Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 11 Aug 2026
Revised 07 Sep 2026
Accepted 26 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (23s): 955–971

KEYWORDS:
Lithium-ion battery; State of Charge; State of Health; Transformer; BiLSTM; Multi-task learning; Battery management system.

A Data-Driven Transformer–BiLSTM Multi-Task Learning Framework for Joint State-of-Charge and State-of-Health Estimation in Lithium-Ion Battery Energy Systems

Thirupathamma Sonti[1], Kudupudi Satya Ganesh[2], Pachikura Charan Teja[2], Sudheer Babu Devarapalli[1], Anuj P. Deshpande[1], Manjula R.[2], Rahul Das Gupta[2], Malay Jana[3], Sibendu Samanta[1]*

[1]Department of Electronics and Communication Engineering, SRM University–AP, Amaravati, Andhra Pradesh, India
[2]Department of Computer Science and Engineering, SRM University–AP, Amaravati, India
[3]Head – Battery Pack Design, Adani New Industries Ltd., India
*Corresponding author: sibendu.s@srmap.edu.in

ABSTRACT: Reliable operation of lithium-ion battery energy storage systems requires continuous, data-driven monitoring of two internal states that cannot be measured directly: State-of-Charge (SOC), which reflects instantaneous energy availability, and State-of-Health (SOH), which reflects long-term capacity degradation. Digital battery management systems increasingly rely on machine learning to extract these states from routinely logged voltage, current, and temperature measurements, yet most existing approaches estimate SOC and SOH independently, missing the shared temporal information embedded in a common measurement stream. This paper puts forward a Transformer–BiLSTM multi-task learning framework that jointly estimates SOC and SOH from a single shared representation of battery operating data: a linear embedding layer projects the three input channels into a 32-dimensional space, a two-layer Transformer encoder with four attention heads models dependencies across a 50-step measurement window, and a Bidirectional Long Short-Term Memory (BiLSTM) layer captures forward and backward temporal structure, producing a 64-dimensional representation shared by two task-specific prediction heads. The framework used here is trained and evaluated on the publicly available NASA lithium-ion battery dataset (cells B0005, B0006, B0007, and B0018; 636 discharge battery-cycle combinations), with SOC references generated via Coulomb counting and SOH references derived from smoothed capacity-degradation trajectories. On a held-out test partition, it reaches an SOC RMSE of 0.01659 (MAE 0.00901, R2 0.99592) and an SOH RMSE of 0.01821 (MAE 0.01205, R2 0.97095), with lower prediction error than ten baseline deep-learning architectures across both SOC and SOH estimation on all reported metrics, and notable differences confirmed under the adopted paired-window evaluation protocol using Wilcoxon signed-rank tests with Holm correction. These results show that a shared Transformer–BiLSTM representation gives an effective, computationally efficient data-driven pathway for joint battery-state monitoring, with a measured average inference processing time of 0.04564 ms per test sample under the adopted batched inference configuration. An additional target-battery few-shot adaptation study further shows that updating only 1.766% of the model's parameters matches or beats full-model fine-tuning when adapting to a specific battery with as little as 1% additional

Natural Resources for Human Health
eISSN: 2583-1194

data. The reported cycle-level train/validation/test partitioning strategy, and what it means for cross-battery generalization, are discussed as a point worth keeping in mind for deployment-oriented battery informatics systems. Keywords: Lithium-ion battery, State of Charge, State of Health, Transformer, BiLSTM, Multi-task learning, Data-driven energy systems, Battery informatics

1 INTRODUCTION

Because of their favourable combination of energy density, cycle life, and electrochemical stability, lithium-ion batteries (LIBs) now underpin much of the world's electrified transport, portable electronics, and grid-scale renewable storage. The ability to continuously and reliably infer a cell's internal condition from routinely logged operating data has become a core requirement for battery management systems (BMS), and for data-driven energy-system operation in general, as battery storage becomes further embedded within digital energy-management infrastructure [1, 2]. Two internal states matter most here, neither of which can be measured directly during normal operation: State-of-Charge (SOC), the available charge relative to usable capacity, and State-of-Health (SOH), the long-term degradation of that capacity. Safe operation, sound energy-management decisions, and condition-based maintenance all depend on estimating both accurately [3–5]. Battery-state estimation is hard precisely because the mapping from measurable signals, voltage, current, temperature, to the underlying states is strongly nonlinear and time-varying, and is shaped further by operating temperature, load profile, and cumulative aging [6, 7]. Classical approaches such as Coulomb counting, open-circuit-voltage methods, equivalent-circuit models, and Kalman-filter-based observers are cheap to compute, but their accuracy depends on well-chosen initial conditions and model parameters, both of which get harder to maintain as a battery ages and its dynamics drift [8–11].

