Chapter 6

Temporal Convolutional Transformer Framework for Life Estimation in SCs

Supercapacitors are an advantageous technology in the field of energy storage devices due to their high-power density, low ESR, long life cycle, and compatibility with high energy density devices. However continuous and rigorous use of these devices still affects its life cycle. Therefore, estimating the lifespan of these energy storage devices is essential for maintaining their dependable operation. For the SC degradation prediction, this chapter proposes a Temporal Convolutional Transformer (TCT), a hybrid model that combines Temporal Convolutional Networks (TCNs) with Transformers' self-attention mechanism. TCT facilitates robust modelling of non-linear ageing by using self-attention to capture long-range dependencies and dilated convolutions for localized feature extraction. TCT accurately estimates the remaining useful life (RUL) of SCs, performing better than the state-of-the-art models with lower Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) in both charging and discharging phases at different current rates. Specifically, it achieves an MAE of 0.587 and RMSE of 0.869 during charging, and

an MAE of 1.412 and RMSE of 2.448 in discharging, demonstrating its effectiveness in capturing complex temporal dependencies. An experimental setup is prepared to collect real-time data using Hall sensors, low-pass filters, and 10-bit microcontrollers by charging and discharging the 18 V, 61.7 F old and new SC modules at a constant current rate.

6.1 Introduction

Supercapacitors (SCs) are widely recognized for their ability to deliver quick bursts of power and endure a large number of charging and discharging cycles. Despite their robustness, they do not last indefinitely. Factors such as operating voltage, temperature fluctuations, and continuous cycling gradually wear down the cells, leading to a loss in performance. This degradation is usually observed as an increase in resistance or a reduction in capacitance, both of which limit the efficiency of the device. In real applications, multiple cells are connected in series to form modules capable of handling higher voltage and energy demands. Although these modules can operate reliably through millions of cycles, they eventually reach a point where their performance declines significantly. At this stage, the reduced efficiency can affect the overall stability and reliability of the energy storage system. For this reason, the ability to estimate the remaining useful life (RUL) of SCs has become a crucial aspect of ensuring dependable operation in practical systems.

RUL prediction strategies for SCs can be categorized into two main approaches: model-based techniques and data-driven techniques [183]. Model-based approaches seek to develop mathematical models to describe the degradation and lifespan of SCs. Among these, the Ac-Impedance Spectroscopy Method stands out as the standard technique for lifespan estimation due to its simplicity. However, its application is meticulous, requiring numerous experimental cycling tests for each application scenario and often additional instrumentation [184], [185]. Several studies have proposed various methodologies for

predicting the RUL of SCs under aging conditions. In [93], a particle filter-based model is presented to estimate the posterior values of capacitance and resistance, which serve as aging indicators. Another approach, discussed in [177], proposes an online SC lifetime estimation algorithm that uses an adaptation law based on Lyapunov. In [186], the Extended Kalman Filter (EKF) is employed to estimate the SOH of the SC, utilizing an RC equivalent circuit model that accounts for capacitance variation. While the EKF provides valuable estimations, its performance may be compromised due to the first-order linearization of the SC's nonlinear model, which could lead to occasional filter divergence.

Additionally, some researchers have explored alternative aging indicators. In [187], the degree of self-discharge is considered as a potential aging indicator, while [184] notes that leakage current may not always reliably reflect SC degradation based on test results. Lastly, [188] introduces a state estimation method for the joint prediction of the SC's SOH and state of energy (SOE), leveraging the Unscented Kalman Filter (UKF) algorithm. This method incorporates a first-order equivalent circuit model that accounts for capacitance changes and self-discharge. Although these model-based techniques are widely used in RUL estimation, they struggle to capture the sophisticated patterns in the degradation of a SC module.

The capacitance of SCs degrades nonlinearly over time. To address these complex trends, data-driven techniques employing machine learning and deep learning can be utilized [179], [178]. Traditional recurrent neural network (RNN) methods like long-short-term memory (LSTM) and gated recurrent unit (GRU) are compelling choices in life estimation due to their ability to work well in time-dependent SC data [182]. These RNN models are not suited for extensive real-time monitoring because of their intrinsic sequential nature which restricts parallelism and increases the overall training time. These issues are resolved by temporal convolutional network (TCN).

TCNs are well suited for these time-varying tasks like life estimation in SCs because

of their parallel processing capability and ability to capture long-range dependencies using dilated convolutions [171], [172]. Another advantage is its ability to process the entire input data sequence simultaneously thereby improving the efficiency of the life estimation. However, these convolutional models struggle to adaptively focus on critical ageing patterns since they lack dynamic attention mechanisms. As a result, they are less effective in non-stationary degradation trends or sudden performance shifts in SCs.

While machine learning models such as RNNs and TCNs have advanced SC lifespan estimation, they exhibit significant limitations in capturing the nonlinear and dynamic nature of degradation processes. A critical drawback is their inability to effectively differentiate between gradual aging trends and abrupt performance drops, leading to suboptimal predictions. Transformer-based architectures, leveraging self-attention mechanisms, offer a promising alternative by dynamically weighting historical data to better model complex degradation patterns. However, transformers inherently lack inductive biases for capturing local temporal patterns, making them highly dependent on large amounts of training data to learn short-term dependencies. This limitation reduces their effectiveness in scenarios with limited or noisy datasets, where degradation signals exhibit strong localized trends. This highlights the need for a hybrid model that combines the computational efficiency of convolutional approaches with the adaptive capabilities of attention mechanisms. To address these challenges, this chapter proposes the Temporal Convolutional Transformer (TCT), a novel architecture that integrates TCNs with self-attention mechanisms. By harnessing the efficiency of TCNs in capturing temporal dependencies and the dynamic weighting capability of transformers, the TCT enhances the predictive accuracy of SC life prediction.

