IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 January 2025, accepted 26 January 2025, date of publication 6 March 2025, date of current version 21 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3549402

RESEARCH ARTICLE

Detailed Architectural Design of a Multi-Head Self-Attention Model for Lithium-Ion Battery Capacity Forecasting

JUYEON PARK[1], GYEONG HO LEE[2], JANGKYUM KIM[3], YOON-SIK YOO[4], AND IL-WOO LEE[5]

[1]Department of Data Science, Sejong University, Seoul 05006, Republic of Korea
[2]Department of Artificial Intelligence and Information Technology, Sejong University, Seoul 05006, Republic of Korea
[3]Department of Artificial Intelligence and Data Science, Sejong University, Seoul 05006, Republic of Korea
[4]Energy ICT Research Section, ETRI, Daejeon 34129, Republic of Korea
[5]Industrial and Energy Convergence Research Division, ETRI, Daejeon 34129, Republic of Korea

Corresponding authors: Gyeong Ho Lee (gyeongho@sejong.ac.kr) and Jangkyum Kim (jk.kim@sejong.ac.kr)

This work was supported in part by Korea Institute of Energy Technology Evaluation and Planning (KETEP) through the Ministry of Trade, Industry and Energy (MOTIE), Republic of Korea, under Grant 20202020900290; and in part by the Culture, Sports and Tourism Research and Development Program through Korea Creative Content Agency (KOCCA) Grant funded by the Ministry of Culture, Sports and Tourism (MCST), in 2025 (Project Name: Cultivating Masters and Doctoral Experts to Lead Digital-Tech Tourism, Contribution Rate: 50%), under Project RS-2024-00442006.

ABSTRACT As the adoption of lithium-ion batteries increases, concerns over safety incidents and replacement costs have become increasingly pressing. Accurate battery state prediction is thereby vital for reducing costs and ensuring safety. This paper introduces a novel system based on a transformer encoder architecture, leveraging Multi-Head Self-Attention (MHSA) layers to enhance battery capacity prediction. To address variability in battery data collection, we implement robust preprocessing techniques and a sliding window method to standardize data input. Positional encoding is applied to embed sequence order information at the input stage, while residual connections and layer normalization between MHSA layers optimize the learning process. Final predictions are generated through a fully connected neural network (FNN). Experimental results on the NASA dataset, obtained from repeated charge-discharge cycles, demonstrate that the proposed model mostly achieves superior predictive accuracy across multiple cell datasets, as measured by MAPE and RMSE. The proposed model showed enhanced performance, achieving a MAPE of 0.9918% and an RMSE of 0.02542 in experiments on specific cells. Furthermore, compared to other models in most experiments, it showed a performance improvement of at least 41% in MAPE and 29% in RMSE. Notably, the model maintains excellent performance while reducing complexity. Additionally, to prevent overestimation of overall performance due to high accuracy in only the early stages, we conducted interval-based performance evaluations, confirming that the proposed model consistently provides accurate predictions with minimal variance across different stages compared to other models.

INDEX TERMS Battery management system, energy storage system, electric vehicle, lithium-ion battery, multi-head self-attention, SoH prediction.

I. INTRODUCTION

The rising use of lithium-ion batteries in transportation and energy storage, driven by environmental and economic factors, has amplified concerns about safety and cost. Given their significant expense and associated risks, effective management and precise forecasting of battery lifespan are essential for minimizing costs and ensuring safety [1], [2], [3], [4]. Here, monitoring and predicting the State of Health (SoH) and Remaining Useful Life (RUL) are significant for

The associate editor coordinating the review of this manuscript and approving it for publication was Vitor Monteiro.

optimizing performance, preventing unforeseen failures and reducing operational costs. Such advanced predictive capabilities are crucial for informed decision-making throughout the battery's lifecycle, promoting environmentally responsible energy management and advancing sustainable development.

To accurately predict the state of lithium-ion batteries, two primary approaches are commonly leveraged: model-based and data-driven methods. Initially, model-based approaches [5], [6] involve creating mathematical models that reflect the physical and electrochemical characteristics of lithium-ion batteries to forecast their SoH and RUL. Although effective at capturing detailed chemical processes, model-based methods face several notable limitations. They require complex partial differential equations, which are computationally intensive and challenging to solve in real-time, especially under rapidly changing conditions. Furthermore, such model-based approaches often struggle to adapt to diverse operational environments, as they require extensive parameter tuning and lack flexibility. These constraints can thus limit their practical application in real-world battery systems, where conditions vary dynamically. On the other hand, data-driven approaches [7], [8] use regression or optimization techniques applied to data collected during battery charge-discharge cycles, which are often measured using the C-rate, a metric representing the rate of charge or discharge relative to the battery's nominal capacity. These methods eliminate the need for complex electrochemical models and enable real-time analysis of charge-discharge patterns, temperature fluctuations, and internal resistance to accurately estimate battery lifespan. Thus, data-driven methods facilitate precise real-time predictions that are often challenging for model-based approaches to achieve.

With advances in computing power, deep learning methods have become central to data-driven approaches. These techniques utilize large datasets to identify patterns, yielding more accurate predictions of battery states than traditional methods. For instance, Wu et al. [9] employed importance sampling with neural networks, while Kang et al. [10] developed an FNN-based model that integrated a cycle life model to enhance capacity predictions and mitigate the impact of battery degradation. However, FNNs faced challenges in capturing sequential patterns inherent in time-series data. To address this limitation, Recurrent Neural Networks (RNNs) were introduced, excelling in the analysis of time-series data. Advanced RNN architectures, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs), effectively manage long-term dependencies. Research by [11] showed that stacking GRUs not only improved predictive accuracy but also reduced computational complexity, while the work in [12] further optimized LSTM performance through refined data preprocessing and network configurations. Recently, the attention mechanism, popular in natural language processing, has introduced a novel approach to battery state prediction by focusing on critical data aspects, capturing complex temporal patterns, and enhancing accuracy. This paper uses Multi-Head Self-Attention (MHSA) as the core mechanism and experimentally verifies the impact of techniques such as positional encoding, layer normalization, residual connections, and sliding window size on performance, thereby providing a transparent and interpretable model. Additionally, by conducting segment-based prediction experiments, it prevents result distortion caused by initial prediction performance.

This paper is structured as follows: Section II reviews related work and contributions, Section III details the proposed methodology, Section IV evaluates performance, and Section V concludes the findings. Lastly, to enhance readability, a list of abbreviations is provided in Table 1.

II. RELATED WORK

Battery data is inherently complex, characterized by sequential patterns between consecutive samples [13], [14]. Recurrent models are particularly effective for predicting such time-series data. To overcome the limitations of traditional RNNs, various enhancements have been proposed. A notable advancement is the Bi-directional approach. For example, Bi-LSTM and Bi-GRU models, as discussed in [15] and [16], utilize information from both past and future time steps, thereby mitigating the information loss typically associated with unidirectional RNNs. Herein, single algorithms often fall short in providing both robust generalization and high predictive accuracy, leading to extensive research into hybrid models. Thus, the work in [17] introduced a hybrid approach combining state-space estimation with RNNs to address battery degradation. Similarly, the authors in [18] combined traditional statistical methods with machine learning, using Autoregressive Integrated Moving Average (ARIMA) for

TABLE 1. List of abbreviations.

Abbreviation	Description
LIB	Lithium-Ion Battery
EV	Electric Vehicle
ESS	Energy Storage System
BMS	Battery Management System
FNN	Fully Connected Neural Network
SOH	State of Health
EOL	End of Life
RUL	Remaining Useful Life
MLP	Multi-Layer Perceptron
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
MHSA	Multi-Head Self-Attention
MHSAP	Multi-Head Self-Attention for Prediction
Add & Norm	Residual Connections and Layer Normalization
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
MSE	Mean Squared Error

linear predictions and Bi-LSTM for capturing non-linear relationships. In another approach [19], they utilized Extreme Gradient Boosting (XGBoost) to select critical temporal features and feed them into Bi-LSTM, enhancing predictive performance.

Combining multiple deep learning techniques has proven effective for enhancing battery state prediction. For instance, the authors in [20] and [21] utilized a 1-D Convolutional Neural Network (CNN) to preprocess data, capturing nonlinear relationships between state measurements such as current and voltage, before feeding the data into an LSTM for prediction. Similarly, Mazzi et al. [22] combined 1-D CNN with Bi-GRU to capture bidirectional temporal dependencies more effectively. In [23], authors employed the CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) algorithm for efficient time-series decomposition and noise reduction prior to applying a CNN-Bi-LSTM model. These hybrid methods have notably enhanced learning efficiency and improved the detection of both global and local patterns within the data.

