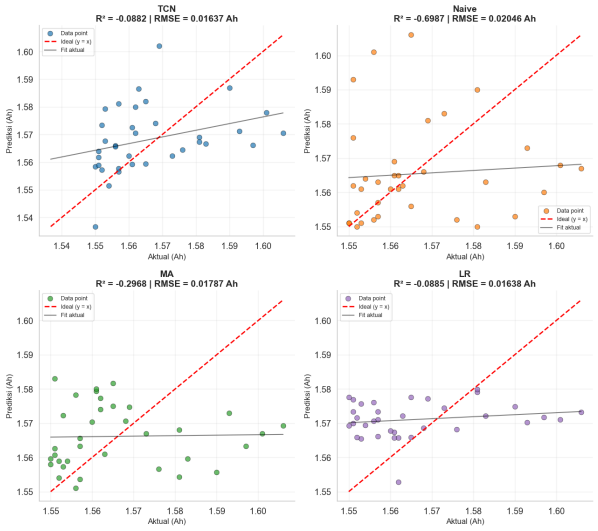

Figure 8. Scatter plot of predicted vs actual for all four models

In contrast, the Naive model (top-right panel) displays a near-flat fit line with considerably wider scatter (R^2 = −0.6987, RMSE = 0.02046 Ah), confirming that simply repeating the most recent observed value introduces significant prediction error when the capacity exhibits short-term fluctuations. The MA model (bottom-left panel) similarly shows a flat fit line (R^2 = −0.2968, RMSE = 0.01787 Ah), as averaging past observations suppresses the responsiveness to sudden capacity changes. The LR model (bottom-right panel) achieves a positive fit slope comparable to TCN (R^2 = −0.0885, RMSE = 0.01638 Ah), though its predictions tend to cluster near a fixed mid-range value, suggesting underfitting to the local degradation dynamics. Overall, the scatter plot analysis corroborates the quantitative metrics in Table X, affirming that the TCN model provides the most accurate and least biased one-step-ahead capacity predictions among the evaluated models.

The temporal prediction trajectories of all models across the aggregated test samples are depicted in Figure 9. The actual capacity sequence (black line) ex-hibits considerable cycle-to-cycle fluctuation within the range of approximately 1.54–1.62 Ah, reflecting the inherent measurement noise and electrochemical variability characteristic of $LiFePO_4$ battery cycling. The TCN predictions (blue dashed line) demonstrate the greatest responsiveness to these fluctuations, tracking the general trend and local minima/maxima of the actual sequence more closely than the compe-ting models. Notably, the TCN successfully captures several sharp capacities drops and recoveries that the Naive and MA baselines fail to anticipate.

The Naive model (orange dotted line) exhibits a one-step phase lag behavior, as expected from a persistence predictor, and substantially overshoots or undershoots at turning points in the degradation trajectory. The MA model (green dotted line) produces smoother predictions that systematically lag behind rapid capacity changes, owing to its averaging nature. The LR model (purple line) produces a relatively stable and subdued prediction trajectory that captures the general mean level but fails to track short-term variations. These observations are consistent with the quantitative results in Table 4 and further validate the advantage of the TCN's dilated causal convolution mechanism in capturing short-term temporal dependencies in battery degradation data.

