Chapter 5

XAI-Enhanced Dilated Convolutional Model for SC Life Prediction

Supercapacitors are compelling choices for energy storage devices in diverse applications due to their high power density, long lifespan, and wide operating temperature range. The accurate prediction of the life of SCs plays a vital role in ensuring the reliable operation of these energy storage systems. This chapter presents a novel hybrid model that combines a custom dilated temporal convolutional network (CDTCN) and a gated recurrent unit (GRU) to predict the remaining useful life (RUL) of SCs. The developed testbed collects real-time experimental data using hall sensors under specific operating conditions from the SC. The novel CDTCN architecture, with its customized dilations, captures short-term, mid-term, and long-term dependencies via separate branches, while the GRU model captures the sequential and temporal information in the capacity degradation pattern of the SC module. The proposed CDTCN-GRU model efficiently captures capacity degradation patterns at different current rates for the charging and discharging of the SC module. The experimental results show that the proposed model demonstrates

better performance compared to existing models in RUL estimation, with reduced mean absolute error and root mean square error values, along with parameter compatibility. The use of explainable artificial intelligence in the proposed model identifies critical input features for life estimation in SCs and increases the interpretability of the model.

5.1 Introduction

Reliable estimation of the remaining useful life (RUL) of supercapacitors (SCs) is vital for ensuring safe and dependable operation in applications such as intelligent grids, electric vehicles, and clean energy systems. [159], [160]. Accurate RUL prediction enables proactive maintenance, prevents unexpected failures, and maintains performance within safe operating limits [161]. The RUL describes how long an energy storage system, such as a SC, can function effectively before reaching the end of its useful life [162], [163].

The SC's performance deteriorates over time, as shown by a reduction in capacitance or a rise in internal resistance [164]. Two widely used approaches for estimating RUL are model-based methods [165], [166] and data-driven methods [167], [168]. Model-based methods use mathematical or analytical models to characterize the degradation behavior of SCs [169]. However, these methods lack adaptability to unforeseen or fluctuating degradation patterns in time-series data. They also demand significant domain expertise and time for development [170]. Additionally, SC exhibit nonlinear degradation patterns influenced by varying operating conditions. Data-driven models operate without relying on physical models, instead establishing a relationship between input variables and the degradation of SCs. These methods are particularly effective in using machine learning or deep learning to capture complex degradation patterns in SC data. Unlike traditional machine learning approaches, deep learning techniques utilize various strategies to extract more sophisticated information from time-series data.

SC data typically includes temporal features such as voltage, current, and temperature

that evolve over time. This time series data relies on past values to predict future ones. Recurrent neural networks (RNNs) are a deep learning method that uses both current and historical data for predictions, but they often suffer from vanishing gradient issues [171]. Temporal convolutional networks (TCNs) address these problems through residual connections and dilated convolutions, enabling them to effectively capture degradation patterns without the gradient-diminishing issues associated with RNNs [172]. Although gated RNN architectures like long short-term memory (LSTM) and gated recurrent unit (GRU) help mitigate gradient problems, they are still limited by their reliance on hidden states. In contrast, TCNs avoid memory bottlenecks caused by hidden states, making them more suitable for applications involving variable-length sequences.

Traditional TCN using the uniform receptive field fails to capture sophisticated temporal dynamics during the degradation of SCs. In real-world data on capacitors, the dependencies vary across different time scales. In this chapter, an enhanced version of TCN is proposed based on the dilation rates, i.e., a custom dilated temporal convolutional network (CDTCN) to deal with varying dependencies. This novel CDTCN architecture contains multiple branches, each consisting of various TCN blocks employing varying dilation rates. This CDTCN uses different formulas for dilations within each branch, thereby capturing short-term, mid-term and long-term dependencies. These varying dilations help to understand complex dynamics within the degradation mechanisms of SCs. The CDTCN design captures multi-scale temporal patterns but fails to understand the sequential order across a broader time horizon. To learn sequential information and temporal context over time, the GRU unit is integrated into the CDTCN architecture. The proposed hybrid model, which combines CDTCN and GRU, performs exceptionally well in capacitance-degrading tasks simultaneously capturing intricate dependencies and sequential patterns.

A single SC is insufficient for high-power applications, so multiple SCs are connected in series or parallel to meet these demands [173]. The dataset used in this work is derived

from such series configurations, and the RUL is estimated at the pack level rather than for each SC. While data-driven artificial intelligence (AI) methods can effectively predict RUL, they often operate as black boxes, lacking transparency in how these factors influence the output capacitance. Incorporating Explainable AI (XAI) approaches, such as SHapley Additive exPlanations (SHAP), improves the interpretability of the model by clearly delineating the impact of each input feature on RUL predictions [174]. By integrating XAI techniques, the model becomes more transparent and allows for the identification of critical features that contribute to predicting the SC's lifespan [175].

The major contributions of this chapter are:

- The novel CDTCN architecture leverages customized dilation rates to enable multiscale pattern analysis in SCs. By employing branches with varying dilation rates, CDTCN effectively captures dependencies across short-term, mid-term, and long-term timescales. This help to identify short-term fluctuation and sudden variations in voltage and current, as well as aid in capturing dependencies over an extended time interval.
- The proposed hybrid CDTCN-GRU model effectively captures complex degradation dependencies and sequential patterns in SC data. This approach effectively enhances predictive accuracy while maintaining parameter compatibility by achieving a lower root mean square error (RMSE) and mean absolute error (MAE) compared to state-of-the-art models.
- For the first time, XAI techniques have been applied to SCs to identify critical features influencing their RUL predictions. This shows the influence of voltage and current on the output predicted RUL values. Integrating the SHAP method improves the interpretability and transparency of the predictive model, thus promoting trust and understanding among stakeholders and users.

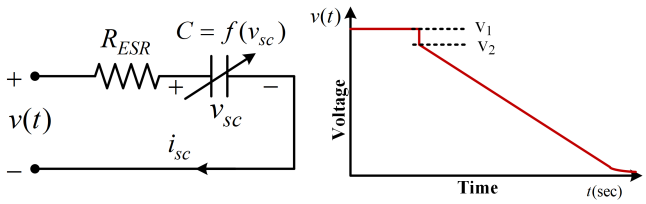

Figure 5.1: The Equivalent circuit representing the discharging characteristic of the SC.