Recent research has increasingly used the attention mechanism to enhance battery capacity prediction. Often, as shown in [24], attention is applied after an LSTM layer to capture relationships between hidden states, thus improving performance. For example, Zhang et al. [25] first generated hidden states with an initial LSTM layer and then used concatenated attention to evaluate the significance of each hidden state, producing a weighted context vector that a subsequent LSTM layer utilizes for estimation. Similarly, Li et al. [26] employed multi-head attention mechanism with scaled dot-product attention in parallel after the LSTM layer, allowing multiple attention operations to further refine predictions.

Despite their high accuracy, recurrent models continue to face challenges with long-term memory retention. Furthermore, as highlighted in [21], combining multiple algorithms often increases computational costs and complicates model interpretation, and the benefits of hybrid models remain unclear. Additionally, the interplay between attention mechanisms and other methods is often not well understood. To address these issues, this paper proposes the use of Multi-Head Self-Attention (MHSA) as an alternative, avoiding the need for recurrent or hybrid structures. The necessary notations for the proposed methodology are fully provided in Table 2.

A. CONTRIBUTIONS OF THIS PAPER

In this paper, our contributions can be summarized as follows:

- This paper presents a model that employs Multi-Head Self-Attention (MHSA) as the core mechanism to reduce model complexity while achieving high performance.
- We clarify performance improvements often obscured by high parameter counts and complex structures in existing models by providing a detailed structural analysis.

TABLE 2. List of notations.

Notation	Description
Q	Query
K	Key
V	Value
d_k	Dimension of K
H_h	h-th attention output
μ	Mean of a specific feature X (e.g., Voltage)
σ	Standard deviation of a specific feature X
Z	Value of a specific feature X after z-score transformation
x	Dataset transformed into matrix form
z	x after min-max normalization
k	Look-back window size
S	Size of the 1D cycle profile
i	Column index of matrix z
m_l	Input of the l-th attention layer
M_l	Output of the l-th residual connection
γ, β	Weights used in layer normalization
W_Q	Weight for generating Q
W_K	Weight for generating K
W_V	Weight for generating V
$C(k)$	Actual capacity for the k-th cycle
$\hat{C}(k)$	Predicted capacity for the k-th cycle

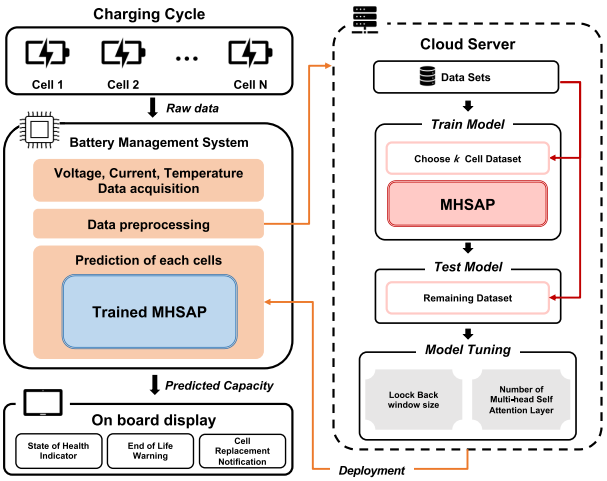
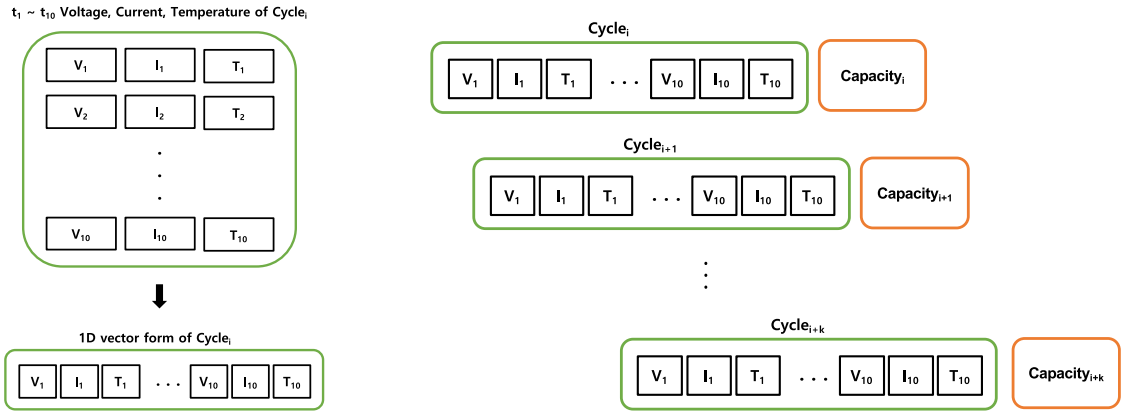


FIGURE 1. System architecture for predicting battery cell capacity.

- We incorporate robust data preprocessing techniques, including outlier removal, normalization, and methods that account for time-series characteristics, to ensure consistent performance across diverse battery datasets.
- We assess the impact of techniques, such as positional encoding, layer normalization, and residual connections, to identify the most effective combinations that enhance prediction accuracy.
- We verify that our optimal configuration delivers high performance throughout the battery lifecycle, ensuring reliable and consistent operation, particularly excelling in the late stages near the end of life (EoL).

III. PROPOSED METHODOLOGY

As illustrated in Fig. 1, throughout each charging cycle of an electric vehicle (EV) or energy storage system (ESS), the



(a) Charging cycle as 1D vector (b) Sequence construction through sliding window

FIGURE 2. Charging profile as 1D vector and sequence via sliding window.

Battery Management System (BMS) continuously monitors and records critical parameters such as voltage, current, and temperature. These raw measurements undergo preprocessing to align with the requirements of time-series data analysis before being securely transmitted to a cloud-based server. Utilizing a cloud server enables efficient handling of large data volumes from multiple distributed sources, supports real-time processing, and facilitates continuous re-training and validation of the advanced Multi-Head Self-Attention for Prediction (MHSAP) model. This setup typically allows the model to be dynamically updated and optimized with the latest data, ensuring robust predictive performance across diverse operational scenarios [27]. Additionally, the cloud infrastructure supports rigorous model optimization, including hyperparameter tuning and comprehensive performance evaluations, which would be computationally challenging to achieve on a local server without compromising processing speed and predictive accuracy. Given that the proposed methodology extensively employs the multi-head self-attention mechanism, a detailed overview of this technique is initially addressed to elucidate its application.

A. OVERVIEW OF MULTI-HEAD SELF ATTENTION

In this study, we propose a model that utilizes multi-head self-attention, a key mechanism in the transformer architecture as introduced by [28], to achieve high predictive performance and computational efficiency without relying on recurrent structures. This self-attention mechanism is specifically designed to measure the degree of association between various points within a sequence. The values derived from this mechanism are integral to the prediction process. The core mathematical formulation of this mechanism is expressed as

$$\text{Attention}\,(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \qquad (1)$$

where, Query (Q), Key (K), and Value (V) are obtained via matrix multiplication with the model's weight matrices W_Q, W_K, W_V, respectively. The dot product of Q and K is scaled by the square root of the dimension of K and is then transformed into a probability distribution via the softmax function. In this context, Q serves as the reference point for the computation, while K represents the elements against which Q is evaluated. The resulting distribution is multiplied by V associated with K, which influences the model's prediction output. When this mechanism is applied within a single sequence where Q, K, and V are derived from same dataset, it is referred to as self-attention. This approach allows the model to allocate greater weights to more pertinent K with respect to a given Q.

Now, multi-head self-attention involves executing the attention mechanism in parallel across multiple heads [29]. This approach is designed to capture a range of attention outcomes from the input sequence and is defined as

$$\text{Multi-Head}\,(Q, K, V) = \text{Concat}\,(H_1, \ldots, H_h)\,W_h, \qquad (2)$$

where H_h is given by

$$H_h = \text{Attention}\left(QW_Q^{(h)}, KW_K^{(h)}, VW_V^{(h)}\right). \qquad (3)$$

By employing this method, the model can capture associations missed by individual attention operations. Each head performs its own attention operation, and the results from all heads are concatenated and multiplied by the learned weight matrix W_h. In practice, multi-head self-attention is typically implemented by dividing the input matrices Q, K, and V into segments for each head, performing self-attention computations, and then concatenating the resulting attention values. Consequently, multi-head self-attention can be viewed as a matrix multiplication-based process, offering superior parallel processing capabilities compared to Recurrent Neural Networks (RNNs) [30]. Furthermore, it effectively isolates the influence of irrelevant data points from other time steps, focusing solely on the association between specific pairs of data points for prediction purposes.