Battery-monitoring systems now log far more voltage, current, and temperature data than before, and this has driven interest in data-driven and deep-learning methods that learn the signal-to-state mapping directly from experimental data instead of from an explicit electrochemical model. Recurrent architectures, especially LSTM and its bidirectional variant BiLSTM, are a natural fit here, since a battery's present state depends heavily on its recent operating history [12–15]. Transformer-based self-attention offers something extra on top of that: modelling dependencies between measurements that sit far apart in a sequence, which purely recurrent models struggle to do efficiently. SOC and SOH describe different timescales of battery behaviour, instantaneous charge against cumulative degradation, but both arise from the same electrochemical process and show up in the same voltage, current, and temperature signal, so training separate models for them throws away information the two tasks could share. Multi-task learning offers a natural way to exploit that sharing: a shared representation is learned from the common input stream, while separate output heads keep the flexibility each task's own behaviour needs. It is not trivial, though, because SOC can shift a lot within a single cycle while SOH changes gradually over hundreds of cycles, so a shared model has to capture both short-term dynamics and slow degradation trends without either task's signal washing out the other. A few studies point toward combining Transformer-based and recurrent architectures for this, but a framework that unifies global dependency modelling, bidirectional temporal processing, and multi-task learning in one place is still comparatively rare [17, 19–21]. This paper puts forward exactly such a framework, a Transformer–BiLSTM multi-task architecture for jointly estimating SOC and SOH from voltage, current, and temperature sequences. A linear embedding first lifts the three input channels into a higher-dimensional space, a two-layer multi-head-attention Transformer encoder then models dependencies across the input window, and a BiLSTM layer extracts bidirectional temporal structure from the result, with the shared representation feeding two task-specific heads, one for SOC and one for SOH. We evaluate the framework on the public NASA lithium-ion battery dataset against ten baseline deep-learning architectures spanning convolutional, recurrent, attention-based, temporal-convolutional, and Transformer-based designs.

The main contributions of this work are summarised as follows:

1. An integrated Transformer–BiLSTM multi-task architecture for joint, data-driven SOC and SOH estimation, combining attention-based dependency modelling with bidirectional temporal feature extraction in a single framework.

2. A shared multi-task learning strategy that learns common representations from voltage, current, and temperature measurements while retaining dedicated prediction heads for SOC and SOH.

3. A capacity-based SOH label-generation procedure, combining causal rolling averaging with exponential smoothing to build degradation-consistent training targets from the raw capacity trajectories of the evaluated cells.

4. An evaluation across four NASA battery cells against ten baseline deep-learning models on MSE, RMSE, MAE, and R2, supported by paired statistical significance testing.

5. An explicit discussion of the train/validation/test partitioning strategy and its implications for deployment-oriented battery-state estimation, offering a transparent basis for future cross-battery evaluation studies.

6. A parameter-efficient few-shot adaptation strategy that specializes the pretrained model to a specific target battery by updating only the SOC and SOH prediction heads (1.766% of total parameters), matching or beating full-model fine-tuning with as little as 1% additional target-battery data.

2 MATERIALS AND METHODS

2.1 NASA battery dataset

This work makes use of the NASA Ames Prognostics Center of Excellence (PCoE) lithium-ion battery repository, which records experimental voltage, current, temperature, and time measurements from cells subjected to repeated charge–discharge cycling under controlled laboratory conditions. Four cells from this repository — B0005, B0006, B0007, and B0018 were selected for the present joint SOC/SOH estimation task, chosen for their complete discharge-cycle records under consistent cycling conditions. The first three cells were each cycled through 168 discharge runs, while B0018 was cycled 132 times, giving 636 battery-cycle combinations in total. Voltage, current, and temperature form the three input features supplied to the proposed Transformer–BiLSTM framework, while SOC and SOH serve as the prediction targets.

2.2 Data extraction and preprocessing

The raw measurements obtained from the selected NASA battery cells are processed to extract the discharge-cycle data required for SOC and SOH estimation. For each battery cell, the recorded operating cycles are examined and only cycles identified as discharge cycles are retained. The extracted measurements include voltage, current, temperature, time, and discharge-capacity information. The discharge capacity associated with each cycle is used for generating the SOH reference values, whereas the measured current and time information is used for generating the SOC reference values. The processed battery records are afterward normalized and transformed into fixed-length temporal sequences for model training, validation, and testing.

2.3 SOC reference generation

The State of Charge (SOC) reference values are generated independently for each discharge cycle of the selected NASA lithium-ion batteries. SOC represents the remaining charge condition of the battery and is used as the reference target for the SOC estimation task. For each discharge cycle, the measured current and time information is used to calculate the charge removed from the battery. The initial SOC of each discharge cycle is set to 1. The SOC at a given sample is then calculated by integrating the magnitude of the measured current over time and normalizing the accumulated charge by the measured discharge capacity of the corresponding cycle. For a discharge cycle with capacity Q(k), the SOC at sample t is calculated as

$$SOC(t) = 1 - \frac{1}{Q(k)} \int_{t_0}^{t} |I(\tau)| d\tau \quad (1)$$

where I(τ) denote the measured battery current at time τ, Q(k) represents the measured discharge capacity of the corresponding cycle, and t0 denotes the beginning of the discharge cycle. In the implementation, the integration is performed numerically over successive measurements using the corresponding time intervals. The calculated SOC values are

constrained to the range [0, 1] to maintain physically meaningful reference values. This procedure is repeated independently for every discharge cycle and for each selected battery cell. The generated SOC reference values are aligned with the corresponding voltage, current, and temperature measurements. So, each temporal input sequence has an SOC target corresponding to its final time step. The resulting SOC references are afterward used together with the capacity-based SOH references to construct the multi-task input-target sequences.