5.2 Methodology

5.2.1 RUL estimation in SC

The SC's performance and its lifespan are mainly influenced by the electrical and thermal stress it experiences during the high-current operation. These thermal changes significantly affect the properties of an electrolyte. As the temperature increases, the viscosity of an electrolyte decreases, leading to ion mobility enhancement and thus allowing ions to penetrate deeper into the porous carbon material. Such variations in the ion accessibility affect the equivalent series resistance (ESR) and the capacitance of the SC. Repeated thermal cycling over time leads to electrolyte decomposition and electrode wear, which accelerates the degradation of both ESR and capacitance. These are key indicators of a SC's health and lifespan or RUL or end of life (EOL). An increase in ESR and/or a decrease in capacitance have been reported t541o indicate a reduction in the lifetime of the SC.

The capacitance of a SC determines its ability to store the charge, which is completely based on the electrostatic principle, whereas the ESR is the internal resistance of the SC, which significantly influences its efficiency and thermal behavior. A low ESR is desirable as it minimizes energy losses and heat generation during high current operation. The equivalent circuit with the discharging characteristics of the SC is depicted in Fig. 5.1,

from where (1) and (2) can be determined.

$$\begin{cases} v_{sc} = v(t) - i_{sc} R_{\text{ESR}} \\ i_x = C \dfrac{dv_{sc}}{dt} \end{cases} \tag{5.1}$$

$$R_{\text{ESR}} = \frac{V_1 - V_2}{I} = \frac{\Delta V}{I} \tag{5.2}$$

Where C is the capacitance, R_{ESR} is the equivalent series resistance of the SC, and the V_{SC} is the voltage across the SC. ΔV is the immediate voltage drop at the start of the discharge, and I is the discharge current.

The lifetime (EOL) criterion for a SC can be defined based on its ESR and capacitance. When the ESR doubles, the SC is considered to have reached its EOL, i.e., R_{EOL} is 200% of its initial (brand-new) value R_{new}. Similarly, from the capacitance point of view, EOL is defined as the point where the capacitance drops to 80% (C_{EOL}) of its original rated value C_{new}. On a lighter note, the life span indicator of a SC is given by (3) and (4) [176].

$$LSI(\%) = \frac{R_{\text{EoL}} - R_{\text{old}}}{R_{\text{EoL}} - R_{\text{new}}} \times 100\% \tag{5.3}$$

$$LSI(\%) = \frac{C_{\text{EoL}} - C_{\text{old}}}{C_{\text{EoL}} - C_{\text{new}}} \times 100\% \tag{5.4}$$

However, it is challenging to estimate the ESR in real-time applications due to sensor noise and the limited accuracy of other system components. Hence, the capacitance of the SC can be considered to be the more reliable parameter for lifespan indication [177], which can be calculated using equation (4), as it serves as the preferred metric for this purpose.

5.2.2 Data Collection and Acquisition

In this study, both a new and an aged Eaton 18 V, 61.7 F SC module were selected for data acquisition. Hall sensors were employed to measure voltage and current for

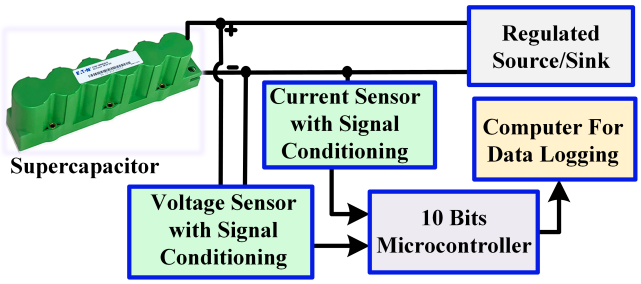

Figure 5.2: Illustration for the data acquisition of the SC.

real-time monitoring and data logging. Specifically, the LV 25-P sensor was used for voltage measurement, while the LA 55-P sensor was utilized for current measurement. To make sure the output accurately reflects the voltage and current levels found in the real world, the sensors are calibrated. The output signals from the sensors are passed through the signal conditioning circuits to mitigate the noises. A 10-bit ADC microcontroller has been used to interface with the workstation and facilitate data logging at a sampling rate of 2 samples per second. The setup illustration for the data acquisition of the SC is depicted in Fig. 5.2.

This integrated configuration enables the storage of voltage and current data. The characterization process involves the charging and discharging of the SC at the constant current profile rate, i.e., between 2A and 20A at room temperature.

5.2.3 Temporal Dependency Modeling with CDTCN

The proposed novel neural network variant, CDTCN, works with a multi-branch TCN, where each branch has varying dilations. The dilation rate defined in each branch is different, with different customised equations for capturing the short-term, medium-term, and long-term dependencies. The short-term dependencies use dilation rates in terms of 2*m, mid-term dependencies capture the dilation rate in the Fibonacci series order, and long-term dependencies take dilation as powers of 2, i.e. 2^m. These branches contain dilated TCN blocks within them for capturing dependencies. These blocks use the dilation defined in the branch. For the layer m in each branch b, the dilation rate $D_{b,m}$

can be expressed formally as follows:

$$D_{b,m} = \begin{cases} D_{\text{short},m} & \text{if } b \mod 3 = 0 \\ D_{\text{med},m} & \text{if } b \mod 3 = 1 \\ D_{\text{long},m} & \text{if } b \mod 3 = 2 \end{cases} \tag{5.5}$$

Here, $D_{short,m}$, $D_{med,m}$, $D_{long,m}$ represent the short-term dilation, medium-term dilation, and long-term dilations, which are selected following the needs of the corresponding branch. Based on the dilations chosen above, the dilated convolutional output is calculated for the dilated TCN blocks within each branch. The dilated convolutional output $V_{b,m,j}$ for layer m in brach b at time step j is calculated as:

$$V_{b,m,j} = \text{ReLU}\left(\sum_{u=1}^{Q} W_{b,m,u} \cdot S_{j-D_{b,m,u}} + b_{b,m}\right) \tag{5.6}$$

where $W_{b,m,u}$ is the weight of the u-th element in the filter for branch b and layer m, $S_{j-D_{b,m,u}}$ represents the input sequence at a specific time step j shifted by the dilation rate. The value of the dilation rate $D_{b,m}$ depends on the branch being selected. $b_{b,m}$ is the bias term used within the branch b and layer m, and Q is the kernel size used. The goal of aggregating this dilated output $V_{b,m,j}$ within a branch is to allow each branch to mix temporal information retrieved by separate layers. The accumulated output of layer m at branch b is given by:

$$O(b)_j = O(b)_j + V_{b,m,j} \tag{5.7}$$

Algorithm 3 Custom Dilated-TCN (CD-TCN) Algorithm

Input: Sequential data $\mathbf{S} = \{s_{pjq}\}$, where p is the sample index, j represents the time steps, and q denotes the feature dimension.