B. TIME-SERIES DATA PREPROCESSING

Lithium-ion battery data is recorded at multiple measurement points throughout each charging cycle in general. Given that the duration of a charging cycle may differ based on user behavior and BMS charging strategies, it is imperative to standardize data from all cycles through meticulous preprocessing to ensure uniformity for model inputs.

Initially, to address potential data fluctuations due to measurement errors, outlier removal is performed. This involves eliminating certain data points to ensure smooth training. The outlier removal is conducted using the z-score method, also known as the standard score [31], [32], as defined in

$$Z = (X - \mu)/\sigma. \tag{4}$$

This method normalizes data using z-scores and removes outliers outside a specified range. Cycles with missing capacity records are excluded, and data points after a steady state are discarded. Subsequently, to standardize the data, we follow the approach detailed in [1], choosing a fixed number of measurement points from each cycle. Here, ten measurement points are chosen per cycle, and these values are assembled into a 1D vector of size 30, as illustrated in Fig. 2(a). This process is applied to all cycles, resulting in a dataset formatted as a matrix with dimensions.

Min-Max normalization is then applied to scale the matrix values to a range between 0 and 1, as described in [1] and [32]. This ensures consistency across datasets with varying value ranges. The min-max normalization is thereby defined as

$$z_r^c = \frac{x_r^c - \min(x)}{\max(x) - \min(x)}, \tag{5}$$

where $r \in \{1, \ldots, N\}$ and $c \in \{1, \ldots, 30\}$. Here, x is the dataset in matrix form, and x_r^c denotes the value in the r-th row and c-th column. N signifies the total number of cycles. After prediction, the values are transformed back to their original range using the scaling method applied during training. Following this, sequences required for prediction are constructed using the matrix, as depicted in Fig. 2(b). This process involves utilizing a sliding window approach, as referenced in [33] and [34], to select an appropriate look-back window size k. This method allows the construction of input data matrices of size (k, 30) for capacity prediction. This approach facilitates detailed analysis by capturing subtle variations in battery behavior, thereby enhancing prediction accuracy. The optimal look-back size will be determined through iterative experimentation in the subsequent stages of the research.

C. PROPOSED MULTI-HEAD SELF ATTENTION FOR PREDICTION (MHSAP)

The data obtained through the pre-processing progress is used as input for the model. In this study, the input data consist of voltage, current, and temperature readings collected during the battery charge-discharge cycles. These data are converted into time series format using a sliding-window technique and standardization before being applied to the model. Additionally, Positional Encoding is applied to capture the sequential characteristics of the data, enabling the model to learn while considering temporal context. Here, output is a predicted battery capacity for the next charge-discharge cycle, which can be utilized for battery performance evaluation and future performance prediction. This section provides a detailed explanation of the proposed model's structure.

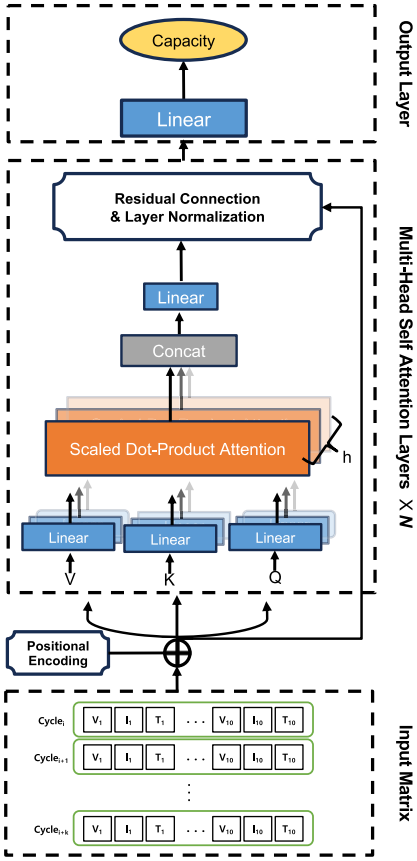


FIGURE 3. Architecture of the proposed model.

The architecture of the proposed model is detailed in Fig. 3. At its core, the model employs a series of layers built around multi-head self-attention, a pivotal mechanism from the transformer framework as outlined by [28]. This design choice ensures the model achieves superior predictive accuracy and computational efficiency, circumventing the need for recurrent structures. The number of layers is a critical hyperparameter, and performance evaluations have demonstrated significant variations with different layer configurations, necessitating a careful selection to optimize model performance.

Prior to feeding the input values into the first multi-head self-attention layer, the positional encoding technique introduced by [28] is applied. This technique incorporates sequential order information into the input data, an essential step given that the model lacks inherent mechanisms to interpret the order of input data as recurrent models do.

Positional encoding is thereby computed as

$$\mathrm{PE}\,(\mathrm{pos}, 2i) = \sin\left(\frac{\mathrm{pos}}{10000^{2i/S}}\right), \tag{6}$$

$$\mathrm{PE}\,(\mathrm{pos}, 2i+1) = \cos\left(\frac{\mathrm{pos}}{10000^{(2i+1)/S}}\right). \tag{7}$$

These two equation use periodic functions, sine and cosine, to generate a matrix of the same dimensions containing positional information, which is then added to the input matrix. Here, *pos* denotes the row index of the input matrix, i denotes the column index, and S is the size of the 1D cycle profile, which is 30 in this study. Following this, the input matrix is processed through the multi-head self-attention layer, where it is transformed into Q, K, and V using the layer specific weight matrices W_Q, W_K, W_V. These weight matrices are trained individually for each layer. As the model employs the multi-head self-attention mechanism, the Q, K, and V matrices are partitioned according to the number of attention heads, each performing attention operations independently before being concatenated. The results are then concatenated and passed through a linear layer to reduce dimensionality, with this process being iteratively applied across multiple multi-head self-attention layers. Between consecutive multi-head self-attention layers, we employ residual connections and layer normalization (Add & Norm) processes, as proposed by [28]. The residual connection adds the input of the layer to its output, defined by

$$M_l = m_l + \text{Multi-Head}(m_l) \tag{8}$$

This approach enhances training efficiency by focusing residual mapping on specific layers, mitigating the effects of initial weights initialized near zero and thereby accelerating the learning process [35], [36]. Layer normalization [37] is a normalization technique predominantly used for sequence data and is defined as

$$\text{Layer-Normalization}(M_{l,n}) = \gamma \hat{M}_{l,n} + \beta \tag{9}$$

where γ and β are learned parameters, and $\hat{M}_{l,n}$ represents the normalized value of the n-th input data $M_{l,n}$, adjusted for its mean and standard deviation. This method mitigates issues related to internal covariate shift by standardizing the distribution of activation. The normalization improves performance in multi-head self-attention mechanisms, as each head performs computations independently [38].

Ultimately, the output layer decodes the latent state vector into scalar through a fully connected (FC) layer, providing the final output of the model. A summary of the input and output data used to build the model is provided in Table 3.

IV. PERFORMANCE EVALUATION

In this section, we undertake a comprehensive evaluation and analysis of the proposed methodology by conducting an array of extensive experiments and comparisons with alternative approaches. We begin by detailing the experimental framework, including a description of the dataset, the performance metrics employed, and the comparative methodologies used. Following this, we present and rigorously analyze the experimental results to assess the efficacy and advantages of our proposed method.

TABLE 3. Description of input and output data used to build the model.