2.4 SOH reference generation

The State of Health (SOH) reference values are generated from the discharge-capacity degradation characteristics of the selected NASA battery cells. SOH represents the remaining health condition of a battery relative to its initial healthy capacity and is used as the reference target for the SOH estimation task. For each battery cell, the extracted discharge measurements are organized according to their corresponding discharge cycles. The capacity value associated with each discharge cycle is used to

represent the available discharge capacity of that cycle. The cycle-level capacity values are processed independently for each battery cell to preserve the individual degradation characteristics of the batteries. To reduce short-term fluctuations in the measured capacity, a causal rolling mean with a window size of three cycles is first applied to the cycle-level capacity sequence. The resulting capacity sequence is then further smoothed using an exponentially weighted moving average (EWMA) with a smoothing factor of α = 0.3. The EWMA is defined as

$$\tilde{Q}(k) = \alpha Q_{roll}(k) + (1-\alpha)\tilde{Q}(k-1) \quad (2)$$

where $\tilde{Q}(k)$ denotes the smoothed capacity at cycle k, $Q_{roll}(k)$ represents the capacity obtained after applying the three-cycle causal rolling mean, and α denotes the EWMA smoothing factor. The initial healthy capacity is determined from the mean of the first five smoothed discharge-cycle capacities. The SOH value at cycle k is subsequently calculated as

$$SOH(k) = \frac{\tilde{Q}(k)}{Q_{initial}} \quad (3)$$

where $Q_{initial}$ denotes the initial healthy capacity obtained from the first five smoothed cycles. The calculated SOH values are constrained to the range [0, 1] to maintain physically meaningful health-state values. The resulting SOH value is a cycle-level reference and is assigned to all measurement samples belonging to the corresponding discharge cycle. Thus, all samples within a particular discharge cycle share the same SOH reference, whereas the SOC reference varies according to the measurement time within the cycle. The complete SOH reference-generation procedure is performed independently for each of the selected battery cells. This preserves the individual capacity degradation trajectories of B0005, B0006, B0007, and B0018. The generated SOH references are subsequently associated with the corresponding voltage, current, and temperature measurements for constructing the input-target sequences used in the simultaneous SOC and SOH estimation framework.

2.5 Feature normalization

To ensure that the input features have comparable numerical ranges, min–max normalization is applied to the voltage, current, and temperature measurements. The normalization parameters are learned exclusively from the training data to prevent information from the validation and test sets from influencing the preprocessing stage. For a feature value x, the normalized value is calculated as

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4)$$

where $x_{min}$and x_{max} denote the minimum and maximum values computed from the training set for the corresponding feature. The same training-derived parameters are subsequently applied to the validation and test data. Only the voltage, current, and temperature features are normalized. The SOC and SOH values are retained as prediction targets. The fitted scaler is stored for consistent preprocessing during model evaluation an inference.

2.6 Temporal sequence construction

After the battery-cycle-level partitioning and feature normalization, fixed-length temporal sequences are generated independently within each battery-cycle combination. A sliding window of 50 consecutive measurement samples is used to construct each input sequence. Each sequence therefore contains three input features, namely voltage, current, and temperature, resulting in an input dimension of 50 × 3. For a sequence beginning at sample index t, the input is represented as

$$X_t = [x_t, x_{t+1}, \ldots, x_{t+49}], X_t \in \mathbb{R}^{50\times 3} \quad (5)$$

where xt denotes the three-dimensional feature vector at time step t. The corresponding SOC and SOH targets are taken from the last sample of the sequence. So,

$$y_t^{SOC} = SOC_{t+49},\ y_t^{SOH} = SOH_{t+49}. \quad (6)$$

The sliding windows are generated only within individual discharge cycles, and no sequence is allowed to cross a battery-cycle boundary. This preserves the temporal structure of each discharge cycle and prevents information from adjacent cycles from being combined within a single input sequence. Sequence generation is performed separately for the training, validation, and test partitions after the battery-cycle split. The resulting datasets contain 107,539 training sequences, 23,321 validation sequences, and 23,697 test sequences.

Natural Resources for Human Health
eISSN: 2583-1194

3 PROPOSED TRANSFORMER–BILSTM MULTI-TASK LEARNING FRAMEWORK

3.1 Overall architecture

The framework used here is built to perform simultaneous State of Charge (SOC) and State of Health (SOH) estimation from the same sequence of battery measurements, combining a linear input embedding layer, a Transformer encoder, and a Bidirectional Long Short-Term Memory (BiLSTM) network within a shared feature-learning structure, followed by task-specific prediction heads for SOC and SOH. The input consists of fixed-length sequences of voltage, current, and temperature measurements; for a sequence length of 50 time steps, this is $X \in R50\times3$. Each time step's three-dimensional feature vector is first projected into a 32-dimensional representation through a linear embedding layer, and the embedded sequence then goes through a two-layer Transformer encoder with four attention heads, which learns dependencies among observations across the temporal sequence. This Transformer representation is afterward passed to the BiLSTM component, which processes it in both forward and backward temporal directions using a hidden size of 32 per direction, so the output at the final time step forms a 64-dimensional shared representation carrying information from both directions. That shared representation then goes to two task-specific prediction branches, the first estimating SOC and the second estimating SOH, each implemented as a multilayer perceptron (MLP). This lets the two tasks keep separate prediction mappings while still sharing the learned feature representation. The overall processing flow of the framework used here is expressed as

$$X \rightarrow Linear\ Embedding \rightarrow Transformer \rightarrow BiLSTM \rightarrow Shared\ Representation \rightarrow \{SOC\ Head, SOH\ Head\} \quad (7)$$

3.2 Input feature representation

The framework used here receives the pre-processed battery measurements as a multivariate temporal sequence. At each time step, three measurable battery variables, namely voltage, current, and temperature, are used to represent the instantaneous operating condition of the battery. The input vector at time step t is defined as