Initialization: Define parameters:

- B: Number of branches.
- M_b: Number of dilated convolution layers in branch b.
- $D_{\text{short},m}$: Dilation rate pattern $(2 \cdot m)$ for short-term branch (branch 1).
- $D_{\text{med},m}$: Dilation rate pattern (Fibonacci sequence) for medium-term branch (branch 2).
- $D_{\text{long},m}$: Dilation rate pattern (2^m) for long-term branch (branch 3).
- Q: Kernel size.

CD-TCN Operations:

for $j = 1$ to J **do**
  Initialize $F_j = 0$
  /Initialize fused output for time step j.
  for $b = 1$ to B **do**
    Initialize $O_j^{(b)} = 0$
    /Initialize branch output for time step j.
    (a) Dilation rate selection for branch b
    for $m = 1$ to M_b **do**
      if $b \mod 3 = 0$ **then**
        $D_{b,m} = D_{\text{short},m}$
        */*Short-term branch uses $2m$ dilation.*/*
      else if $b \mod 3 = 1$ **then**
        $D_{b,m} = D_{\text{med},m}$
        */*Medium-term branch uses Fibonacci dilation.*/*
      else if $b \mod 3 = 2$ **then**
        $D_{b,m} = D_{\text{long},m}$
        */*Long-term branch uses powers of 2 dilation.*/*
      end if
      (b) Dilated convolution layer output
      $V_{b,m,j} = \text{ReLU}\left(\sum_{u=1}^{Q} W_{b,m,u} \cdot S_{j-D_{b,m,u}} + b_{b,m}\right)$
      */*Compute the output of the m^{th} dilated convolution layer in branch b at time step j.*/*
      Update the branch output for time step j:
      $O_j^{(b)} = O_j^{(b)} + V_{b,m,j}$
      */*Accumulate the output of layer m in branch b.*/*
    end for
    Add branch output to fused output:
    $F_j = F_j + O_j^{(b)}$
    */*Aggregate branch b contribution into fused output.*/*
  end for
end for

Output: Fused representation F_j capturing diverse temporal patterns across all branches for each time step.

This output aggregate across layers helps the branches capture sophisticated capacity degradation patterns over multiple temporal resolutions. Finally, the output of all branches is combined, with each branch being specialized on focusing a range of temporal dependencies as follows:

$$F_j = \sum_{b=1}^{B} O(b)_j \tag{5.8}$$

Where F_j is the fused representation capturing temporal patterns across all branches for

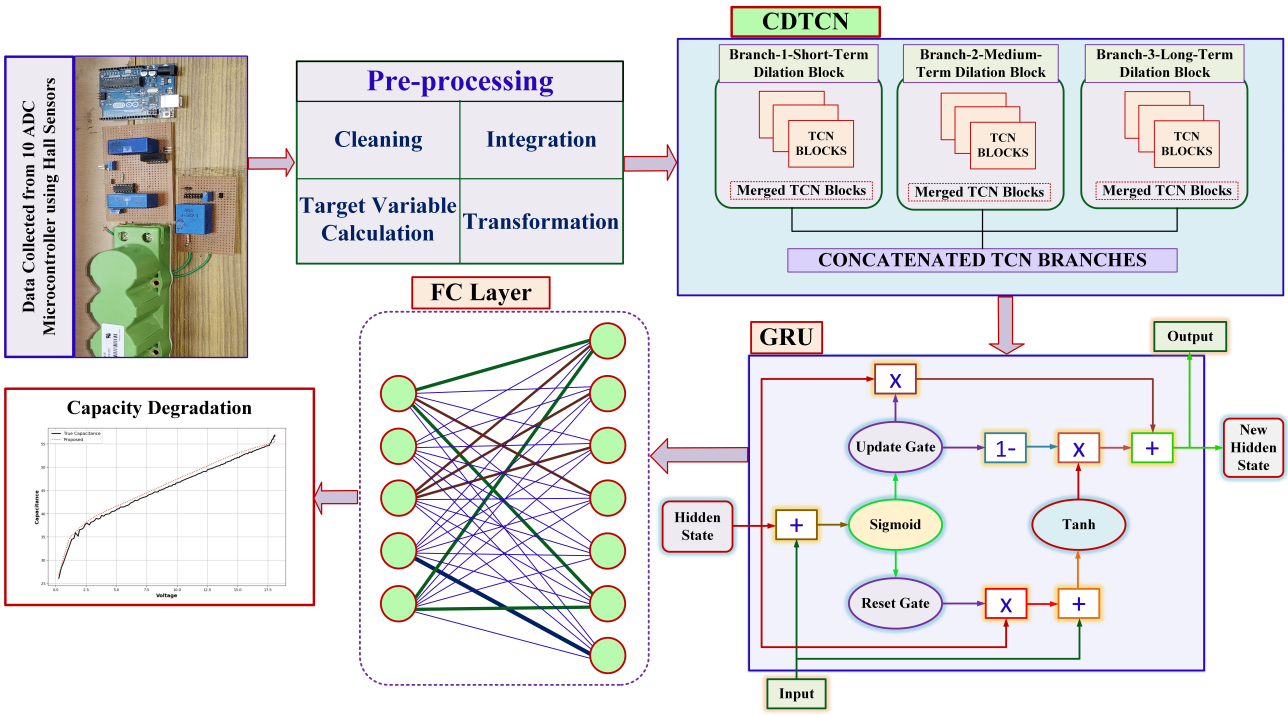

Figure 5.3: The framework of the proposed hybrid model for RUL estimation.

each time step. Each branch focuses on varying temporal dependencies within capacity degradation patterns, such as short-term, mid-term, and long-term dependencies. The complete steps involved in CDTCN are shown in Algorithm 3. The fused output from the CDTCN is passed to the GRU layer for sequential temporal modeling. GRU is versatile and memory efficient because of its gated architecture, which decides whether a portion of fused output F_j should be retained or forgotten. Thus, GRU avoids vanishing gradient issues while preserving crucial information over lengthy temporal sequences in capacitor data.