Component	Description
Input	Voltage (V), Current (A), Temperature (°C)
Data Preprocessing	Outlier removal (z-score method)
	10 measurement points selection per cycle
	Min-Max normalization
	Sliding window approach (optimal size: 5)
Positional Encoding	Sine and cosine functions for sequence order information
Multi-Head Self-Attention Layers	Multiple layers (number tuned as hyperparameter, optimal: 2)
	Parallel attention mechanism across multiple heads
	Query (Q), Key (K), and Value (V) matrices from input
Add & Norm	Residual connections
	Layer normalization
Output Layer	Fully connected layer for final prediction
Output	Scalar value of predicted battery cell capacity
Total Parameters	5,257

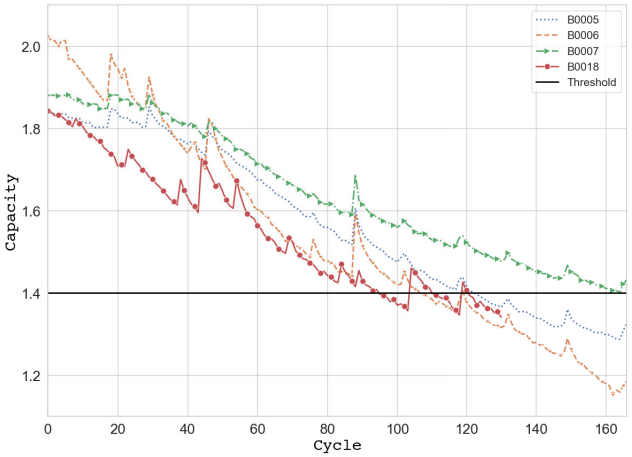


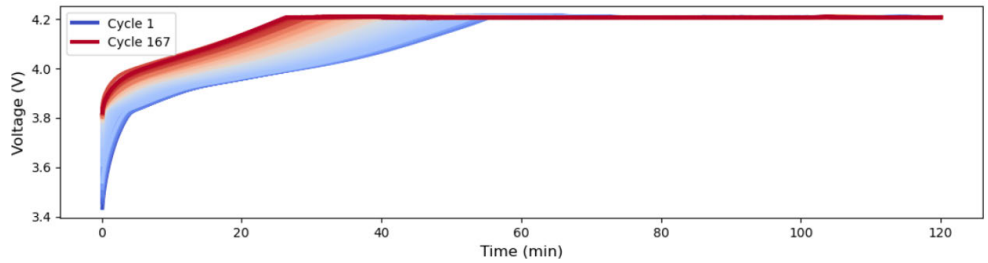
FIGURE 4. Changes in capacity for each battery cell.

A. DATA DESCRIPTION

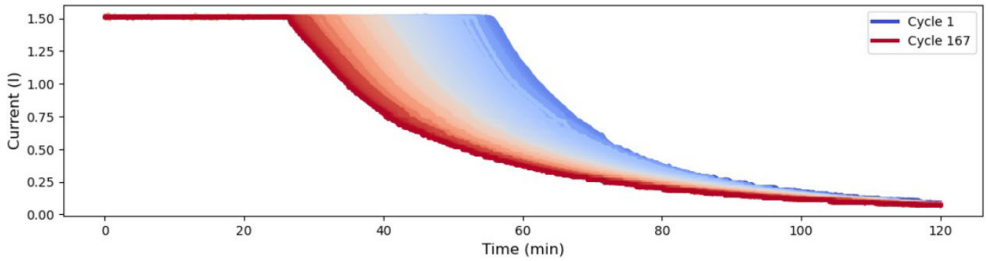
In this paper, the Li-ion battery aging datasets from NASA Open Data Portal were fully used, specifically the datasets labeled B0005, B0006, B0007, and B0018 [39].

Here, each dataset comprises cell profiles measured across multiple charging cycles. As shown in Fig. 4, the capacity degradation patterns of each cell vary distinctly, with B0018 cell showing a notably earlier cessation of measurement compared to the other cells.

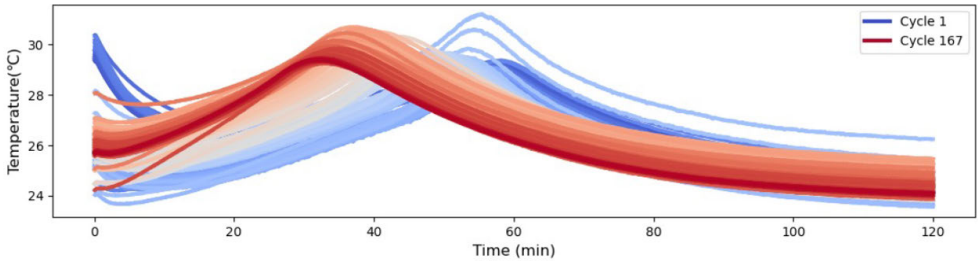
To mitigate the risk of overfitting to specific cells, we employed a K-fold cross-validation approach, wherein three of the four datasets were used for training and the remaining one was used for testing. For instance, when B0007 served as the test set, B0005, B0006, and B0018 were utilized for training. This procedure was repeated for each dataset, resulting in a total of four training and testing



(a) Voltage in initial and final stages



(b) Current in initial and final stage



(c) Temperature in initial and final stage

FIGURE 5. Cell B0005: voltage, current, and temperature in initial and final Stages.

configurations to robustly evaluate model performance. For capacity prediction, we employed profiles measured during the charging process, voltage (V), current (A), and temperature (°C). These measurements were obtained during the constant current/constant voltage (CC/CV) charging phase. During this phase, the battery is charged with a constant current of 1.5A until the voltage reaches 4.2V, at which point charging continues at a constant voltage until the current drops to 20mA. Fig. 5 illustrates the changes in voltage, current, and temperature as the battery ages. The observations indicate that aged batteries exhibit a faster drop in current, reach 4.2V more quickly, and attain peak temperatures sooner.

B. PERFORMANCE METRICS

To successfully evaluate cell capacity predictions, we employ two key metrics. These are commonly used metrics for evaluation [40]. Given that the predicted and actual values are relatively small, we utilize the Mean Absolute Percentage Error (MAPE) as one of our primary evaluation criteria. The MAPE is then defined as

$$\text{MAPE}(\%) = \frac{100}{K}\sum_{k=1}^{K}\frac{\left|C(k)-\hat{C}(k)\right|}{|C(k)|} \tag{10}$$

TABLE 4. Number of parameters for each model.

Model	Number of parameters
MHSAP	5257
LSTM	7471
Attention + LSTM	11197
MHSAP (No Positional Encoding)	5257
MHSAP (No Add & Norm)	5167

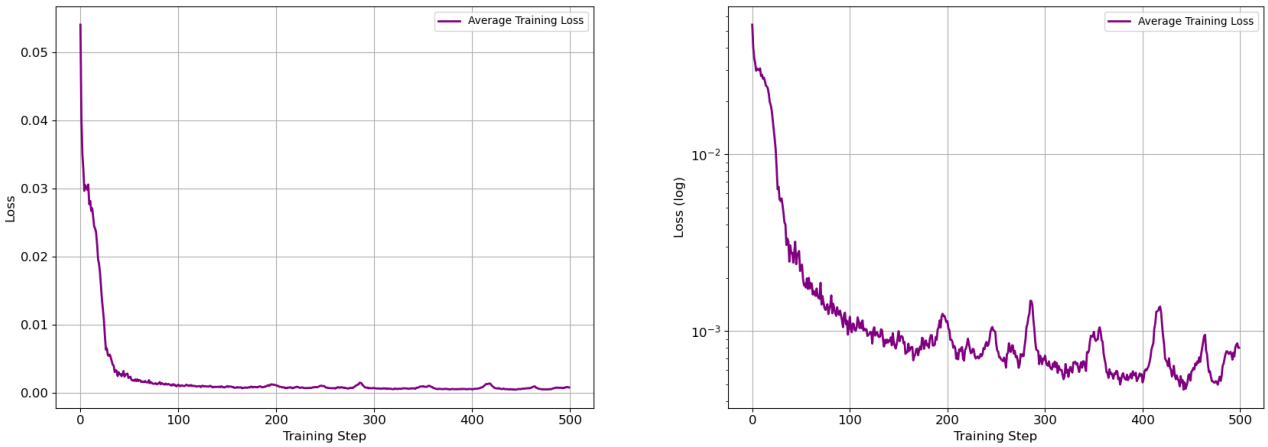
Additionally, we assess performance using the Root Mean Square Error (RMSE), which is given by

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(C(k)-\hat{C}(k)\right)^2} \tag{11}$$

In these two key metrics, $C(k)$ is the actual capacity for the k-th cycle, and $\hat{C}(k)$ is the predicted capacity for the same cycle. Both metrics provide a comprehensive assessment of the model's accuracy in predicting battery cell capacity.