The dilated convolution of TCT enables it to excel at localised feature extraction and the self-attention mechanism captures global dependencies through adaptive weighting of historical data. This proposed hybrid model provides a strong robust degradation modeling when non-linear aging is influenced by varying operational conditions. Ad-

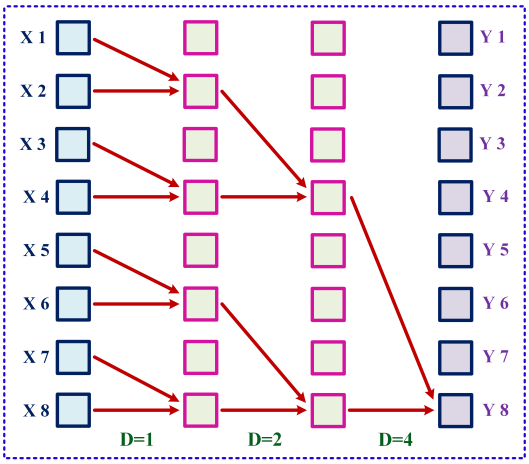

Figure 6.1: Dilation in TCN.

ditionally, the residual connection combined with layer normalization in transformers improves training stability by reducing issues related to vanishing gradients. Through the integration of efficient parallel computation with long-range dependency modelling and adaptive attention mechanisms, TCT makes substantial progress in predicting the degradation pattern of SCs. The major contributions of the proposed work are:

- TCT for robust SC degradation modeling: Novel TCT that integrates TCNs with self-attention mechanisms in transformers enhances localised feature extraction and global dependency modeling. The dilated convolution in TCN effectively captures fine-grained temporal patterns, while adaptive attention mechanisms dynamically weight historical data, ensuring robust degradation modeling even under varying operational conditions.
- Higher predictive performance and enhanced stability: The proposed TCT framework performs better than the existing hybrid machine learning models by leveraging efficient parallel computation, residual connections, and layer normalization to improve training stability and prevent vanishing gradient issues.

6.2 Proposed Methodology

TCN is a one-dimensional convolutional model for capturing sequential dependencies. It is a recurrent model that uses causal convolution and a one-dimensional fully connected network [40]. Dilated convolutions are a key feature of TCNs, enabling the model to capture long-range dependencies, as illustrated in Fig. 6.1 [169]. The output of each dilated convolution layer is added to the feature representation through a residual connection. This residual connection is a common technique in TCNs that helps stabilize training and improve gradient flow. Another advantage is its ability to process the input data sequence in parallel without relying on a gated structure. The input data S is processed by TCN at each time step j using M dilated convolutional layers, as shown below.

$$V_{m,j} = \text{ReLU}\left(\sum_{u=1}^{Q} W_{m,u} \cdot S_{j-D_{\text{dilated},m,u}} + b_m\right) \tag{6.1}$$

where $V_{m,j}$ is the output of the dilated convolution layer at time step j and layer m, $W_{m,u}$ are the convolutional weights, $S_{j-D_{\text{dilated},m,u}}$ is the input sequential data with dilation applied, b_m is the bias term, and ReLU is the Rectified Linear Unit activation function. These individual dilated convolutional layers adds to the final output F_j as shown in the following equation.

$$F_j = F_j + V_{m,j} \tag{6.2}$$

where F_j is the feature representation at time step j. This phase aggregates the retrieved feature information from all the convolutional layers. Before being passed to the multi-head attention within the transformer, the feature information in layer is normalized to facilitate faster training and convergence.

Transformers follow an encoder-decoder structure based on attention mechanisms. The transformer block processes input through a self-attention mechanism, which captures relationships between elements. A feedforward network further transforms the

data, with residual connections ensuring stable gradient flow. Layer normalization is applied after each stage to maintain numerical stability and improve training efficiency. The encoder converts the input sequence into a continuous representation that captures essential information. The decoder then processes this representation step by step, generating an output while also considering previous inputs. Unlike models with fixed memory constraints, transformers use attention with an unlimited range, allowing them to retain long-term dependencies. A crucial component is multi-head attention, which assigns attention weights to the input and produces output vectors that determine how each word relates to others in the sequence. A fundamental transformer function called multi-head self attention enables the proposed model to capture the significance of several temporal stages into the prediction process. The multi-head self attention in transformers is computed as:

$$A_j = \text{MultiHeadAttention}(F_j, F_j, F_j) \tag{6.3}$$

The query, key and value are same as F_j. This enables model to concurrently focus on the data required for capacitor prediction.

$$F_j = F_j + A_j \tag{6.4}$$

This stage incorporates the learnt attention based representation into the feature map. This step is followed by a dense layer for refining the feature representation. The residual connection is used with dense layer to prevent information loss. The final predicted capacitance value is stabilized by applying layer normalization. The final temporal feature representation for each of the steps is obtained after passing through dilated convolutions, attention, and dense layers, which predict the degradation values of capacitance. The complete steps of proposed TCT model is shown in Algorithm 4.

Algorithm 4 Temporal Convolutional Transformer (TCT) Algorithm

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

Initialization: Define parameters:

- N: Number of time steps.
- M: Number of dilated convolution layers.
- Q: Kernel size for convolution layers.
- H: Number of attention heads for MultiHeadAttention.
- F_{out}: Final output size.
- D_{dilated}: Dilated convolution rates.