5.2.4 System Design

The input data, representing electrical parameters such as voltage and current, are collected from Hall sensors. These sensors are carefully calibrated to ensure an accurate representation of real-world voltage and current levels. The output signals from the sensors are then passed through signal conditioning circuits to filter out noise and

improve the data quality. The sensor-collected data undergoes a preprocessing stage, which involves cleaning, integration, transformation, and target variable generation. Preprocessing begins with data cleaning by removing missing values and identifying outliers. The data is collected at different ampere currents at intervals of 0.5 A, starting from 2 A and ending at 20 A, for both charging and discharging. The parameters of the models are selected through hyper-band optimization. The details of the target variable calculation are provided in the RUL estimation section. The test data includes 5 A, 10 A, 15 A, and 20 A, while the remaining data is used for training and validation. A sliding window of size 50 with a step size of 5 is used within the preprocessed input data, generating overlapped segments of multivariate time series data, with voltage and current as the two input characteristics per each time step. The core component of the hybrid model is the CDTCN for capturing temporal dependencies of varying scales. The CDTCN employs three branches of varying dilation rates to capture short-term, mid-term, and long-term dependencies. This dilated TCN blocks uses 32 filters with a kernel size of 5, RELU activation, along with causal padding for maintaining the temporal sequences. The output of the CDTCN is passed to a GRU layer of 64 units, which learns the intricate and sequential patterns in the capacitor data. Afterwards, the GRU's output is passed through a series of fully connected layers of 64 units and finally, a single dense layer for capacitance value estimation. The complete steps involved in the life estimation of the SC are depicted in Fig. 5.3.

5.3 Results and Discussion

5.3.1 Experimental setup

An experiment testbed has been prepared using the SCs, hall sensors, signal conditioners, and a 10-bit microcontroller, as depicted in Fig. 5.4, where the data logging has been done at the sample rate of 2 samples per second. Fig. 5.5 illustrates the characteristics

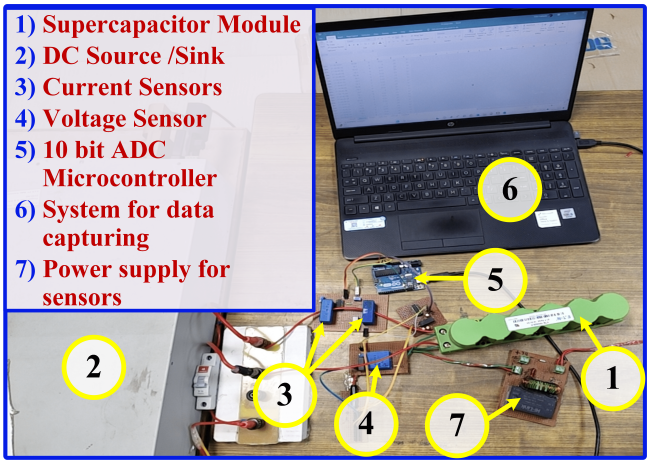

Figure 5.4: Experimental testbed for data logging of the SC.

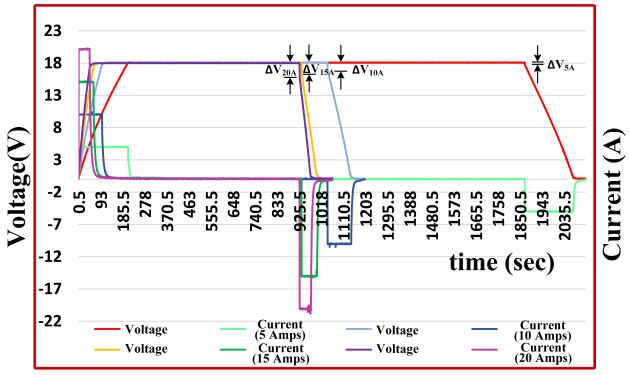

Figure 5.5: Charging/discharging characteristics of the SC module.

of the SC during constant current charge, rest, and discharge phases, using the stored sampled data for voltage and different rating of currents, i.e., 5 A, 10 A, 15 A, and 20 A.

To calculate the capacitance value, experimental data has been used, and column counting has been applied. The total charge stored in the SC is given by (9). Using this, the capacitance value can be formulated as:

$$\begin{cases} Q_t = Q_1 + Q_2 + Q_3 + \cdots + Q_n \\ Q_t = i_1 t_1 + i_2 t_2 + i_3 t_3 + \cdots + i_n t_n \end{cases} \tag{5.9}$$

here, $Q_1 = i_1 t_1$; $Q_2 = i_2 t_2$; $Q_3 = i_3 t_3$; $Q_n = i_n t_n$; Since the time span between each

Table 5.1: Performance comparison with existing models.

Model	Charging		Discharging		Parameters
	MAE	RMSE	MAE	RMSE	
Stacked-BiLSTM [178]	1.68	1.98	2.06	3.20	257765
ANN [179]	1.82	2.16	2.51	3.32	94469
CNN+Transformer [180]	1.08	1.36	1.54	2.78	308069
TCN [181]	1.69	2.13	1.70	2.63	138917
LSTM-RNN [182]	1.72	2.18	1.73	2.58	62255
CDTCN-GRU (Proposed)	**0.94**	**1.31**	**1.46**	**2.35**	**29206**

Table 5.2: Performance metrics of proposed model for different current rates during charging and discharging.

Current Rate	Charging		Discharging	
	MAE	RMSE	MAE	RMSE
5 A	0.76	0.82	0.83	1.25
10 A	1.01	1.29	3.37	3.44
15 A	1.11	1.21	0.81	0.87
20 A	1.77	1.85	3.97	5.18

sampled data is equal, therefore,

$$t = t_1 = t_2 = t_3 = \cdots = t_n \tag{5.10}$$

Hence the charge stored is given by (11) and (12),

$$Q_t = t \sum_{k=1}^{n} i_k \tag{5.11}$$

$$Q_t = C v_{sc} \tag{5.12}$$

So, the capacitance of the SC can be formulated as given in (13)

$$C = \frac{t \sum_{k=1}^{n} i_k}{v_k} \tag{5.13}$$

Here, v_k is the SC voltage v_{sc} at the instant of sample data k.