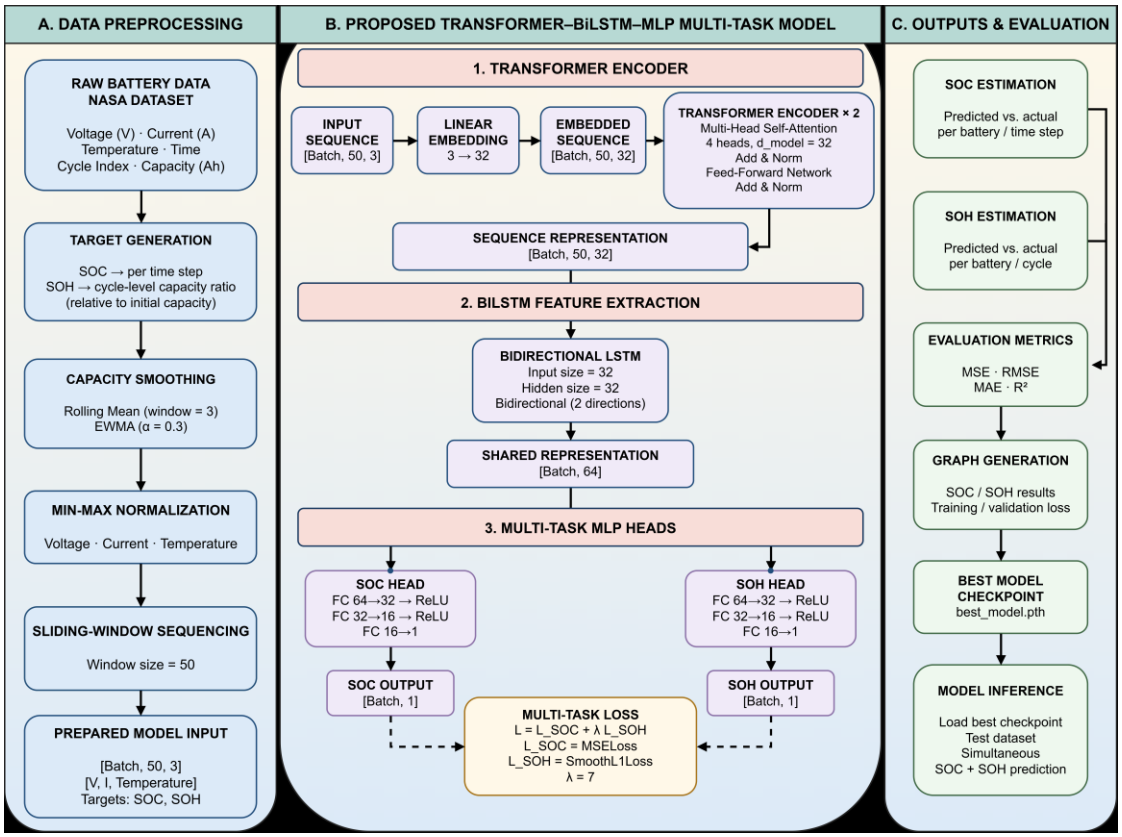


Fig. 1 Overall architecture of the proposed Transformer–BiLSTM multi-task learning framework for simultaneous SOC and SOH estimation.

$$x_t = [V_t, I_t, T_t] \quad (8)$$

where Vt, It, and Tt denote the voltage, current, and temperature measurements, respectively. For a sequence length of 50-time steps, the complete input sequence is represented by

$$X = [x_1, x_2, \ldots, x_{50}] \in \mathbb{R}^{50\times 3} (9)$$

The three-dimensional feature vector at each time step is projected into a 32-dimensional representation using a linear embedding layer. The embedding operation can be expressed as

$$e_t = W_e x_t + b_e,\ e_t \in \mathbb{R}^{32} (10)$$

where We and be denote the learnable weight matrix and bias of the linear embedding layer, respectively. The resulting embedded sequence preserves the temporal ordering of the original measurements and is supplied to the Transformer encoder for subsequent feature extraction. Using the same measured variables for both estimation tasks provides a

common information source for the multi-task learning framework, allowing the shared Transformer and BiLSTM components to learn representations before the task-specific SOC and SOH prediction heads generate their respective outputs.

3.3 Transformer encoder

The Transformer encoder is used after the linear input embedding to learn dependencies among observations across the complete input sequence. The embedded input has a feature dimension of 32, and the Transformer encoder is configured with a model dimension of 32, four attention heads, and two encoder layers. Self-attention within each encoder layer operates by projecting the embedded sequence E onto three learned subspaces queries, keys, and values — via

$$Q = EW_Q,\ K = EW_K,\ V = EW_V (11)$$

where WQ, WK, and WV are learnable projection matrices. Attention weights are then obtained by scaling the query-key dot product by √dk before applying softmax, and the resulting weights combine the value vectors as

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V (12)$$

where dk denotes the dimensionality of the key representation. Repeating this computation across four parallel heads lets the encoder attend to different positional relationships within the input sequence simultaneously and stacking two such layers progressively refine the sequence representation before it is passed to the temporal (BiLSTM) stage. No explicit positional encoding is added before the Transformer encoder in the implemented framework. Temporal ordering is afterward modeled by the BiLSTM, which processes the Transformer-derived sequence representation in forward and backward directions. Through this operation, the Transformer encoder captures dependencies among voltage, current, and temperature measurements across the 50-step temporal window. The resulting sequence representation retains the temporal dimension and is afterward passed to the BiLSTM component for bidirectional temporal processing.