C. COMPARISON METHODS

All models were implemented using the Python programming language, specifically with the PyTorch framework. Next,



(a) Average training loss (linear scale)

(b) Average training loss (log scale)

FIGURE 6. **Average training loss shown in original and log scales.**

TABLE 5. **Hyperparameter settings applied uniformly to all models.**

Hyperparameters	Settings
Training Parameters	Epochs: 500
	Batch size: 50
	Dropout rate: 0.5
Performance Metrics	Mean Absolute Percentage Error (MAPE)
	Root Mean Square Error (RMSE)
Loss Function	Mean Squared Error (MSE)
Optimizer	Adam (learning rate: 0.001)

to mitigate overfitting, dropout was applied to the main structure of each model, with a dropout rate set to 0.5. The loss function utilized was Mean Squared Error (MSE), and the optimizer employed was the Adam optimizer with a learning rate of 0.001. The training was conducted over 500 epochs with a batch size of 50. Fig. 6 shows the training loss reduction pattern that occurs during the training process of the MHSAP model for the experiments. A summary of the model parameters is provided in Table 4, while the hyperparameter settings, applied uniformly across all models, are outlined in Table 5.

1) LSTM

The first model employed for performance comparison is the Long Short-Term Memory (LSTM) [41], designed to address long-term dependency issues inherent in traditional RNNs. Compared to Multi-Layer Perceptrons (MLP) and Convolutional Neural Networks (CNN), LSTM demonstrates superior performance in battery capacity prediction tasks [1]. LSTM effectively manages long-term memory through its cell state and short-term memory through its hidden state, utilizing three types of gates: forget gate, input gate, and output gate. The forget gate regulates the retention or discarding of information in the cell state, the input gate controls the integration of the current input and previous hidden state into the cell state, and lastly the output gate adjusts the hidden state based on the current cell state to produce the output for the current time step. Similar to other models, the output layer of the LSTM includes a FC layer to transform the latent vector into a one-dimensional scalar value, producing the final prediction.

2) ATTENTION + LSTM

The second model for comparison integrates LSTM with a self-attention mechanism [24], [26]. This model sequentially processes input data through both LSTM and attention layers, performing attention operations on the LSTM's hidden states. Similar to other models, a FC layer is used to derive scalar predictions. This model leverages LSTM for encoding the input into a latent space, capturing positional information and local dependencies without the need for positional encoding [42]. The use of both LSTM and multi-head self-attention results in a relatively larger number of parameters.

3) MHSAP (NO POSITIONAL ENCODING)

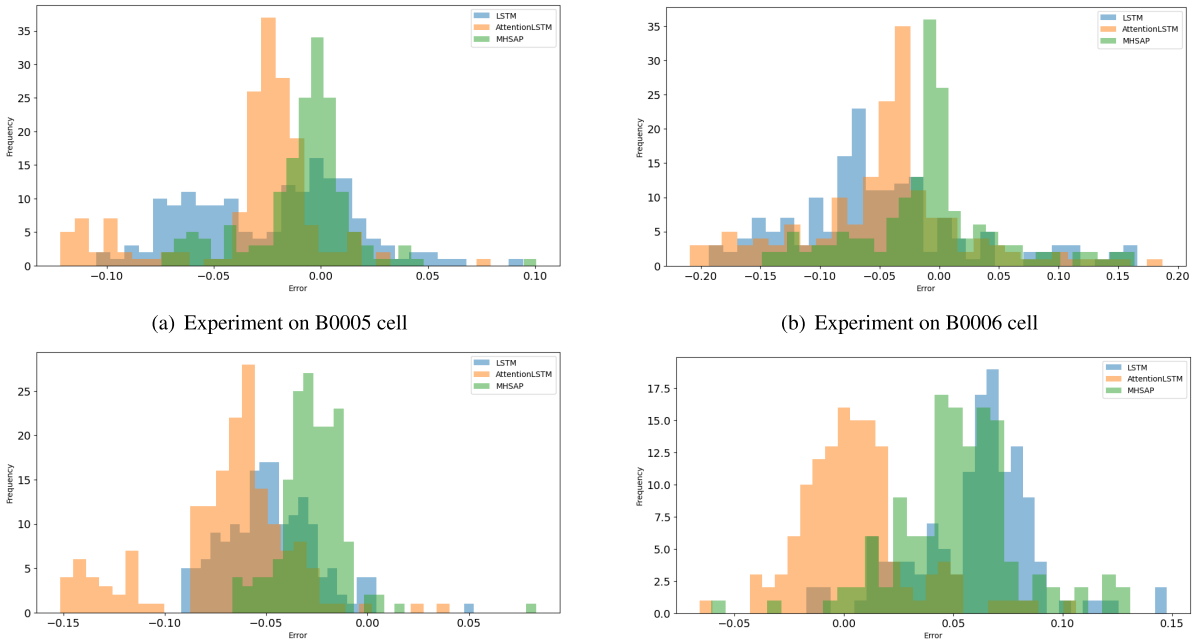
This model retains the core structure of the proposed model but excludes the positional encoding component. This comparison aims to assess the impact of injecting positional information versus not using it.

4) MHSAP (NO ADD & NORM)

This model is similar in structure to the proposed model but does not incorporate residual connections and layer normalization. This comparison helps evaluate the effects of including or excluding these techniques. Since residual connections and layer normalization involve additional learnable parameters, their absence reduces the total number of model parameters.

TABLE 6. Predicted capacity of individual battery cells.

B0005			**B0006**		
Model	MAPE(%)	RMSE	Model	MAPE(%)	RMSE
MHSAP	**0.9918**	**0.02542**	MHSAP	**2.7450**	**0.05879**
LSTM	2.1669	0.03961	LSTM	5.4042	0.09021
Attention+LSTM	2.1835	0.04779	Attention+LSTM	4.6832	0.08323
MHSAP (No Positional Encoding)	2.2908	0.04844	MHSAP (No Positional Encoding)	4.8884	0.08403
MHSAP (No Add & Norm)	1.2525	0.03419	MHSAP (No Add & Norm)	2.8314	0.06853
B0007			**B0018**		
Model	MAPE(%)	RMSE	Model	MAPE(%)	RMSE
MHSAP	1.7096	**0.03168**	MHSAP	3.5699	0.05993
LSTM	3.0469	0.05313	LSTM	3.9417	0.06576
Attention+LSTM	4.2978	0.07720	Attention+LSTM	**1.1696**	**0.02601**
MHSAP (No Positional Encoding)	3.0159	0.05163	MHSAP (No Positional Encoding)	7.6636	0.12349
MHSAP (No Add & Norm)	**1.4048**	0.03316	MHSAP (No Add & Norm)	1.8224	0.03316



(a) Experiment on B0005 cell

(b) Experiment on B0006 cell

(c) Experiment on B0007 cell

(d) Experiment on B0018 cell

FIGURE 7. Error histograms for capacity prediction models across 4 experiments.