Once the capacitance is calculated then the estimated life span of the SC can be calculated using equation (4).

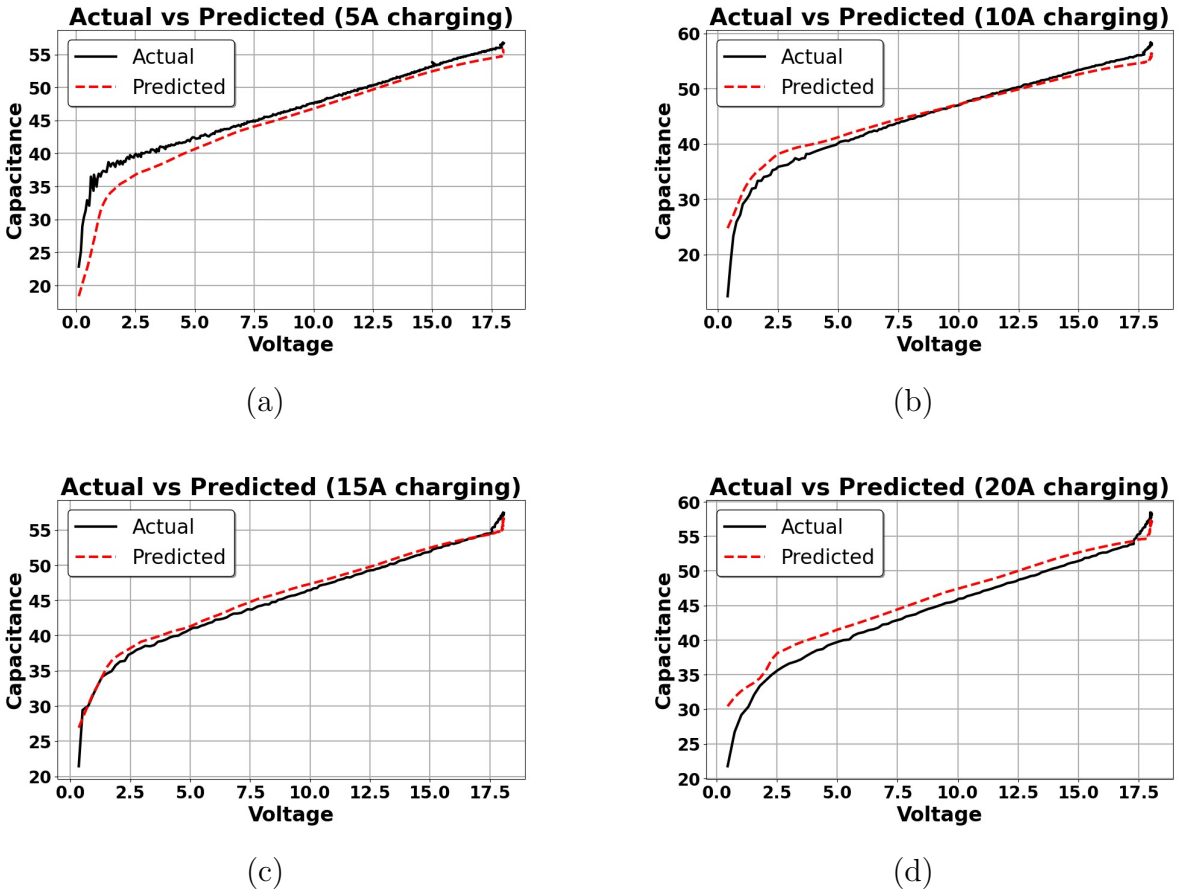

Figure 5.6: Actual vs predicted capacitance for charging current of (a) 5A (b) 10A (c) 15A (d) 20A.

5.3.2 Comparison with state-of-the-art models

The effectiveness of the proposed model is validated by its estimation of RUL in SCs during the charging and discharging scenario. The proposed CDTCN-GRU is compared with existing models that excel in SCs' life estimation. The state-of-the-art model includes stacked bidirectional long-short-term memory (Bi-LSTM), artificial neural network (ANN), convolutional neural networks (CNN) with transformer, TCN, and LSTM with RNN.

The proposed model achieved better results than the state-of-the-art models with parameter efficiency in both the charging and the discharge scenarios, as shown in Table 5.1. In the charging phase, the proposed CDTCN-GRU model obtains an MAE of 0.94