TCT Operations:

for $j = 1$ to N **do**
  Initialize $F_j = 0$
  /Initialize temporal feature representation for time step j.
  (a) Dilated Convolution Block
  for $m = 1$ to M **do**
    Compute the output of the dilated convolution layer:
    $V_{m,j} = \text{ReLU}\left(\sum_{u=1}^{Q} W_{m,u} \cdot S_{j-D_{\text{dilated},m,u}} + b_m\right)$
    */*Compute output of the m^{th} dilated convolution layer at time step j.*/*
    Update feature representation:
    $F_j = F_j + V_{m,j}$
    */*Accumulate outputs from convolution layers.*/*
  end for
  (b) Normalization Step
  $F_j = \text{LayerNorm}(F_j)$
  */*Stabilize and normalize feature representation.*/*
  (c) Multi-Head Self-Attention
  $A_j = \text{MultiHeadAttention}(F_j, F_j, F_j)$
  */*Apply self-attention to capture long-range dependencies.*/*
  Update feature representation:
  $F_j = F_j + A_j$
  */*Incorporate attention output using residual connection.*/*
  (d) Dense Projection with Residual Connection
  $D_j = \text{Dense}(F_j)$
  */*Transform features into higher-level representation.*/*
  $F_j = D_j + F_j$
  */*Add residual connection with dense output.*/*
  (e) Final Normalization
  $F_j = \text{LayerNorm}(F_j)$
  */*Normalize final feature representation for time step j.*/*
end for

Output: Final temporal feature representations F_j for each time step, capturing both local temporal features (via dilated convolutions) and long-range dependencies (via self-attention).

The proposed model effectively captures temporal trends in life estimation of the SCs. In order to improve prediction performance, the model combines TCN with Multi-Head Self-Attention processes, taking advantage of their complementing advantages. The SC data collected from sensors consists of two important input features, voltage and current, which are preprocessed with MinMaxScaler normalization. To aid in learning temporal patterns, the time-series data is segmented into sliding windows of size 50 with

a step size of 5. This approach ensures that the proposed model analyzes sequential data in a structured manner while preserving the temporal order. The model includes causal convolutional layers with exponentially increasing dilation rates (1, 2, 4, 8) to capture short- and long-term dependencies. A Multi-Head Self-Attention layer with four attention heads helps focus on important time steps and identify relationships across distant points, complementing the local feature extraction of TCN. A dense layer with 64 units and GELU activation introduces non-linearity, followed by dropout for better generalization. The final output layer predicts the capacitance value.

The model is trained using the Adam optimizer, which adjusts the learning rate dynamically to enhance convergence. Since the task involves regression, Mean Squared Error (MSE) is used as the loss function. To minimize overfitting, early stopping is applied with a patience of fifty epochs, monitoring validation loss and restoring the best model weights. Training runs for a maximum of 250 epochs with a batch size of 32, balancing computational efficiency and performance. The integration of causal convolutions, attention mechanisms, and regularization techniques strengthens the model's ability to capture intricate temporal patterns while ensuring robustness.

6.3 Life estimation in SCs

The SC generally consists of several components, viz. collector, electrolytes, electrodes, and separators. The separator functions as it does in batteries by providing isolation between the electrodes, preventing short circuits while allowing ions to pass through. The fundamental principle of a SC is to store electrical energy through the electric double-layer capacitance formed by the charge separation at the interface between the electrolyte and the solution. However, the SC's performance deteriorates with the time of operation, i.e., its energy capacity decreases. This is because the SCs experience thermal and electrical stress during the continuous surge and transient current operation, resulting in a change in the electrolyte properties of SCs.

An increase in thermal stress reduces the viscosity of electrolyte resulting in enhanced ion mobility and allowing ions to penetrate more deeply into the porous carbon material. As an effect, this increases the equivalent series resistance (ESR) and reduces the capacitance of the SC. Prolonged exposure to thermal stress due to continuous surge and transient current on the SC leads to electrolyte decomposition and electrode wear, which accelerates the degradation of both ESR and capacitance. The ESR increase is generally caused by the organic electrolyte decomposition on the active surface of the carbon substrate [189].

$$C(v_{sc}) = C_0 + k_v v_{sc} \tag{6.5}$$

$$v = v_{sc} + i_{sc} ESR \tag{6.6}$$

$$i_{sc} = C\frac{dv_{sc}}{dt} \tag{6.7}$$

$$ESR = \frac{\Delta V}{I_{dis}} \tag{6.8}$$

Fig. 6.2 shows the SC equivalent model with the constant current discharging behaviour, where C represents the capacitance of the SC, which is linearly dependent on voltage, ESR is the equivalent series resistance of the SC, and v is the terminal voltage across the SC, v_{sc} is the voltage across the capacitor. ΔV is the instant voltage dip during the constant current discharge, and I_{dis} is the discharge current. The capacitance of the SC is given by (5), whereas the terminal voltage is given by (6). The current through the SC is given by (7), and the ESR of the SC can be estimated using (8). During the charging/discharging process, it stores/releases electrical energy, which is entirely based on the electrostatic principle. However, a small resistance occurs during the charging/discharging due to the electrochemical process because of the electrolyte, which is also called the ESR of the SC. It mainly impacts the efficiency and is influenced by the thermal behaviour. Hence, the need for low ESR is mandated to minimize the losses and mitigate the heat dissipation during the high current operation. ESR and the