3.4 Bidirectional Long Short-Term Memory

Following Transformer-based feature extraction, a Bidirectional Long Short-Term Memory (BiLSTM) network is used to capture temporal dependencies in both forward and backward directions. The BiLSTM receives the 32-dimensional sequence representation produced by the Transformer encoder and uses a hidden size of 32 for each direction. For an input sequence, the forward LSTM processes the sequence from the beginning toward the end, while the backward LSTM processes the same sequence in the opposite direction. Their hidden representations are combined to form the bidirectional representation as

$$h_t = [\overrightarrow{h}_t; \overleftarrow{h}_t] (13)$$

where $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ denotes the forward and backward hidden states, respectively. Since the hidden size is 32 for each direction, the resulting BiLSTM representation has a dimensionality of 64. The BiLSTM therefore provides a temporal representation that considers information from both directions of the input sequence. The output at the final time step is used as the shared 64-dimensional representation for the two task-specific prediction branches. This shared representation combines the dependencies learned by the Transformer with the bidirectional temporal information captured by the BiLSTM and is afterward provided to the SOC and SOH prediction heads.

3.5 Shared feature representation

The final hidden representation from the BiLSTM serves as a shared feature representation for both SOC and SOH estimation tasks. The shared representation is directly provided to two task-specific multilayers perceptron (MLP) heads, one for SOC estimation and the other for SOH estimation. So, the Transformer and BiLSTM layers constitute the shared feature extraction component, whereas the subsequent MLP branches independently transform the shared 64-dimensional representation into the corresponding SOC and SOH predictions.

3.6 State-specific prediction heads

The shared 64-dimensional representation obtained from the BiLSTM is provided to two independent multilayer perceptron (MLP) heads for task-specific prediction. Each head follows the same fully connected structure, consisting of layers with dimensions 64, 32, 16, and 1. The final scalar output of each head represents the estimated battery state. For SOC estimation, the first MLP head maps the shared representation to the predicted state-of-charge as

$$\widehat{SOC} = f_{SOC}(h_{shared}) \quad (14)$$

where h shared ∈R64 denotes the shared BiLSTM representation. Similarly, the second MLP head produces the predicted state-of-health as

$$\widehat{SOH} = f_{SOH}(h_{shared}) \quad (15)$$

The two prediction heads allow the model to learn task-specific mappings while retaining the common temporal representation learned by the Transformer–BiLSTM

feature extraction layers. The resulting scalar outputs are used to compute the corresponding SOC and SOH losses during multi-task training.

3.7 multi-task learning strategy

The framework used here adopts a multi-task learning strategy to estimate SOC and SOH simultaneously from a shared temporal representation. The Transformer encoder and BiLSTM layers are jointly optimized for both tasks, while the state-specific MLP heads generate the corresponding SOC and SOH predictions. By sharing the feature extraction layers, the model learns a common representation from voltage, current, and temperature measurements that is used by both estimation tasks. During training, the SOC and SOH objectives are optimized jointly using a weighted loss function. The SOH loss is assigned a weighting factor of 7 in the final implementation to control its contribution to the overall multi-task objective.

3.8 Joint loss function

The proposed multi-task framework jointly optimizes SOC and SOH estimation using a weighted combination of task-specific loss functions. Mean Squared Error (MSE) is used for SOC estimation, while Smooth L1 loss is used for SOH estimation. The overall training objective is defined as

$$L_{total} = L_{SOC} + \lambda_{SOH} L_{SOH},\ \lambda_{SOH} = 7 \quad (16)$$

where LSOC and LSOH denote the SOC and SOH losses, respectively, and λSOH controls the contribution of the SOH objective. In the final implementation, λSOH is set to 7. The SOC loss is calculated using Mean Squared Error as

$$L_{SOC} = \frac{1}{N}\sum_i \left(\widehat{SOC}_i - SOC_i\right)^2 \quad (17)$$

where N is the number of samples, $\widehat{SOC}_i$ is the predicted SOC, and SOC_i is the corresponding reference value. For SOH estimation, Smooth L1 loss is employed to provide a robust regression objective. For an error

$$e_i = \widehat{SOH}_i - SOH_i \quad (18)$$

the loss is expressed as

$$L_{SOH} = \frac{1}{N}\sum_i SmoothL1(e_i) \quad (19)$$

The weighted joint objective enables simultaneous optimization of both battery-state estimation tasks while assigning a higher contribution to the SOH objective through the weighting factor of 7.

4 EXPERIMENTAL SETUP

4.1 Dataset and experimental configuration

The experiments used the NASA Ames Prognostics Center of Excellence (PCoE) lithium-ion battery dataset, with four battery cells considered: B0005, B0006, B0007, and B0018. The first three cells have 168 discharge cycles each, while B0018 has 132, giving 636 discharge cycles in total, and only discharge-cycle measurements were used. For each discharge cycle, voltage, current, and temperature were the input features; the reference SOC came from Coulomb counting, while the reference SOH was derived from the smoothed discharge-capacity degradation (preprocessing and reference-target generation are described in Section 2). The dataset was split at the battery-cycle level into training, validation, and test subsets at a 70%, 15%, 15% ratio, before temporal sequence construction, so samples from the same battery cycle would not end up spread across different subsets. With a fixed random seed of 42, this partition gave 444 training cycles, 96 validation cycles, and 96 test cycles. This produced 107,539 training sequences, 23,321 validation sequences, and 23,697 test sequences. Feature normalization used MinMax scaling, with the scaling parameters estimated only from the training data and then applied unchanged to the validation and test sets, which avoids information leakage from the evaluation subsets into model training.