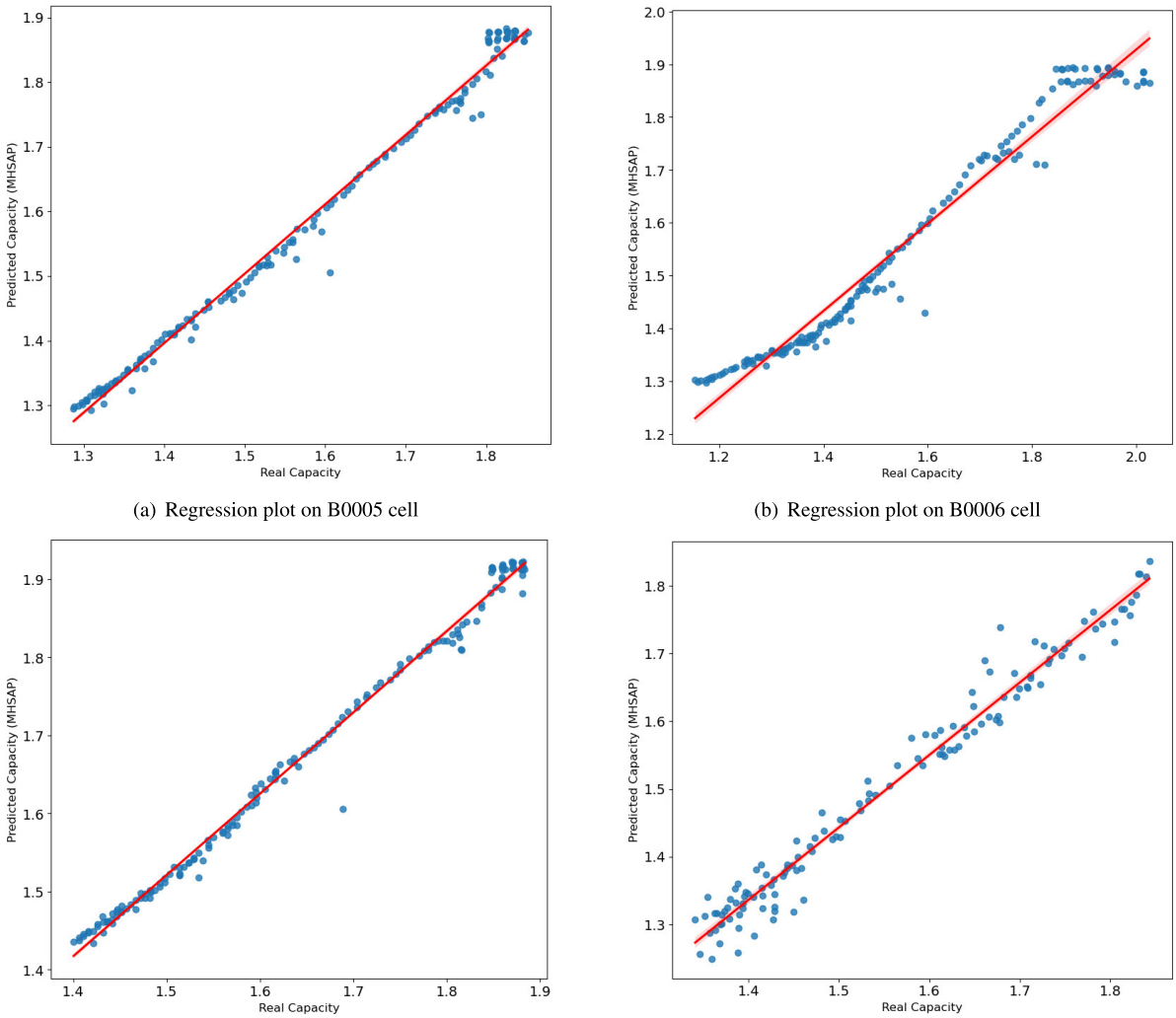
D. EXPERIMENTAL RESULTS AND DISCUSSION

This section provides a comprehensive analysis of the experimental results obtained from the proposed method against four alternative methods. The improvements are thoroughly examined using MAPE and RMSE metrics. A more detailed discussion of these results is provided below. This analysis not only highlights the relative strengths of each method but also offers insights into the effectiveness of various model components and their impact on predictive performance.

1) EFFECTIVENESS OF SUB-TECHNIQUES

As shown in Table 6, the proposed MHSAP model demonstrates superior performance in predicting the capacity of B0005 and B0006 batteries across both metrics. However, for the B0007 dataset, the model without residual connections and layer normalization achieves better MAPE performance. The Attention+LSTM model delivers the best results on the B0018 dataset, which is characterized by high volatility and the shortest cycle count among the datasets. As can be seen in the Error Histogram in Fig. 7, MHSAP's errors are most closely and densely distributed near zero for all data except B0018. In the case of B0018, the collected data was not only insufficient in quantity, but also found to contain some bias, resulting in limited impact on assessing the suitability of the overall model.

Despite having no learnable parameters, positional encoding significantly enhances performance in all cases when it



(a) Regression plot on B0005 cell

(b) Regression plot on B0006 cell

(c) Regression plot on B0007 cell

(d) Regression plot on B0018 cell

FIGURE 8. **Regression plot comparing predicted and actual values for MHSAP model across 4 experiments.**

is utilized, indicating that providing information about the sequence order and relative positions within the data greatly improves the model's ability to capture temporal dynamics and dependencies [43]. This means that by pre-embedding positional information, the model reduces the need to learn the order-related information from scratch, thereby lowering complexity and enhancing predictive performance under the same experimental conditions.

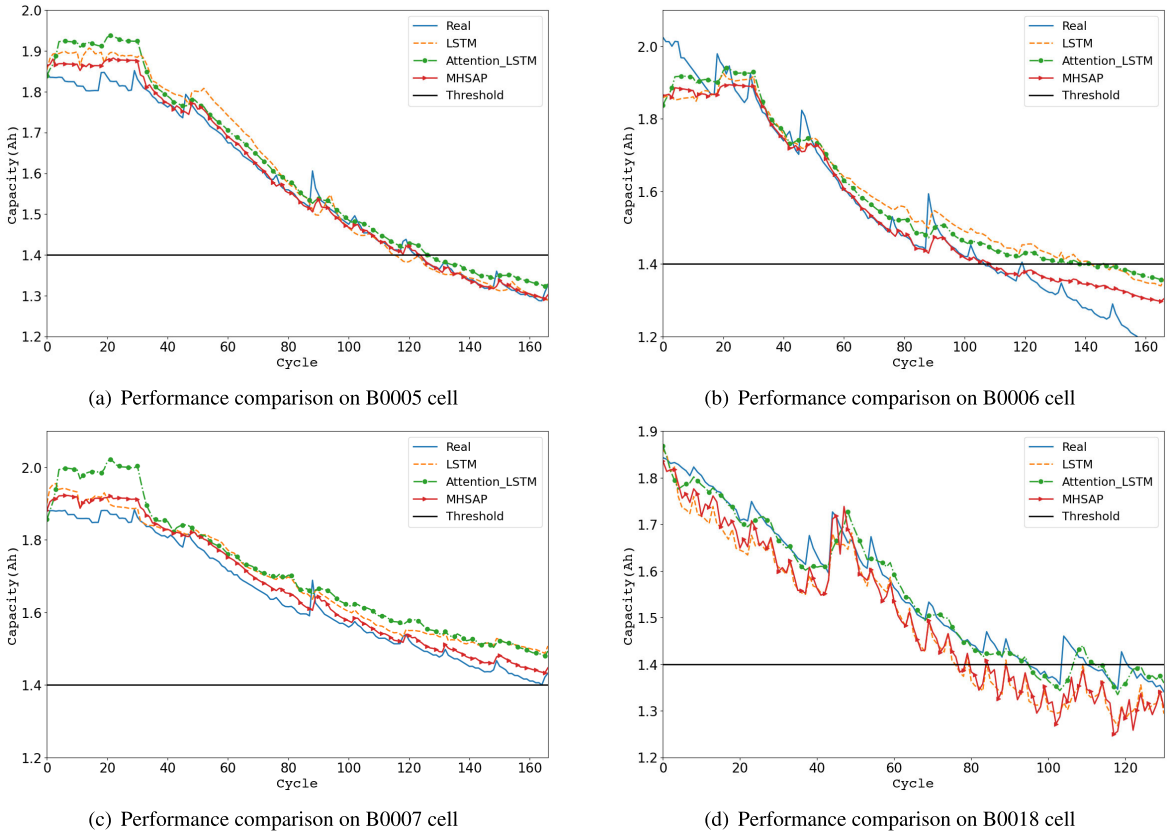
The Add & Norm technique, though its impact is comparatively minor, shows meaningful performance improvements in capacity predictions for the B0005, B0006, and B0007 cells. This suggests that maintaining consistent internal covariate shifts across layers and preventing the model from fully mapping each training example contributes to better learning efficiency. Consequently, the results confirm that the simultaneous use of both Positional Encoding and Add & Norm, is effective in enhancing the model's performance.

TABLE 7. **Performance variation across look-back window sizes.**

	2	3	4	5	6
MAPE(%)	1.4000	1.9680	1.1440	**0.9918**	1.2190
RMSE	0.0345	0.0434	0.0265	**0.0254**	0.0311

2) IN-DEPTH STRUCTURE CONFIGURATION

This study also investigated the impact of varying the look-back window size of the sliding window and the number of layers in the multi-head self-attention mechanism on model performance using B0005. The look-back window size was adjusted between 2 and 6, and the model's performance was compared and analyzed for each setting. As shown in Table 7, a look-back window size of 5 produced the most effective results, addressing that selecting an appropriate range of past data is crucial for predictive accuracy. Notably, increasing the look-back window size from 3 to 5 resulted in a significant reduction in MAPE and RMSE by 49.6%



(a) Performance comparison on B0005 cell

(b) Performance comparison on B0006 cell

(c) Performance comparison on B0007 cell

(d) Performance comparison on B0018 cell

FIGURE 9. Comparative performance of MHSAP, LSTM, and Attention+LSTM in four different cells.

TABLE 8. Performance with varying multi-head attention layers.