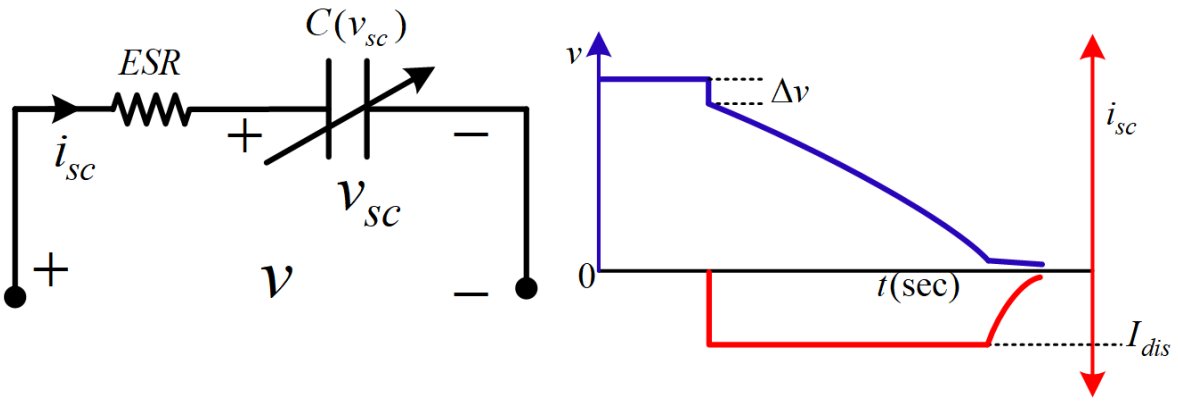

Figure 6.2: Equivalent circuit with constant current discharge behaviour of the SC.

capacitance of the SC play a crucial role in indicating the health of the SC, which can be correlated by the remaining useful life (RUL) of a SC. An increase in the ESR of the SC and a decrease in the capacitance of the SC designates a reduction in the lifespan of the SC [176].

The lifetime or end of life (EL) criterion for a SC can be defined when the ESR doubles or becomes 200% compared to the brand-new SC's ESR (ESR_{new}). Whereas for the capacitor, the capacitance of the SC reaches 80% of the brand-new capacitance (C_{new}) value; then the SC is considered to have reached its EL, i.e., ESR_{EL} will be 200% of ESR_{new}, and C_{EL} will be 80% of its brand-new value. The formula for the EL or RUL indicator considering ESR can be expressed as (9), and the EL or RUL indicator considering the capacitance is expressed as (10).

$$L_{IND}(\%) = \frac{ESR_{EL} - ESR_{old}}{ESR_{EL} - ESR_{new}} \times 100\% \tag{6.9}$$

$$L_{IND}(\%) = \frac{C_{EL} - C_{old}}{C_{EL} - C_{new}} \times 100\% \tag{6.10}$$

ESR_{old} and C_{old} are the calculated value of the ESR and capacitance using the captured data from the sensors.

Though, estimating the ESR in real-time applications is difficult due to the internal resistances of the connector and lugs, due to which the instant voltage at the constant

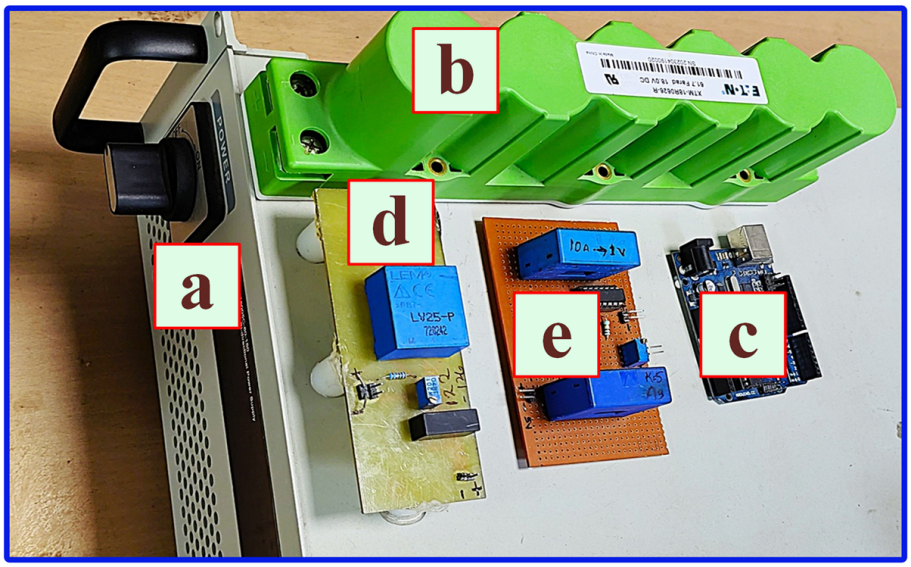

Figure 6.3: Experimental setup for data monitoring of the SC.

current discharging would be higher, resulting in a huge deviation in ESR.

Hence, the capacitance of the SC would be a great choice. Further, it would be a more trustable parameter for EL or the RUL estimation of the SC, which can be calculated using equation (10), as it serves as the preferred metric for this purpose.