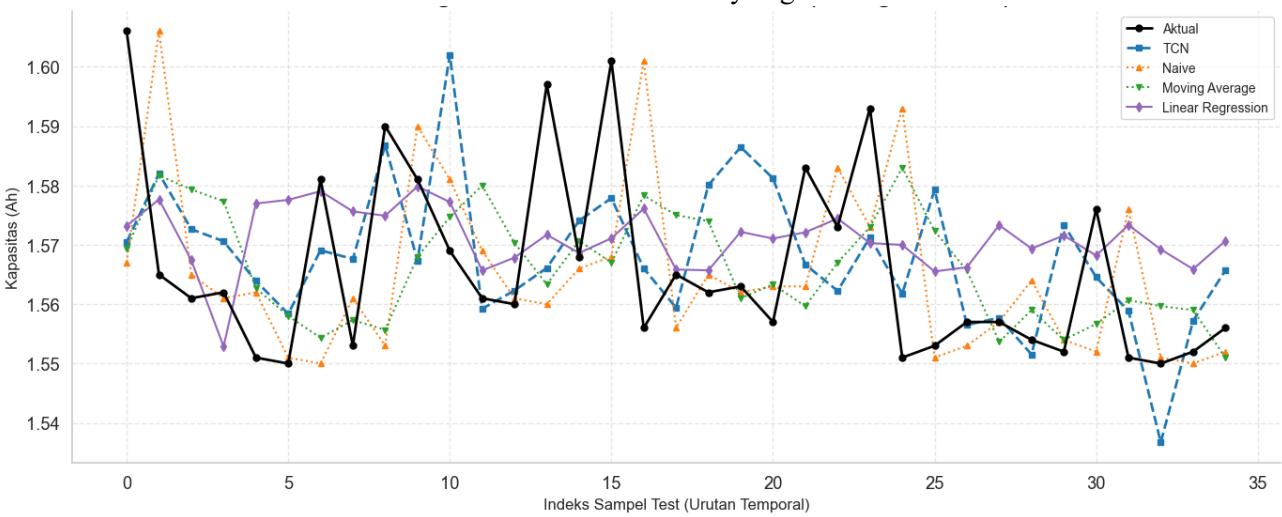

Figure 9. Time series plot of predicted vs actual for all models

To further assess the quality of the TCN predictions, a residual analysis was conducted on the aggregated test predictions across all five folds, with results shown in Figure 10. The left panel displays the residual sequence (actual minus predicted) over the temporal test indices, while the right panel presents the corresponding residual distribution with a kernel density estimate.

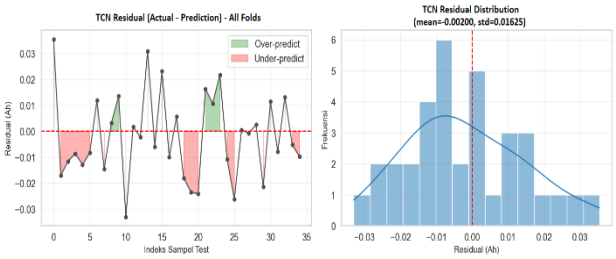

Figure 10. TCN residuals and residual distribution

The residual sequence oscillates around zero without exhibiting any clear systematic trend or structured pattern, suggesting that the TCN model does not suffer from significant systematic bias over successive prediction steps. The mean residual is -0.00200Ah, which is negligibly small relative to the capacity scale (~1.57 Ah), corresponding to a mean relative bias of approximately 0.13%. The standard deviation of the residuals is 0.01625Ah, consistent with the RMSE reported in Table 3.5. Instances of over-prediction (positive residuals, green shaded regions) and under-prediction (negative residuals, red shaded regions) are approximately balanced in magnitude and frequency, further confirming the absence of directional bias in the TCN predictions.

The residual distribution (right panel) is approximately bell-shaped and centered near zero, with slight right skewness. The spread of the distribution within the range of approximately -0.03 to $+0.035$Ah reflects the noise level of the underlying capacity measurements rather than a structural modeling deficiency. This near-zero-mean, approximately symmetric residual distribution provides additional evidence that the TCN model has effectively learned the underlying degradation dynamics without overfitting to a specific fold's data distribution.

Fig. 11 presents the R^2 values of all four models disaggregated by fold, providing a detailed view of the stability and relative ranking of model performance across different training window sizes. Across all five folds, the TCN model consistently achieves the highest (least negative) R^2 value in four out of five folds, with a notable positive $R^2 \approx +0.05$ in Fold 1 the only instance among all models and folds where R^2 exceeds zero. This result indicates that even with a small initial training set (13 samples), the TCN model is capable of extracting meaningful temporal patterns from the limited degradation data.