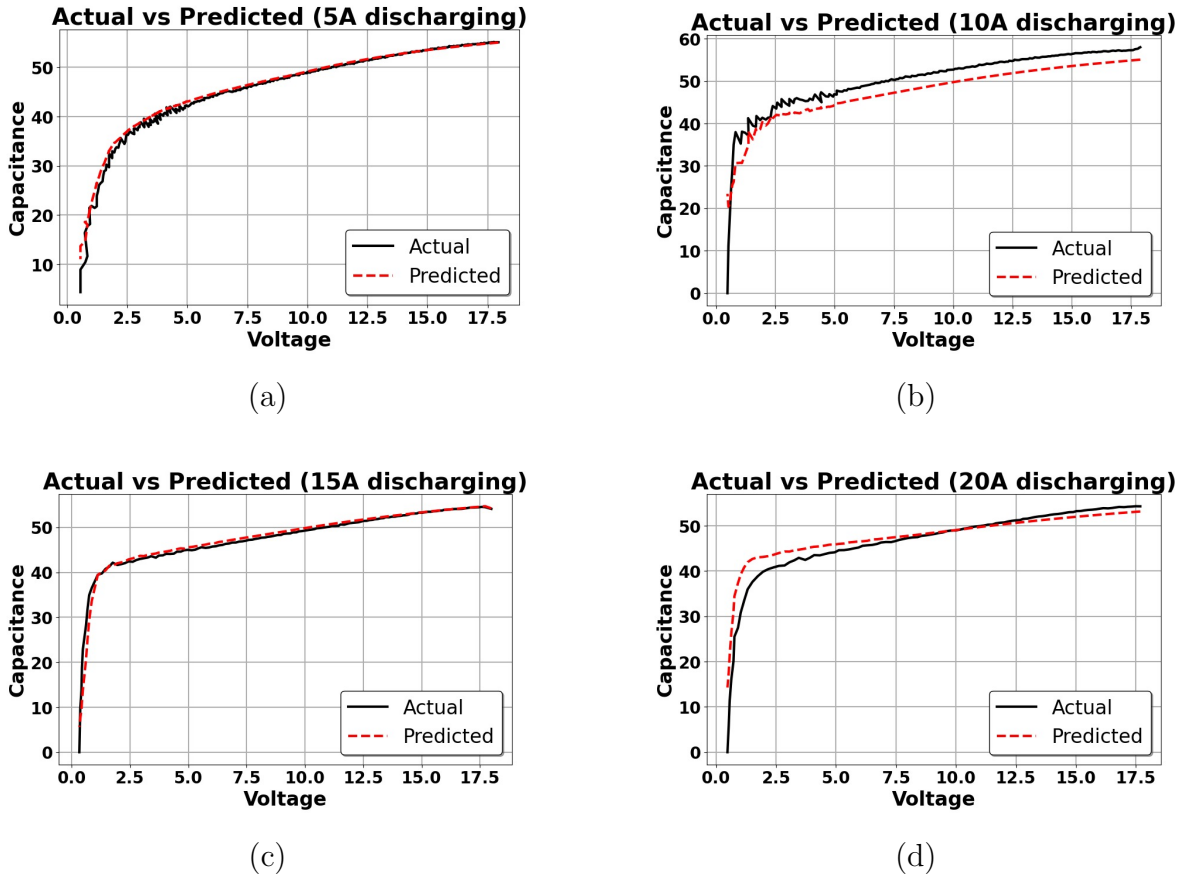

Figure 5.7: Actual vs predicted capacitance for discharging current of (a) 5A (b) 10A (c) 15A (d) 20A.

and an RMSE of 1.31, and in the discharging phase, it obtains an MAE of 1.46 and an RMSE of 2.35. This efficacy results from the model's use of CDTCN to capture temporal dependencies of varying scales and the use of GRU to preserve sequential trends in the degradation pattern of the SC. Charging data, being more stable and less noisy, naturally complement the model's temporal receptive fields and gated memory mechanisms, leading to improved accuracy. Nevertheless, the CDTCN also performs well in discharging scenarios, where fluctuating load conditions cause erratic and noisy data patterns. These results demonstrate the proposed model's robust generalization capabilities across various operating conditions, making it well-suited for real-world predictions in a SC module.

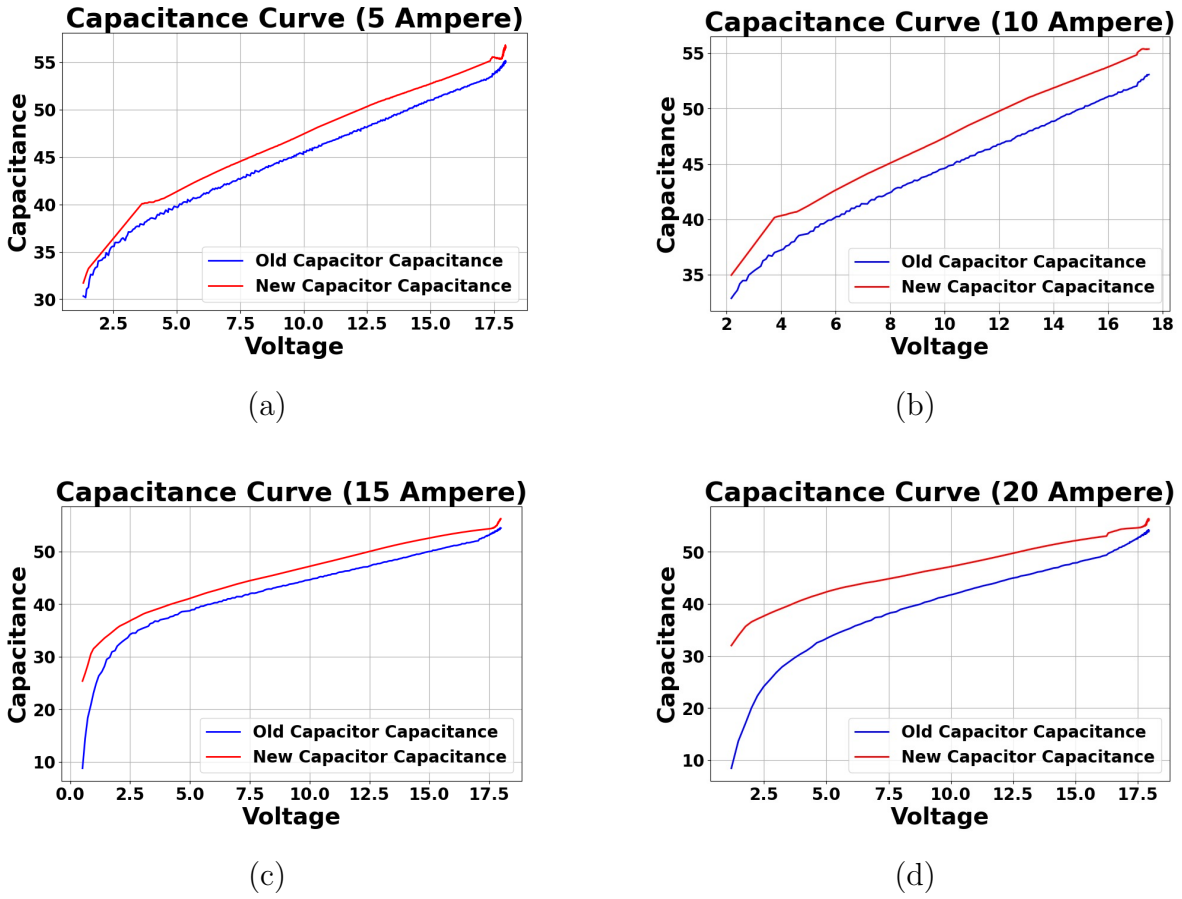

Figure 5.8: Old vs New Capacitance at (a) 5A (b) 10A (c) 15A (d) 20A.