6.4 Results and Discussion

An experiment set is arranged using regulated DC source/sink, the old and new SC modules, hall sensors, low pass filter, signal conditioners, and a 10-bit microcontroller, as illustrated in Fig. 6.3. The experiment was conducted to collect data from both a new and an old Eaton 18 V, 61.7 F SC module. The modules were charged and discharged at a constant current ranging from 2 Amps to 20 Amps, at an interval of 0.5 Amps, all at room temperature. During this process, voltage and current data were recorded. Hall transducer-based voltage and current sensors are used for monitoring and data logging. The LV 25-P sensor model is used for voltage measurement, while the LA 25-P and LA

55-P sensor models are used for current measurement. These sensors are calibrated such that they ensure that the output accurately reflects the measured voltage and current. The hall sensors' output is then passed through a low-pass filter and, thereafter, a signal conditioning circuit to minimize noise, guaranteeing accurate measurements. Further, a 10-bit ADC microcontroller, which is integrated with the workstation, has been utilized for data acquisition purposes at a sampling rate of 2 samples per second. This combined hardware setup enables the data logging of voltage and current data under the constant current charging and discharging of the SC.

The capacitance value has been calculated using the logging experimental data, and charge counting has been applied. The total charge (q_{total}) during the constant current charging or discharging of the SC is given by (11).

$$q_{total} = q_1 + q_2 + q_3 + q_4 + \cdots + q_n \tag{6.11}$$

Where, $q_1 = i_1 t_1$, $q_2 = i_2 t_2$, $q_3 = i_3 t_3$, $q_4 = i_4 t_4$, $q_n = i_n t_n$.

$$q_{total} = i_1 t_1 + i_2 t_2 + i_3 t_3 + i_4 t_4 + \cdots + i_n t_n \tag{6.12}$$

Since the sample interval for each data is equal,

$$t = t_1 = t_2 = t_3 = \cdots = t_n \tag{6.13}$$

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

$$q_{total} = t \sum_{k=1}^{n} i_k \tag{6.14}$$

$$q_{total} = C v_{sc} \tag{6.15}$$

So, the capacitance of the SC during the charging can be calculated as given in (16),

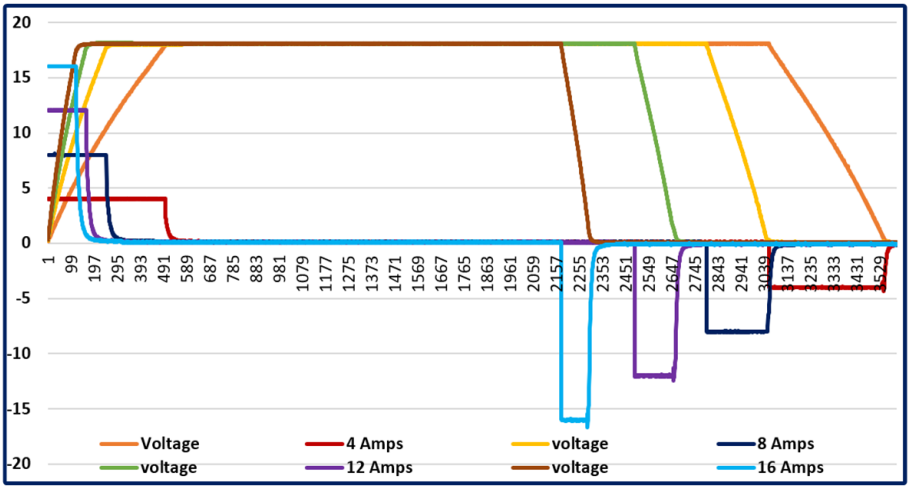

Figure 6.4: Constant current charging and discharging characteristics of the SC module.

whereas the capacitance of the SC during the discharging can be calculated using (17).

$$C = \frac{t \sum_{k=1}^{n} i_k}{v_k} \tag{6.16}$$

$$C = \frac{q_C - t \sum_{k=4}^{m} i_k}{v_k} \tag{6.17}$$

Here, n represents the instant up to which the SC is charged, m represents the instant up to which the SC is discharged, q_C is the total charge stored when the SC is reached to 18 V, and v_k is the SC voltage at the instant of sample data k. Once the capacitance is calculated then the estimated EL or RUL of the SC can be estimated using equation (10). Fig. 6.4 illustrates the characteristics of the SC during constant current charge-rest-discharge phases for the different current profiles i.e., for 4 Amps, 8 Amps, 12 Amps, and 16 Amps, where the SC is charged up to 18 V. The data was collected at different current levels with a 0.5A increment using Hall sensors, ranging from 2A to 20A for both charging and discharging. The current levels of 4A, 9A, 14A, and 19A were used for testing, while the remaining data was used for training.

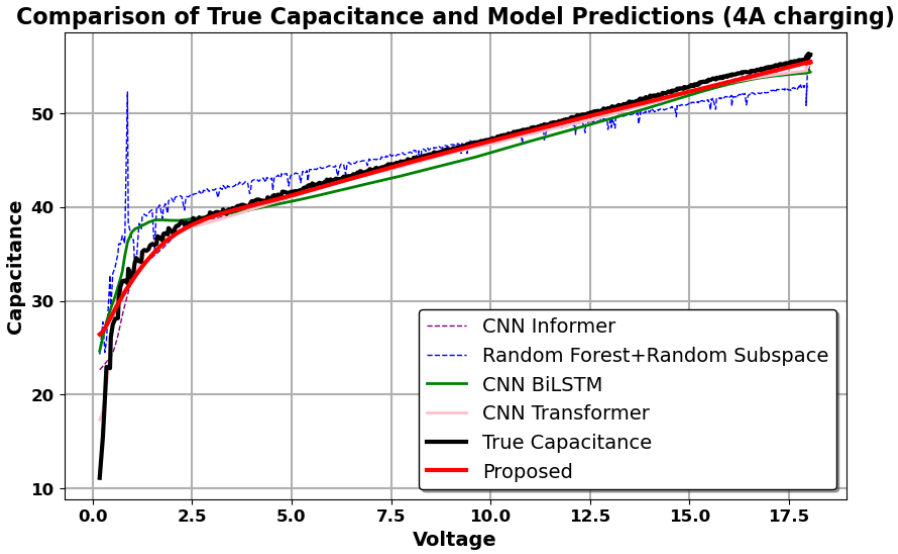

Figure 6.5: Performance comparison of state-of-the-art models for capacitor lifespan prediction under 4A charging.