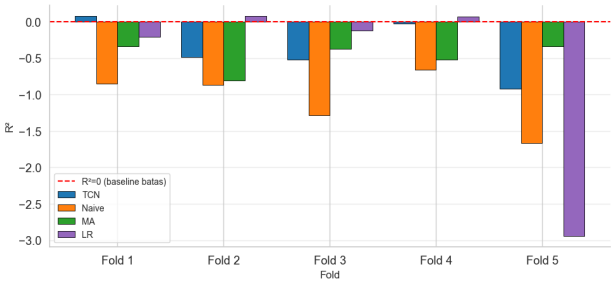

Figure 11. R² per fold all models

The Naive model exhibits the most negative R^2 values across most folds, reaching as low as -1.68 in Fold 5, confirming its fundamental inability to capture the non-stationary fluctuations in the final cycles of the degradation sequence. The MA model shows moderate and relatively stable R^2 values (-0.34 to -0.52), reflecting its smoothing tendency. The LR model produces high variability in R^2 (std $=$ 1.16), including an extreme value of -2.94 in Fold 5, suggesting sensitivity to the specific data distribution encountered in later cycles. In contrast, the TCN's R^2 standard deviation of 0.36 is the lowest among all models, indicating superior fold-to-fold consistency and more robust generalization behavior under the walk-forward validation scheme.

IV. CONCLUSION

This study proposed a short-term capacity estimation framework for a 4S $LiFePO_4$ battery pack based on a Temporal Convolutional Network (TCN) architecture incorporating causal and dilated convolutions, residual connections, and Layer Normalization. To address the limitation of the small dataset size (63 effective cycles), a Time Series Cross-Validation (TSCV) scheme with five walk-forward folds was adopted in place of a conventional static train–test split. This approach ensures that each model instance is trained exclusively on chronologically prior observations and evaluated on strictly future unseen data, thereby preventing temporal information leakage and providing a more robust estimate of generalization performance.

The quantitative evaluation across all five folds demonstrated that the TCN model achieved the best overall predictive performance among the four evaluated models, yielding a mean RMSE of 0.01618 ± 0.00250 Ah, MAE of 0.01355 ± 0.00202 Ah, and MAPE of 0.86 ± 0.13%, compared to the Naive (RMSE = 0.02014 Ah, MAPE = 0.92%), Moving Average (RMSE = 0.01739 Ah, MAPE = 0.93%), and Linear Regression (RMSE = 0.01625 Ah, MAPE = 0.88%) baselines. The average relative prediction error of less than 1% across all folds confirms the capacity of the TCN architecture to learn short-term

temporal degradation patterns from limited battery cycling data with high relative accuracy. Residual analysis further revealed that the TCN predictions exhibit a near-zero mean bias of −0.00200 Ah and a symmetric residual distribution centered around zero, indicating the absence of systematic directional error in the model's capacity estimates. Regarding consistency, the TCN achieved the lowest standard deviation of RMSE (±0.00250 Ah) among all models, demonstrating the most stable fold-to-fold generalization behavior under the walk-forward validation scheme.

The negative R^2 values observed across all models, with the TCN achieving the least negative mean R^2 of −0.3757 compared to −1.0656 for Naive, −0.4753 for Moving Average, and −0.6242 for Linear Regression, are not indicative of a modeling failure. Rather, they reflect two well-known statistical limitations of the R^2 metric when applied to small-scale battery capacity datasets. First, the discharge capacity sequence exhibits an extremely narrow variance of approximately 0.05 Ah across the 63-cycle dataset, which renders the total sum of squares (SS_{tot}) in the R^2 denominator very small, causing even minor absolute prediction errors to produce disproportionately negative R^2 values. Second, the multi-fold concatenation effect inherent in the TSCV evaluation introduces distributional shifts between predictions from independently scaled folds, further suppressing the global R^2 estimate. Consequently, RMSE and MAE in the original physical unit (Ah) were adopted as the primary evaluation criteria in this study, with R^2 serving only as a relative comparative indicator. The fact that the TCN consistently achieves the highest R^2 across four of five individual folds, including a positive R^2 in Fold 1, further supports the conclusion that the proposed architecture is the most effective model among those evaluated for this application. Future work should explore the application of the proposed TCN framework to larger multi-battery datasets, incorporate multi-step-ahead prediction horizons, and investigate transfer learning strategies to improve model generalization across different battery chemistries and operating conditions.

REFERENCE

[1] G. Jin, W. Zhao, J. Zhang, W. Liang, M. Chen, and R. Xu, "High-Temperature Stability of LiFePO4/Carbon Lithium-Ion Batteries: Challenges and Strategies," *Sustainable Chemistry*, vol. 6, no. 1, p. 7, Feb. 2025, doi: 10.3390/suschem6010007.

[2] L. Wang *et al.*, "Insights for understanding multiscale degradation of LiFePO4 cathodes," *eScience*, vol. 2, no. 2, pp. 125–137, Mar. 2022, doi: 10.1016/j.esci.2022.03.006.