The graphs showing the actual vs predicted capacitance at different current rates of 5 A, 10 A, 15 A, and 20 A during charging are presented in Fig. 5.6 (a), 5.6 (b), 5.6 (c), and 5.6 (d) respectively. Similarly, the graphs for actual vs predicted capacitance at different current rates of 5 A, 10 A, 15 A, and 20 A during discharging are shown in Fig. 5.7 (a), 5.7 (b), 5.7 (c), and 5.7 (d) respectively. The capacitance curves of the old and new capacitance at different current rates are depicted in Fig. 5.8. The graph showing the RUL curve for SCs at varying current rates is shown in Fig. 6.7. Table 5.2 shows the performance of the proposed model at different current rates for charging and discharging. The proposed model demonstrates reliable performance at low and moderate current rates but faces challenges at higher rates, particularly during discharging.

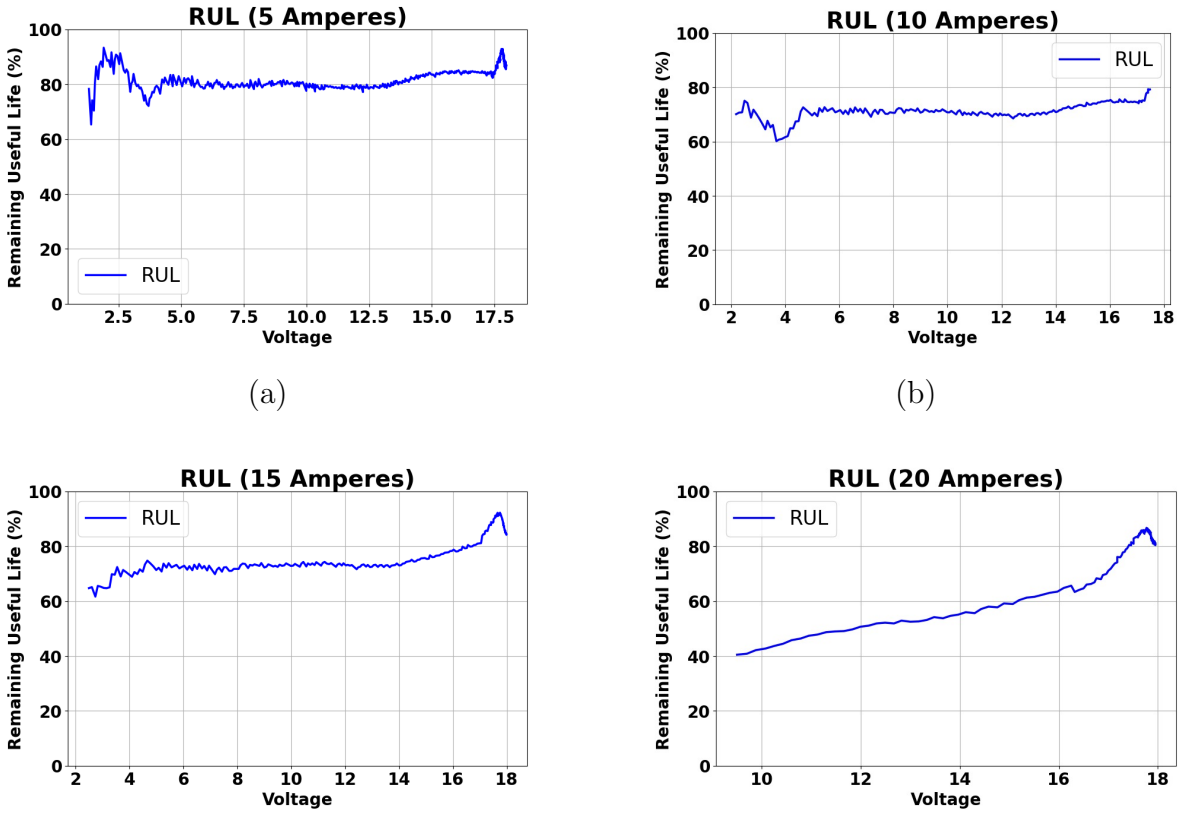

Figure 5.9: RUL curves at (a) 5A (b) 10A (c) 15A (d) 20A.

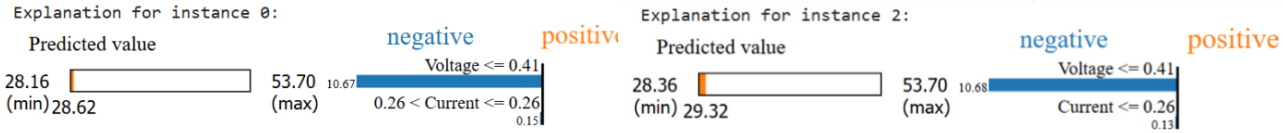

Figure 5.10: LIME instances for two data points, showing the feature contributions (Voltage and Current) to the model's predictions.

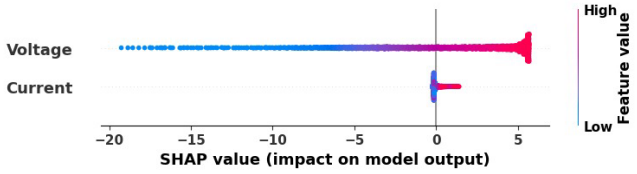

Figure 5.11: The summary plot for all test instances during charging.

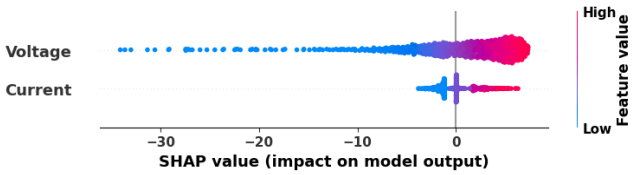

Figure 5.12: The summary plot for all test instances during discharging.

Table 5.3: Ablation study of the components in the CDTCN-GRU.

Model	Charging		Discharging		Parameters
	MAE	RMSE	MAE	RMSE	
TCN [181]	1.69	2.13	1.70	2.63	138917
GRU (baseline)	1.25	1.56	1.50	2.60	29800
CDTCN-GRU (Proposed)	**0.94**	**1.31**	**1.46**	**2.35**	**29206**

5.3.3 Ablation study

An ablation study was conducted to quantify the contribution of each component in the proposed CDTCN-GRU model. The GRU-only variant captures sequential dependencies but lacks the localized multi-scale feature extraction, resulting in higher MAE/RMSE. The TCN-only variant efficiently extracts temporal patterns across multiple scales but cannot adaptively model long-term dependencies as effectively as GRU. The full CDTCN-GRU combines TCN for multi-scale temporal feature extraction with GRU for sequence modeling, yielding the lowest MAE and RMSE. This demonstrates that both components are essential: TCN contributes to capturing diverse temporal features, while GRU enhances sequential modeling and long-term dependency learning. The ablation study results are provided in Table 5.3.