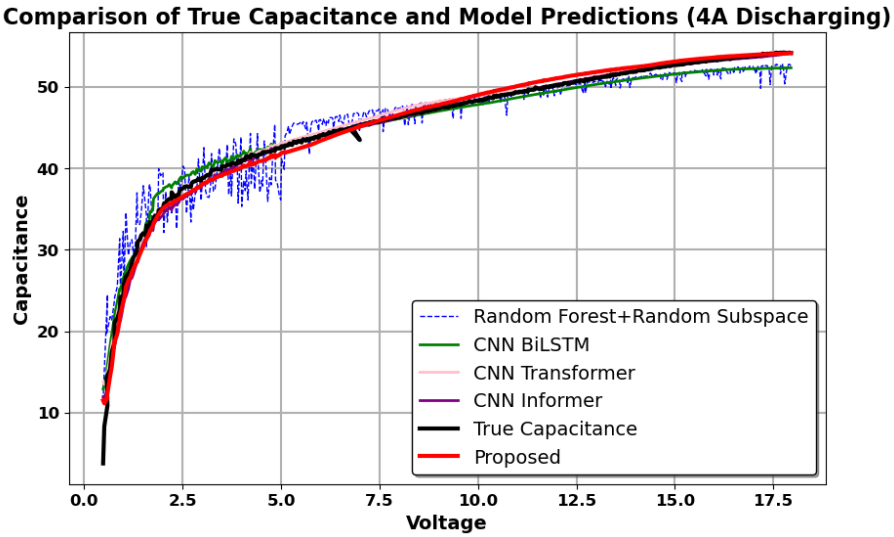

Figure 6.6: Performance comparison of state-of-the-art models for capacitor lifespan prediction under 4A discharging.

6.4.1 Comparison with the state of the art models

Table 6.1 presents the comparative performance of various machine learning and deep learning techniques for both charging and discharging prediction tasks, evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The proposed

Table 6.1: Comparison of different techniques for charging and discharging predictions

Technique	Charging		Discharging	
	MAE	**RMSE**	**MAE**	**RMSE**
CNN-Informer [167]	1.389	1.794	1.439	2.507
CNN-Transformer [180]	0.672	0.941	2.593	3.268
RF-RS [190]	1.925	2.345	2.564	3.363
CNN-BiLSTM [191]	1.008	1.384	2.771	3.516
TCT (Proposed)	**0.587**	**0.869**	**1.412**	**2.448**