4.2 Baseline models

Ten established deep-learning architectures spanning convolutional, recurrent, attention-augmented, temporal-convolutional, and pure Transformer designs were selected as baselines, reflecting the range of approaches commonly applied to sequential battery-state estimation. This allows the contribution of the proposed Transformer– BiLSTM multi-task design to be assessed relative to established alternatives under identical data and training conditions. The evaluated baseline models were: 1. Convolutional Neural Network (CNN), 2. Gated Recurrent Unit (GRU), 3. Long Short-Term Memory (LSTM), 4. Bidirectional Long Short-Term Memory (BiLSTM),

5. LSTM with attention (LSTM-Attention), 6. CNN-BiLSTM, 7. CNN-LSTM, 8. CNN-LSTM-Attention, 9. Temporal Convolutional Network-BiLSTM (TCN-BiLSTM), 10. Transformer. Every baseline was trained and evaluated on the identical pre-processed sequences and identical train/validation/test partitions used for the model used here, so that any observed performance gap reflects the choice of architecture rather than differences in data handling.

4.3 Training configuration

The model used here and baseline models were trained under a common experimental configuration. The batch size was set to 128, the learning rate was 0.0005, and the models were trained for 30 epochs using the Adam optimization algorithm. The same training configuration was maintained across the evaluated architectures to provide a consistent basis for comparison. For the proposed multi-task model, the joint objective combined the SOC and SOH losses as described in Eq. 16. The SOC task was optimized using mean squared error, whereas the SOH task used the Smooth L1 loss. A weighting factor of 7 was assigned to the SOH loss to balance the contributions of the two prediction tasks. The model was trained using the training subset, while the validation subset was used for model evaluation during training. The held-out test subset was used only for the final performance comparison.

4.4 Evaluation metrics

Four standard regression metrics are used to quantify prediction accuracy for both SOC and SOH: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R2). Given reference values yi and model predictions ˆyi over N test samples, MSE penalizes larger errors quadratically as

$$MSE = \frac{1}{N}\sum_i (y_i - \hat{y}_i)^2 \quad (20)$$

and RMSE reports this error in the original units of the estimated state by taking its square root,

$$RMSE = \sqrt{\frac{1}{N}\sum_i (y_i - \hat{y}_i)^2} \quad (21)$$

MAE instead averages the absolute deviation between prediction and reference,

$$MAE = \frac{1}{N}\sum_i |y_i - \hat{y}_i| (22)$$

giving a metric that is less sensitive to occasional large outliers than MSE or RMSE. Finally, R2 compares the residual sum of squares against the total variance of the reference values,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} (23)$$

where $\bar{y}$ denotes the mean of the reference values, with values closer to 1 indicating a better overall fit. Across all four metrics, smaller MSE, RMSE, and MAE together with a larger R^2 jointly signal more accurate state estimation.

4.5 Statistical analysis

To assess whether the observed accuracy gap between the framework used here and each baseline is more than chance variation, the Wilcoxon signed-rank test is applied to the paired absolute prediction errors of the two models being compared. Running this test across ten baselines for each task raises the risk of false positives from multiple comparisons, so the resulting p-values are adjusted with the Holm correction before significance is judged. Complementing this, 95% confidence intervals for the mean absolute-error gap are obtained via bootstrap resampling with 10,000 iterations, giving an interval-based view of the effect size alongside the hypothesis test. Because the sequences are built from overlapping sliding windows, adjacent test windows are not fully independent; the reported significance levels are therefore best read as descriptive of the evaluated test-window population rather than as formal guarantees at the level of independent battery cycles.

5 RESULTS AND DISCUSSION

5.1 Quantitative performance comparison

The performance of the proposed Transformer–BiLSTM multi-task learning framework was evaluated on the held-out NASA test set and compared with ten baseline deep-learning models. The evaluation was performed separately for State of Charge (SOC) and State of Health (SOH) estimation using MSE, RMSE, MAE, and R2.

Table 1 summarizes the complete quantitative comparison. The model used here reaches an SOC MSE of 0.000275, an RMSE of 0.016589, an MAE of 0.009012, and an R2 of 0.995919. For SOH estimation, the framework used here reaches an MSE of 0.000331, an RMSE of 0.018207, an MAE of 0.012046, and an R2 of 0.970953. For SOC estimation, the model used here obtains the lowest MSE, RMSE, and MAE and the highest R2 among all evaluated models. The framework used here therefore achieves the best performance across all four reported SOC metrics among the retained baseline architectures.

Model	SOC MSE	SOC RMSE	SOC MAE	SOC R^2	SOH MSE	SOH RMSE	SOH MAE	SOH R^2
CNN	0.002302	0.047980	0.033441	0.965858	0.003266	0.057147	0.045530	0.713847
GRU	0.001604	0.040048	0.029246	0.976214	0.002239	0.047323	0.036006	0.803768
LSTM	0.001457	0.038174	0.026870	0.978388	0.002296	0.047913	0.037150	0.798847
BiLSTM	0.001935	0.043992	0.030054	0.971299	0.003152	0.056142	0.045452	0.723819
LSTM-Attention	0.001763	0.041994	0.030391	0.973846	0.001957	0.044238	0.034013	0.828524
CNN-BiLSTM	0.000658	0.025659	0.015785	0.990236	0.000636	0.025214	0.017302	0.944293
CNN-LSTM	0.000545	0.023344	0.013974	0.991918	0.000577	0.024017	0.016850	0.949456
CNN-LSTM-Attention	0.000806	0.028391	0.017906	0.988046	0.000694	0.026341	0.019460	0.939201
TCN-BiLSTM	0.000675	0.025989	0.014786	0.989983	0.000629	0.025080	0.017558	0.944883

Natural Resources for Human Health
eISSN: 2583-1194

Model	SOC MSE	SOC RMSE	SOC MAE	SOC R^2	SOH MSE	SOH RMSE	SOH MAE	SOH R^2
Transformer	0.000314	0.017719	0.010861	0.995344	0.000423	0.020556	0.015083	0.962976
Proposed Transformer–BiLSTM	0.000275	0.016589	0.009012	0.995919	0.000331	0.018207	0.012046	0.970953

Table 1. Quantitative comparison of the proposed Transformer–BiLSTM model with ten baseline models on the NASA test set. Lower MSE, RMSE, and MAE values and higher R^2 values indicate better performance.