5.3.4 Generalization of proposed model

To support this analysis, a publicly available dataset hosted on Figshare is utilized. This dataset includes aging profiles for 113 SC cells, divided into four groups—Batch 1 through Batch 4. Batch 1 contains 28 SCs, and Batch 2 includes 25 SCs. These two groups were tested under distinct charge–discharge conditions, while Batches 3 and 4 followed procedures identical to either Batch 1 or Batch 2. For model consistency, only

Table 5.4: Performance Comparison on the Public Dataset

Model	MAE	MSE	RMSE	R^2 Score
ANN	0.021555	0.000663	0.025758	0.478643
CNN Transformer	0.005615	0.000050	0.007053	0.960900
LSTM RNN	0.000955	0.000002	0.001473	0.998294
Stacked BiLSTM	0.000884	0.00000238	0.001543	0.998141
Regular TCN	0.000798	0.000001	0.001000	0.999210
CDTCN-GRU (Proposed)	**0.000707**	**0.000001**	**0.000934**	**0.999311**

the first 10,000 cycles of each cell are retained. In Batch 1, the charging and discharging currents are fixed at 20 mA.

The end-of-life point for each cell is defined using a consistent degradation threshold: 90.3% of the nominal capacitance (0.903 F). This threshold balances the number of usable cells and ensures an even distribution of cycle life across the dataset. All SCs are indexed sequentially from SC1 to SC113, with their degradation trajectories—expressed through declining capacitance—provided in full. The transparency, accessibility, and meticulous annotation of this dataset make it a strong benchmark for testing the transferability of prognostic models trained on proprietary data.

To validate the generalisation ability of the proposed CDTCN–GRU model, and evaluate its performance on the publicly released supercapacitor cycling dataset. The same network configurations and hyperparameters used for earlier evaluations on the proprietary dataset were retained to ensure a fair and consistent benchmark. The proposed model was compared against a range of baseline and state-of-the-art time-series forecasting models, including ANN, CNN- CNN-Transformer hybrids, LSTM RNNs, Stacked BiLSTMs, and standard TCN. The state of the art comparison with proposed model in pulic dataset is provided in Table 5.4.

5.3.5 XAI for feature interpretation

AI-based models serve as powerful tools to solve computationally intensive tasks. The role of AI spread to different sectors such as the health sector, education, energy

management, e-commerce, smart grid, transportation and numerous interdisciplinary fields. AI-based models are black-box models, which means that users often lack an understanding of how the model generates outputs from the given inputs. So, there is a need to increase trust and interpretability among users and developers while using these AI-based models. XAI techniques are employed to make these models interpretable and understandable to the users and stakeholders.

XAI identifies the crucial input features for life prediction within SCs and explains how these features contribute to the output life prediction of SCs. The XAI technique, local interpretable model-agnostic explanations (LIME), focuses on individual instances and shows how much each feature contributes to the predicted capacitance within an instance, as shown in Fig. 6.8. In Fig. 6.8 (a), the predicted value of the data instance is 28.62, with the total range varying between 28.16 and 53.70. The predicted initial value is at the lower end of the range. Voltage contributes negatively to the prediction, whereas current, with a positive value, contributes positively to the output prediction. Similarly, Fig. 6.8 (b) shows the predicted value of another instance. In both samples shown, voltage contributes negatively to the output prediction, while current plays a positive role in the prediction. The predicted positive value is slightly higher in the second instance. For global interpretability and consistent results, the XAI method, SHAP, is employed.

The SHAP summary plot provides consistent global explanations for the features across all individual instances in capacitance prediction. The summary plot in Fig. 6.9 indicates the impact of the input features, voltage and current on the predicted capacitance during charging. The plot shows that voltage is the primary feature influencing the output prediction, with a broad range of values varying from -20 to 5. For lower voltage values, SHAP shows a strong negative relationship with the output. For higher voltage values, SHAP shifts towards zero, contributing positively to the output in a minor way. Current is the next most influential feature, with values ranging

from -5 to 5. Larger current values positively impact the output capacitance, while lower current values slightly reduce the model's output. From the summary plot, we can infer that voltage is the most influential feature, with a nuanced and variable relationship with the predicted output. Meanwhile, the current correlates consistently and positively with the predicted output. From the summary plot of discharging, as shown in Fig. 5.12, it is clear that voltage is the most influential feature with a negative impact on output prediction, especially at lower values. The current has less impact compared to the voltage and has a more negative impact on the predicted output in the discharging plot.

5.4 Conclusion

This chapter presents a hybrid custom dilated temporal convolutional network-gated recurrent unit (CDTCN-GRU) model for predicting the end of life of a SC. The data is collected from calibrated Hall sensors and processed through signal conditioning circuits to ensure accuracy and quality. The novel CDTCN architecture uses customized dilation rates in different branches, thereby capturing multi-scale temporal patterns to understand the degradation behavior of SCs. The addition of the GRU further enhances predictions by capturing sequential relationships while maintaining temporal contexts over longer time horizons. The hybrid model excels in understanding complex patterns in the degradation of SCs, outperforming state-of-the-art models with reduced root mean square error and mean absolute error values. Notably, the proposed model achieves these improved results with a lower parameter count. Moreover, the explainable artificial intelligence (XAI) technique-SHapley Additive Explanations used in the proposed model identify the most important temporal features necessary for predicting the life of SCs. The XAI technique enhances user understanding of the proposed black-box model, thereby increasing the interpretability of the model.

While the CDTCN-GRU provides an effective and efficient framework, there remains

a need to explore more advanced architectures that can model complex nonlinear aging dynamics with even greater precision. To address this, the next chapter introduces the Temporal Convolutional Transformer (TCT), a more sophisticated hybrid approach that leverages self-attention for long-range dependency modeling alongside temporal convolutions for local feature extraction, enabling superior RUL prediction performance under diverse operating conditions.