	1	2	3	4	5
MAPE(%)	2.3060	**0.9918**	1.4910	2.5010	2.8350
RMSE	0.0576	0.0254	**0.0251**	0.0436	0.0470

and 41.5%, respectively. Similarly, the number of multi-head self-attention layers was varied from 1 to 5. These layers, which require learnable parameters, increase the model's complexity as their number grows. The results in Table 8 revealed that two layers provided the optimal balance between model complexity and performance, as measured by the MAPE metric. This finding highlights that excessive complexity can lead to overfitting, thus degrading predictive performance. In Fig. 8, we present a comparison between the actual and predicted values for the MHSAP model when optimally configured. The results indicate that the proposed model learns effectively, exhibiting stable performance without significant bias or variance issues within the given dataset. Overall, tuning the look-back window size and the number of multi-head self-attention layers allows for the identification of optimal model settings, thereby enhancing predictive performance.

3) PREDICTION PERFORMANCE BY SEGMENT

Fig. 9 displays the visual comparison between actual values and predictions from each model. In this visualization, the threshold marks the capacity corresponding to the End of Life (EoL). The EoL represents the point where the battery's performance is no longer guaranteed, determined using the State of Health (SoH), which reflects the performance degradation relative to the rated capacity (C_0). It is computed as

$$\text{SoH}(\%) = \frac{C_k}{C_0} \times 100 \tag{12}$$

According to [1], a SoH of 70% or below is considered indicative of EoL. Accurate prediction of capacity near the EoL is critical for effective long-term battery management. To evaluate prediction performance across different usage stages, this study assesses the MAPE for distinct intervals. The B0005, B0006, and B0007 datasets were analyzed, while the B0018 dataset was excluded due to its shorter cycle length of 36 cycles compared to the others. Additionally, models lacking Positional Encoding and Add & Norm were omitted from consideration due to their inferior performance. The cycles were categorized into three intervals: early (0–55 cycles), mid (56–111 cycles), and late (112–167 cycles), with the late interval representing the EoL region. Table 9 summarizes the MAPE results for each interval, with the model achieving the highest prediction accuracy in each interval highlighted in bold. Although differences in model performance are minimal during the early phase, the proposed model demonstrates notably superior performance in the

TABLE 9. Comparison of capacity predictions across cycle segments.

Cell	Model	0∼55 Cycle	56∼111 Cycle	112∼167 Cycle
B0005	MHSAP	**1.863 %**	**0.654 %**	**0.449%**
	LSTM	2.251 %	1.561 %	2.698 %
	Attention+LSTM	3.551 %	1.336 %	1.654 %
B0006	MHSAP	2.313 %	**0.838 %**	**5.126 %**
	LSTM	2.594 %	4.022 %	9.674 %
	Attention+LSTM	**2.266 %**	2.531 %	9.336 %
B0007	MHSAP	**1.990 %**	**1.680 %**	**1.455 %**
	LSTM	2.004%	3.075 %	4.080 %
	Attention+LSTM	4.550 %	3.693 %	4.667 %

late phase, near EoL. Typically, the chemical and physical changes within a battery accumulate as the number of cycles increases, exerting nonlinear and complex effects on the battery's lifespan. Consequently, it is challenging to predict patterns accurately as the cycles progress, leading to a decline in prediction accuracy. However, the model proposed in this paper addresses such issues and demonstrates stable predictive performance even as the number of cycles increases. Specifically, the proposed method reduced MAPE by 83.4%, 47.0%, and 68.8% compared to the worst-performing method for the B0005, B0006, and B0007 datasets, respectively. A thorough review of Fig. 9 and Table 9 indicates that, with the exception of the early phase, the proposed model consistently provides precise predictions across all intervals, exhibiting minimal deviation relative to other models. This underscores the proposed model's robustness and reliability in predicting battery capacity, particularly as the battery approaches EoL.

V. CONCLUSION

In this paper, we presented the novel model called MHSAP, designed to enhance the accuracy of battery capacity prediction. Our comprehensive experimental evaluation demonstrated that the MHSAP model outperforms in predicting battery capacity, particularly in the challenging late phase near EoL. We also explored various configurations to optimize the model's performance. Initially, we employed a sliding window technique to determine the most effective window size for input data. Following this, we validated the benefits of Positional Encoding, which injects crucial sequential information into the model, enhancing its ability to capture temporal dynamics. Through systematic experimentation, we identified the optimal number of multi-head self-attention layers and confirmed the efficacy of incorporating residual connections and layer normalization between layers to bolster learning performance. Results from our performance metrics, including MAPE and RMSE, confirmed that the proposed MHSAP model mostly provides high accuracy across various datasets and usage stages. Notably, the model excelled in predicting capacity in the late phase, demonstrating its effectiveness in managing long-term battery health. In summary, the MHSAP model offers superior predictive performance by leveraging the advantages of multi-head self-attention mechanisms while circumventing the limitations associated with traditional RNNs, such as parallel processing inefficiencies and structural constraints. This advancement underscores the model's capability to deliver accurate predictions without the inherent drawbacks of RNN-based approaches. The findings of this study contribute to the development of more reliable and efficient battery management systems, paving the way for improved performance and longevity in battery-powered technologies.

As a future research, we plan to enhance the reliability and generalization performance of the model through the large-scale datasets, such as the 128 battery cell dataset provided by MIT. By constructing the model through the various battery cells, we aim to refine the model's performance evaluation under diverse conditions. Additionally, we expect that model could reflect the aging processes and degradation patterns of batteries. Through this approach, we could find a prediction scheme that comprehensively incorporates not only battery operation data but also external environmental factors such as temperature, humidity, and pressure, to reflect real-world operational environments.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

[1] Y. Choi, S. Ryu, K. Park, and H. Kim, "Machine learning-based lithium-ion battery capacity estimation exploiting multi-channel charging profiles," *IEEE Access*, vol. 7, pp. 75143–75152, 2019.

[2] S. Rout and S. Das, "Online state-of-charge estimation of lithium-ion battery using a fault tolerant and noise immune threefold modified adaptive extended Kalman filter," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 4, pp. 9366–9380, Dec. 2024.

[3] G. Crabtree, "The coming electric vehicle transformation," *Science*, vol. 366, no. 6464, pp. 422–424, Oct. 2019.

[4] S. Rout and S. Das, "A noise covariance regulated robust modified adaptive extended Kalman filter for state of charge estimation of lithium-ion battery," *IEEE Access*, vol. 12, pp. 78434–78448, 2024.

[5] Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala, "A model-based method for remaining useful life prediction of machinery," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1314–1326, Sep. 2016.

[6] M. Safari and C. Delacourt, "Simulation-based analysis of aging phenomena in a commercial graphite/$LiFePO_4$ cell," *J. Electrochemical Soc.*, vol. 158, no. 12, 2011, Art. no. A1436.

[7] X. Zheng and H. Fang, "An integrated unscented Kalman filter and relevance vector regression approach for lithium-ion battery remaining useful life and short-term capacity prediction," *Rel. Eng. Syst. Saf.*, vol. 144, pp. 74–82, Dec. 2015.

[8] J. C. Á. Antón, P. J. G. Nieto, C. B. Viejo, and J. A. V. Vilán, "Support vector machines used to estimate the battery state of charge," *IEEE Trans. Power Electron.*, vol. 28, no. 12, pp. 5919–5926, Dec. 2013.

[9] J. Wu, C. Zhang, and Z. Chen, "An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks," *Appl. Energy*, vol. 173, pp. 134–140, Jul. 2016.

[10] L. Kang, X. Zhao, and J. Ma, "A new neural network model for the state-of-charge estimation in the battery degradation process," *Appl. Energy*, vol. 121, pp. 20–27, May 2014.