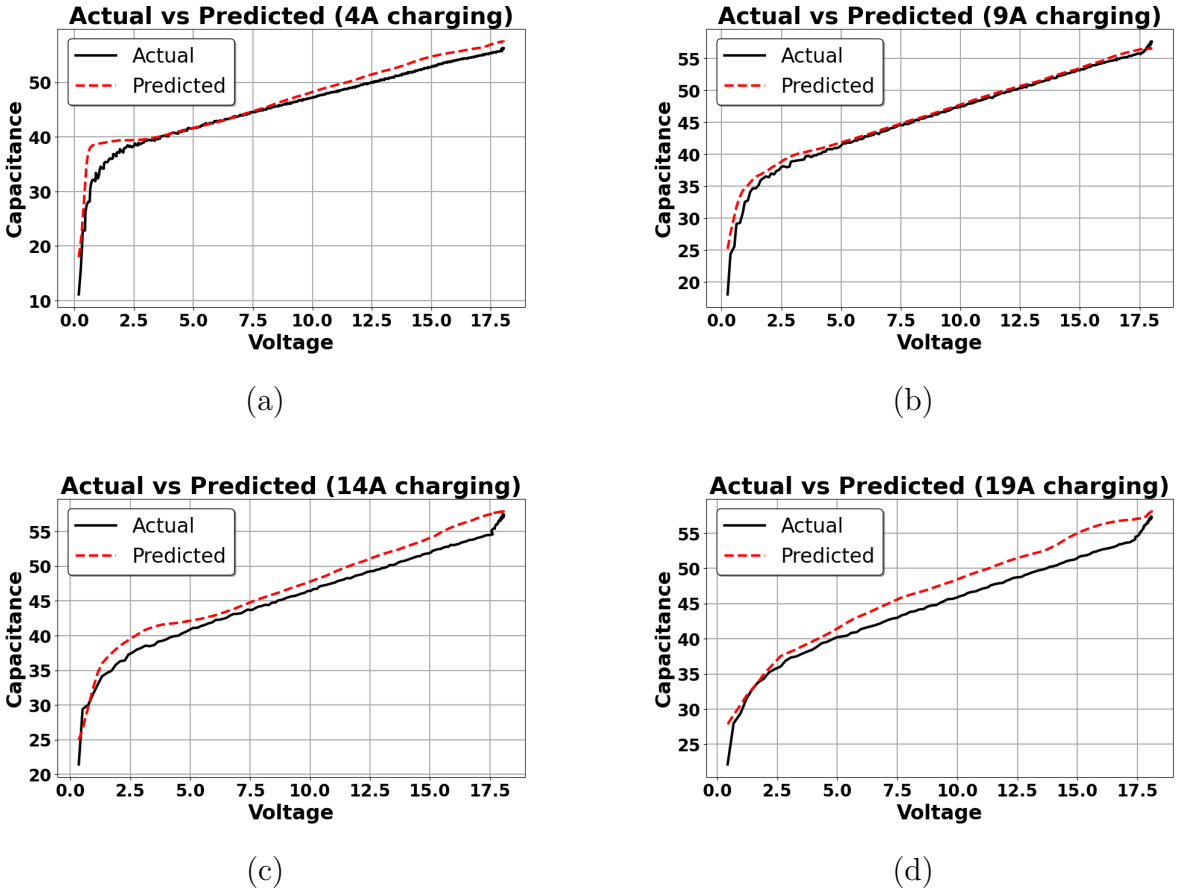

Figure 6.7: Comparison of measured and estimated capacitance for charging currents of (a) 4A, (b) 9A, (c) 14A, and (d) 19A.

hybrid TCT performs better than all baseline models in both scenarios, demonstrating its effectiveness in capturing complex temporal dependencies in SC data. The Fig. 6.5 and Fig. 6.6 show the actual versus predicted capacitance values at 4A charging and

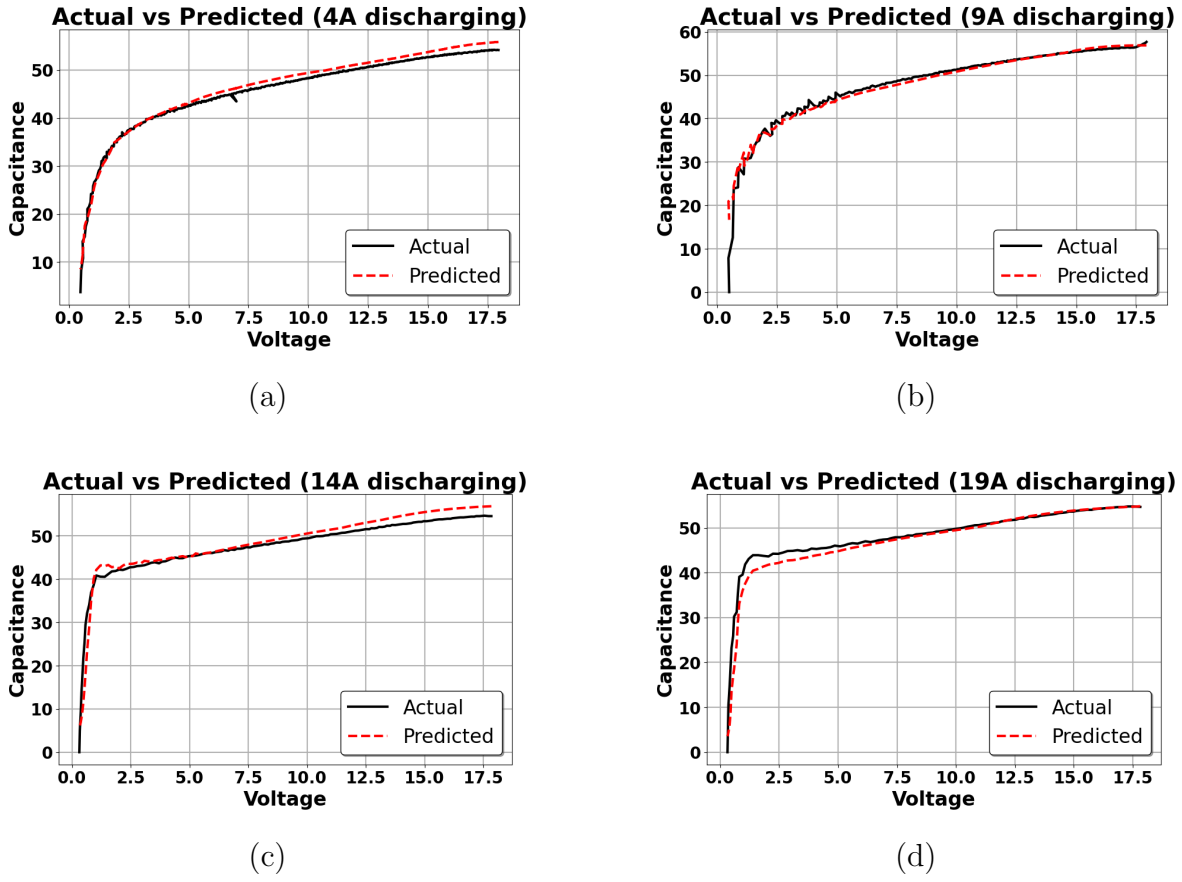

Figure 6.8: Comparison of measured and estimated capacitance for discharging currents of (a) 4A, (b) 9A, (c) 14A, and (d) 19A.

discharging for all the models.

For the charging phase, TCT achieves the lowest MAE of 0.587 and RMSE of 0.869, significantly improving upon the CNN-Transformer and CNN-Informer. The performance of the Random Forest - Random Space (RF-RS) model is notably weaker, indicating that traditional ensemble learning struggles with temporal sequence modeling. In the discharging phase, the TCT model performs better with an MAE of 1.412 and RMSE of 2.448 compared to the existing models, further reinforcing the advantage of transformer-based architectures in sequential battery data analysis.

Fig. 6.7 (a), 6.7 (b), 6.7 (c), and 6.7 (d) illustrate the comparison between actual and predicted capacitance at charging current levels of 4A, 9A, 14A, and 19A, respectively.