5.2 SOC and SOH error comparison

The RMSE values provide a direct assessment of the magnitude of prediction errors across the evaluated architectures. As reported in Table 1, the framework used here reaches an SOC RMSE of 0.016589, which is lower than the Transformer baseline value of 0.017719 and substantially lower than those of the recurrent and convolutional baseline models. For SOH estimation, the model used here reaches an RMSE of 0.018207, followed by the Transformer baseline at 0.020556. The lower SOH error points to the fact that the learned shared representation effectively captures degradation-related patterns in the battery measurements.

5.3 Prediction agreement with reference states

The agreement between the predicted and reference battery states is further examined in Fig. 2. Fig. 2(a) shows the relationship between the predicted and reference SOC values, while Fig. 2(b) presents the corresponding SOH relationship. The SOC predictions are closely distributed around the ideal agreement line, with an R2 value of 0.9959 and an RMSE of 0.01659. The SOH predictions similarly follow the reference values, achieving an R2 of 0.9710 and an RMSE of 0.01821. These results indicate a strong correspondence between the predicted and reference states over the held-out test samples.

5.4 Battery-wise SOH estimation

To examine the consistency of the model used here across individual battery cells, SOH predictions were analyzed separately for B0005, B0006, B0007, and B0018. The corresponding results are shown in Fig. 3. The predicted SOH trajectories generally follow the reference trajectories across the four evaluated cells. The correspondence between the curves is maintained across different portions of the battery trajectories, including regions exhibiting relatively stable degradation and regions with changing SOH behaviour. This battery-wise analysis provides a qualitative assessment of the consistency of the model used here across multiple battery cells and complements the aggregate test metrics.

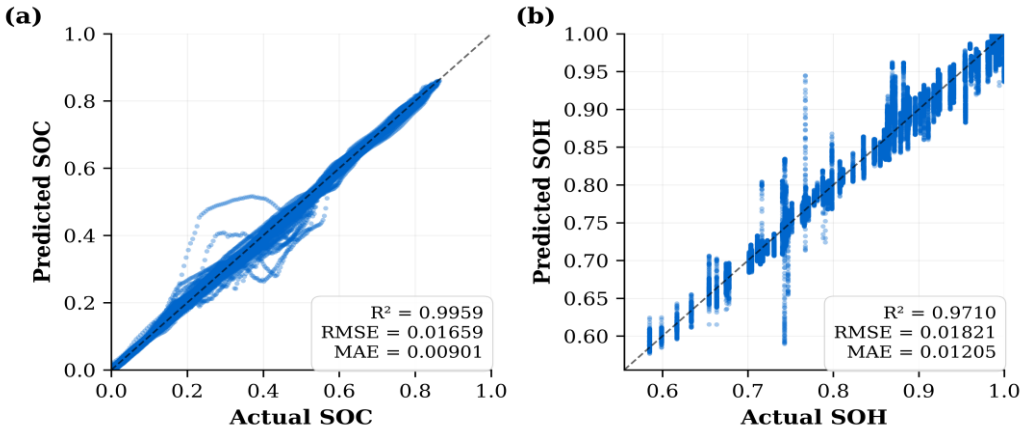


Fig. 2 Prediction agreement of the proposed Transformer–BiLSTM framework on the NASA test set: (a) SOC and (b) SOH. The dashed line represents ideal agreement between reference and predicted values.

5.5 Statistical comparison

To determine whether the observed differences in prediction accuracy are statistically meaningful under the adopted evaluation protocol, paired comparisons were performed between the model used here and each of the ten baseline models. Because ten comparisons were performed for each task, Holm correction was applied to the resulting p-values. The MAE difference was calculated from the paired absolute prediction errors as the absolute error of the model used here minus that of the corresponding baseline. Bootstrap resampling with 10,000 iterations was used to obtain the corresponding 95% confidence intervals. As shown in Table 2, the model used here shows statistically notable differences in SOC prediction error relative to CNN, GRU, LSTM, BiLSTM, LSTM-Attention, batteries B0005, B0006, B0007, and B0018. CNN-BiLSTM, CNN-LSTM, CNN-LSTM-Attention, TCN-BiLSTM, and Transformer after Holm correction. The corrected p-values are below 0.001 for all ten comparisons under the adopted paired-window protocol. For SOH estimation, the model used here shows statistically notable differences relative to all ten baseline models after Holm correction. Because the temporal sequences are built from overlapping sliding windows, neighbouring test windows are not strictly independent of one another. The reported significance levels should accordingly be read as evidence of differences within the adopted paired-window evaluation protocol, rather than as inference at the level of fully independent observations; an additional cycle-level statistical analysis is left for future work.