[11] Y. Song, L. Li, Y. Peng, and D. Liu, "Lithium-ion battery remaining useful life prediction based on GRU-RNN," in *Proc. 12th Int. Conf. Rel., Maintainability, Saf. (ICRMS)*, Oct. 2018, pp. 317–322.
[12] K. Park, Y. Choi, W. J. Choi, H.-Y. Ryu, and H. Kim, "LSTM-based battery remaining useful life prediction with multi-channel charging profiles," *IEEE Access*, vol. 8, pp. 20786–20798, 2020.
[13] J. Chen, P. Kollmeyer, S. Panchal, Y. Masoudi, O. Gross, and A. Emadi, "Experimental results of battery power capability measurement on cells with different state of health levels," in *Proc. IEEE Transp. Electrific. Conf. Expo (ITEC)*, Jun. 2024, pp. 1–6.
[14] E. Samadani, S. Panchal, M. Mastali, and R. Fraser, "Battery life cycle management for plug-in hybrid electric vehicle (PHEVS) and electric vehicles (EVS)," Univ. Waterloo, Waterloo, ON, Canada, Tech. Rep., 2012, vol. 88.
[15] Z. Zhang, M. Xu, L. Ma, and B. Yu, "A state-of-charge estimation method based on bidirectional LSTM networks for lithium-ion batteries," in *Proc. 16th Int. Conf. Control, Autom., Robot. Vis. (ICARCV)*, Dec. 2020, pp. 211–216.
[16] J. Zhang, C. Huang, M.-Y. Chow, X. Li, J. Tian, H. Luo, and S. Yin, "A data-model interactive remaining useful life prediction approach of lithium-ion batteries based on PF-BiGRU-TSAM," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1144–1154, Feb. 2024.
[17] M. Catelani, L. Ciani, R. Fantacci, G. Patrizi, and B. Picano, "Remaining useful life estimation for prognostics of lithium-ion batteries based on recurrent neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
[18] Z. Wang, J. Qu, X. Fang, H. Li, T. Zhong, and H. Ren, "Prediction of early stabilization time of electrolytic capacitor based on ARIMA-Bi_LSTM hybrid model," *Neurocomputing*, vol. 403, pp. 63–79, Aug. 2020.
[19] R. R. Ardeshiri, M. Liu, and C. Ma, "Multivariate stacked bidirectional long short term memory for lithium-ion battery health management," *Rel. Eng. Syst. Saf.*, vol. 224, Aug. 2022, Art. no. 108481.
[20] X. Song, F. Yang, D. Wang, and K.-L. Tsui, "Combined CNN-LSTM network for state-of-charge estimation of lithium-ion batteries," *IEEE Access*, vol. 7, pp. 88894–88902, 2019.
[21] B. Zraibi, C. Okar, H. Chaoui, and M. Mansouri, "Remaining useful life assessment for lithium-ion batteries using CNN-LSTM-DNN hybrid method," *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4252–4261, May 2021.
[22] Y. Mazzi, H. B. Sassi, and F. Errahimi, "Lithium-ion battery state of health estimation using a hybrid model based on a convolutional neural network and bidirectional gated recurrent unit," *Eng. Appl. Artif. Intell.*, vol. 127, Jan. 2024, Art. no. 107199.
[23] X. Guo, K. Wang, S. Yao, G. Fu, and Y. Ning, "RUL prediction of lithium ion battery based on CEEMDAN-CNN BiLSTM model," *Energy Rep.*, vol. 9, pp. 1299–1306, Oct. 2023.
[24] X. Yan, G. Zhou, and H. Wang, "SOC prediction method of wireless sensor nodes batteries based on attention-LSTM," in *Proc. Int. Conf. Sens., Meas. Data Anal. era Artif. Intell. (ICSMD)*, Oct. 2021, pp. 1–6.
[25] J. Zhang, J. Hou, and Z. Zhang, "Online state-of-health estimation for the lithium-ion battery based on an LSTM neural network with attention mechanism," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Aug. 2020, pp. 1334–1339.
[26] X. Li and M. F. Dumlao, "State of charge estimation for lithium battery via attention and LSTM algorithm," in *Proc. Asia–Pacific Conf. Image Process., Electron. Comput. (IPEC)*, Apr. 2023, pp. 197–203.
[27] G. H. Lee, H. Park, J. W. Jang, J. Han, and J. K. Choi, "PPO-based autonomous transmission period control system in IoT edge computing," *IEEE Internet Things J.*, vol. 10, no. 24, pp. 21705–21720, Sep. 2023.
[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jun. 2017, pp. 5998–6008.
[29] Y. Lin, C. Wang, H. Song, and Y. Li, "Multi-head self-attention transformation networks for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 8762–8770, 2021.
[30] Q. Zhang, M. Ge, H. Zhu, E. Ambikairajah, Q. Song, Z. Ni, and H. Li, "An empirical study on the impact of positional encoding in transformer-based monaural speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 1001–1005.
[31] Z. Wang, J. Hong, P. Liu, and L. Zhang, "Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles," *Appl. Energy*, vol. 196, pp. 289–302, Jun. 2017.
[32] M. Mohammadrezaei, Z. Maleki, A. Tabesh, and S. A. Khajehoddin, "A framework for normalizing physical features of Li-ion batteries to form a generic health estimation model," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 3, pp. 6880–6892, Sep. 2024.
[33] J. Shi, M. Jain, and G. Narasimhan, "Time series forecasting (TSF) using various deep learning models," 2022, *arXiv:2204.11115*.
[34] M. Lin, D. Wu, S. Chen, J. Meng, W. Wang, and J. Wu, "Battery health prognosis based on sliding window sampling of charging curves and independently recurrent neural network," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–9, 2024.
[35] X. Wang, M. Xia, and W. Deng, "MSRN-informer: Time series prediction model based on multi-scale residual network," *IEEE Access*, vol. 11, pp. 65059–65065, 2023.
[36] 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.
[37] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
[38] H. Wang, Y. Song, X. Sun, S. Mo, C. Chen, and J. Wang, "Onboard in-situ warning and detection of Li plating for fast-charging batteries with deep learning," *Energy Storage Mater.*, vol. 71, Aug. 2024, Art. no. 103585.
[39] B. Sahaand and K. Goebel, "Battery data set, NASA Ames prognostics data repository," NASA Ames Res. Center, Mountain View, CA, USA, Tech. Rep., 2007.
[40] G. H. Lee, J. Han, and J. K. Choi, "MPdist-based missing data imputation for supporting big data analyses in IoT-based applications," *Future Gener. Comput. Syst.*, vol. 125, pp. 421–432, Dec. 2021.
[41] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[42] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1544–1551, Jul. 2018.
[43] S. A. A. Shah, S. G. Niazi, S. Deng, H. M. H. Azam, K. M. M. Yasir, J. Kumar, Z. Xu, and M. Wu, "A novel positional encoded attention-based long short-term memory network for state of charge estimation of lithium-ion battery," *J. Power Sources*, vol. 590, Jan. 2024, Art. no. 233788.

JUYEON PARK is currently pursuing the B.S. degree in data science with Sejong University, Seoul, South Korea. His academic interests include lithium-ion battery state prediction, machine learning, and time series forecasting.

GYEONG HO LEE received the B.S. degree in mechanical engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2015, and the M.S. degree in industrial and systems engineering (graduate minor in entrepreneurship and innovation) and the Ph.D. degree in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2019 and 2023, respectively. Previously, he was a Postdoctoral Researcher with the Information and Electronics Research Institute, KAIST. He is currently an Assistant Professor with the Department of Artificial Intelligence and Information Technology, Sejong University, Seoul, Republic of Korea. His current research interests include edge computing, data mining, artificial intelligence, and reinforcement learning.

JANGKYUM KIM received the B.S. and M.S. degrees from Sogang University, Seoul, South Korea, in 2015 and 2017, respectively, and the Ph.D. degree from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2022. From 2022 to 2023, he was a Data Scientist in optimization and analytics with SK Innovation. He is currently an Assistant Professor with the Department of Artificial Intelligence and Data Science, Sejong University, Seoul. His current research interests include energy management, energy economics, reinforcement learning, data imputation, data prediction, anomaly detection, biomechatronics, and big data analytic in healthcare.

YOON-SIK YOO received the B.S. and M.S. degrees in electronic engineering from Sungkyunkwan University, Suwon, Republic of Korea, in 1999 and 2001, respectively, and the Ph.D. degree from the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2020. Since 2022, he has been an Assistant Professor with the University of Science and Technology (UST), Daejeon. He joined the Electronics and Telecommunications Research Institute (ETRI), in 2001. He is currently a Principal Researcher with the Energy ICT Research Section. Since 2001, he has been conducting research and development on home networks, context-awareness, wireless LANs, smart grids, energy management systems, virtual power plants, and energy trading systems. His current research interests include carbon neutrality, energy ICT convergence, and data analysis.

IL-WOO LEE received the B.S. and M.S. degrees in computer science from Kyung Hee University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in computer science from Chungnam National University, Daejeon, South Korea, in 2007. He joined the Electronics and Telecommunications Research Institute (ETRI), in 1994, and has been engaged in the research and development of CDMA, TDX-10 ISDN, AIN, and home network systems. He is currently the Assistant Vice President of the division, who has been leading industrial and energy convergence research, since 2023. His research interests include energy efficiency, carbon neutrality, AIoT, autonomous manufacturing, and ICT convergence in agriculture, animal, and aquaculture.