[3] V. Olivero-Ortiz, I. O. Pantoja, and C. Robles-Algarín, "Data-Driven Capacity Modeling of 18650 Lithium-Ion Cells from Experimental Electrical Measurements," *Sustainability*, vol. 17, no. 10, p. 4718, May 2025, doi: 10.3390/su17104718.

[4] D. Zhou, Z. Li, J. Zhu, H. Zhang, and L. Hou, "State of Health Monitoring and Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Temporal Convolutional Network," *IEEE Access*, vol. 8, pp. 53307–53320, 2020, doi: 10.1109/ACCESS.2020.2981261.

[5] S. Nazaralizadeh, P. Banerjee, A. K. Srivastava, and P. Famouri, "Battery Energy Storage Systems: A Review of Energy Management Systems and Health Metrics," *Energies (Basel).*, vol. 17, no. 5, p. 1250, Mar. 2024, doi: 10.3390/en17051250.

[6] M.-F. Ng, J. Zhao, Q. Yan, G. J. Conduit, and Z. W. Seh, "Predicting the state of charge and health of batteries using data-driven machine learning," *Nat. Mach. Intell.*, vol. 2, no. 3, pp. 161–170, Mar. 2020, doi: 10.1038/s42256-020-0156-7.

[7] Y. Shi, "Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Transfer Learning and DAE-LSTM," *Academic Journal of Science and Technology*, vol. 9, no. 3, pp. 181–188, Mar. 2024, doi: 10.54097/pdvk6n65.

[8] X. Pang, R. Huang, J. Wen, Y. Shi, J. Jia, and J. Zeng, "A Lithium-ion Battery RUL Prediction Method Considering the Capacity Regeneration Phenomenon," *Energies (Basel).*, vol. 12, no. 12, p. 2247, Jun. 2019, doi: 10.3390/en12122247.

[9] A. Dineva, "Evaluation of Advances in Battery Health Prediction for Electric Vehicles from Traditional Linear Filters to Latest Machine Learning Approaches," *Batteries*, vol. 10, no. 10, p. 356, Oct. 2024, doi: 10.3390/batteries10100356.

[10] A. Rastegarpanah, J. Hathaway, and R. Stolkin, "Rapid Model-Free State of Health Estimation for End-of-First-Life Electric Vehicle Batteries Using Impedance Spectroscopy," *Energies (Basel).*, vol. 14, no. 9, p. 2597, May 2021, doi: 10.3390/en14092597.

[11] M. A. Hoque, M. K. Hassan, A. Hajjo, and M. O. Tokhi, "Neural Network-Based Li-Ion Battery Aging Model at Accelerated C-Rate," *Batteries*, vol. 9, no. 2, p. 93, Jan. 2023, doi: 10.3390/batteries9020093.

[12] C. Sun, Z. Zhang, M. Wang, H. Yang, and Y. Gao, "Effect of Different Carbon Sources on Electrochemical Performance of LiFePO4/C," *Int. J. Electrochem. Sci.*, vol. 15, no. 11, pp. 11215–11226, Nov. 2020, doi: 10.20964/2020.11.57.

[13] J. Zhu *et al.*, "Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation," *Nat. Commun.*, vol. 13, no. 1, p. 2261, Apr. 2022, doi: 10.1038/s41467-022-29837-w.

[14] D. Irawan, "KOORDINASI PEMBANGKIT LISTRIK TENAGA SURYA (PLTS) BERBASIS ALGORITMA MULTIPLE SEQUENCE ALIGNMENT (MSA)," *MULTITEK INDONESIA*, vol. 17, no. 1, pp. 17–26, Jul. 2023, doi: 10.24269/mtkind.v17i1.7021.

[15] D. Irawan and Z. A. Kinanti, "Performance Analysis of Wind Turbine," *Kontribusia : Research Dissemination for Community Development*, vol. 6, no. 1, p. 129, Jan. 2023, doi: 10.30587/kontribusia.v6i1.4256.

[16] N. Zhang, J. Li, Y. Ma, and K. Wu, "Lithium-Ion Batteries state of health estimation based on optimized TCN-GRU-WNN," *Energy Reports*, vol. 13, pp. 2502–2515, Jun. 2025, doi: 10.1016/j.egyr.2025.02.007.