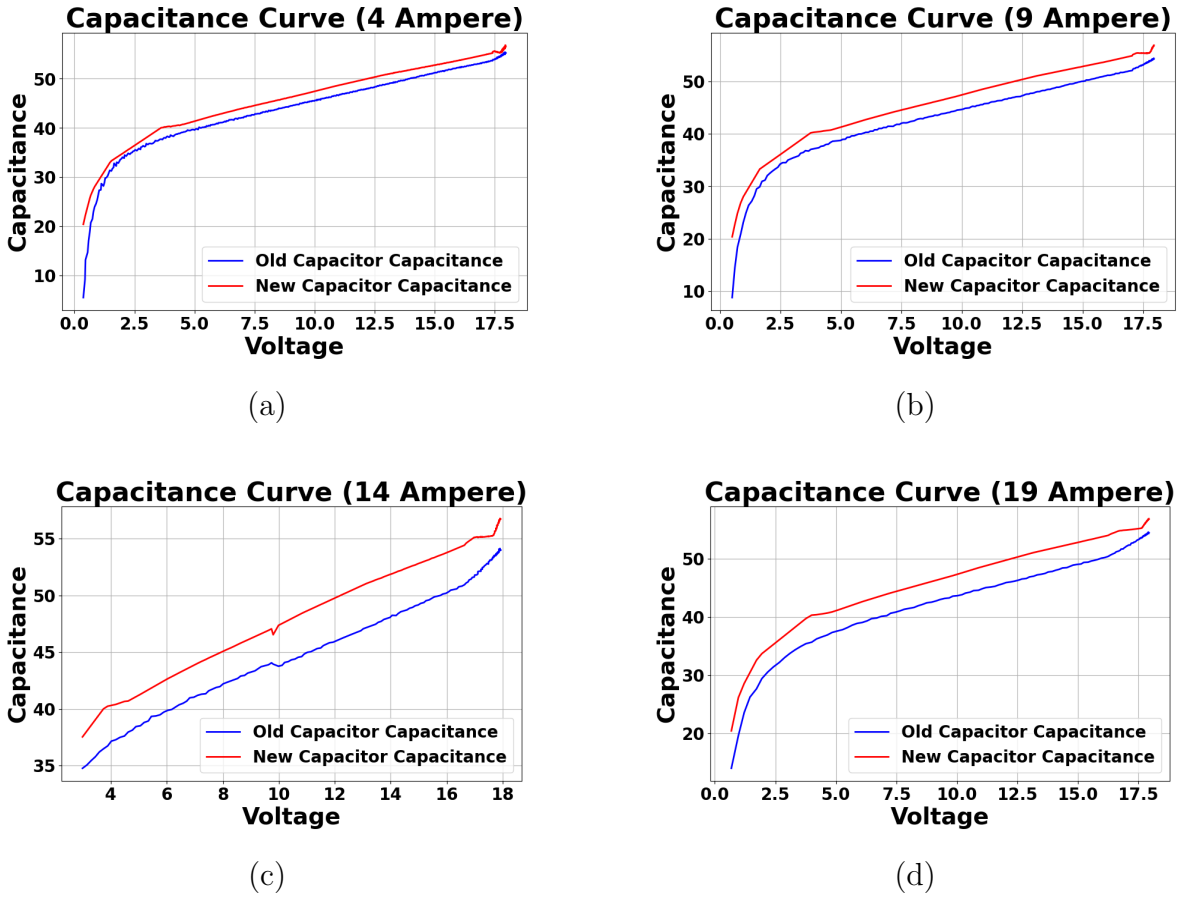

Figure 6.9: Comparison of old and new capacitance at (a) 4A, (b) 9A, (c) 14A, and (d) 19A.

Likewise, Fig. 6.8 (a), 6.8 (b), 6.8 (c), and 6.8 (d) depict the actual vs. predicted capacitance at discharging current levels of 4A, 9A, 14A, and 19A, respectively. Fig. 6.9 illustrates the capacitance variations of the old and new capacitors across different current levels.

The superior performance of TCT can be attributed to its ability to efficiently capture both short-term and long-term dependencies using TCN in conjunction with a self-attention mechanism. Unlike recurrent architectures such as CNN-BiLSTM, which suffer from vanishing gradients and high computational overhead, TCT maintains a balance between computational efficiency and prediction accuracy.

6.5 Conclusion

This chapter presents a novel hybrid robust model for accurately estimating SCs' remaining useful life (RUL). The proposed model uses a temporal convolutional transformer (TCT) that integrates the temporal convolutional network (TCN) with the self-attention mechanism of transformers. The proposed model effectively captures both long-range and localised dependencies in the degradation patterns of the SC module. While the self-attention mechanism adaptively estimates prior data to improve predicted accuracy, the dilated convolutions in TCT enhance feature extraction. Additionally, residual connections and layer normalization improve the training stability by reducing vanishing gradient problems, leading to more effective learning. The model's integration of parallel computation, long-range dependency modelling, and adaptive attention mechanisms enables robust performance in handling non-linear ageing trends. An experimental setup collected the voltage and current data using the hall sensors, low-pass filters and the 10-bit ADC microcontroller during the constant current charge-discharge cycles of an 18V 61.7F SC module. The RUL estimation results demonstrate the effectiveness of TCT, performing better than the state-of-the-art hybrid deep learning and machine learning models with improved prediction accuracy. The results also show its effectiveness in capturing degradation trends, making this a promising solution for SC health monitoring and RUL estimation.