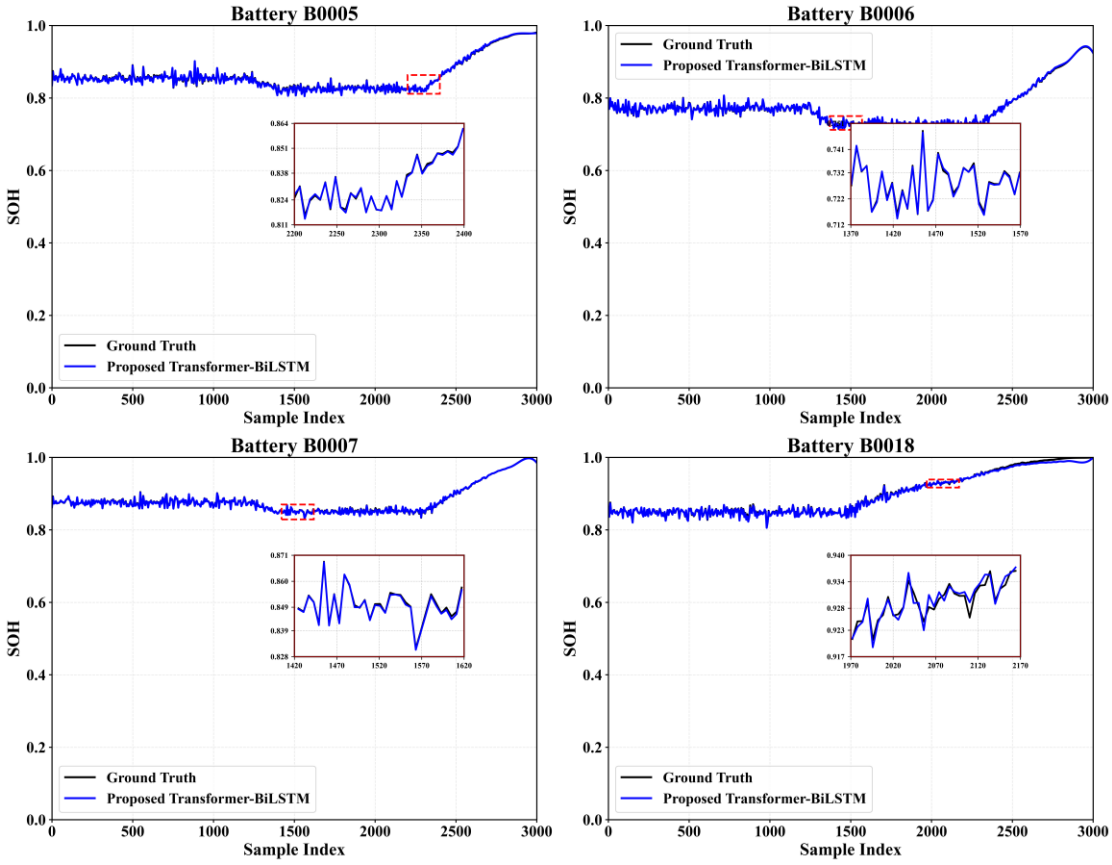


Fig. 3 Battery-wise SOH estimation of the proposed Transformer–BiLSTM framework for NASA

5.6 Prediction-error distribution

The distribution of prediction residuals provides additional information about the stability of the model used here beyond the aggregate error metrics. Figure 4 presents the residual distributions for SOC and SOH estimation. The SOC residuals are concentrated around zero, with a mean residual of approximately −0.00292 and an RMSE of 0.01659. Similarly, the SOH residuals remain concentrated near zero, with a mean residual of approximately −0.00201 and an RMSE of 0.01821. The concentration of errors around zero points to the fact that the framework used here maintains generally small prediction deviations across the held-out test samples. The residual analysis therefore complements the quantitative comparison by demonstrating that the low aggregate error is associated with a generally concentrated prediction-error distribution.

5.7 Computational efficiency

In addition to prediction accuracy, inference processing time was evaluated to assess the computational efficiency of the framework used here. The proposed Transformer– BiLSTM model requires approximately 0.04564 ms per test sample under the adopted batched inference configuration. For comparison, the Transformer baseline requires approximately 0.05744 ms per test sample. So, the framework used here exhibits lower measured average inference processing time per test sample than the Transformer baseline considered in this comparison. The combination of low prediction error and low measured inference processing time indicates the potential suitability of the framework for computationally constrained battery-state monitoring applications.

Comparison	Target	p-value	Holm-adjusted p	MAE difference	95% CI	Significant?
Proposed vs CNN	SOC	<0.001	<0.001	-0.024429	[-0.024841, -0.024029]	Yes
Proposed vs GRU	SOC	<0.001	<0.001	-0.020235	[-0.020568, -0.019902]	Yes
Proposed vs LSTM	SOC	<0.001	<0.001	-0.017858	[-0.018181, -0.017538]	Yes
Proposed vs BiLSTM	SOC	<0.001	<0.001	-0.021042	[-0.021431, -0.020666]	Yes
Proposed vs LSTM-Attention	SOC	<0.001	<0.001	-0.021379	[-0.021733, -0.021016]	Yes
Proposed vs CNN-BiLSTM	SOC	<0.001	<0.001	-0.006773	[-0.007013, -0.006535]	Yes
Proposed vs CNN-LSTM	SOC	<0.001	<0.001	-0.004963	[-0.005155, -0.004772]	Yes
Proposed vs CNN-LSTM-Attention	SOC	<0.001	<0.001	-0.008894	[-0.009132, -0.008665]	Yes
Proposed vs TCN-BiLSTM	SOC	<0.001	<0.001	-0.005775	[-0.006038, -0.005517]	Yes
Proposed vs Transformer	SOC	<0.001	<0.001	-0.001849	[-0.002019, -0.001680]	Yes
Proposed vs CNN	SOH	<0.001	<0.001	-0.033484	[-0.033942, -0.033033]	Yes
Proposed vs GRU	SOH	<0.001	<0.001	-0.023960	[-0.024368, -0.023560]	Yes
Proposed vs LSTM	SOH	<0.001	<0.001	-0.025104	[-0.025518, -0.024698]	Yes
Proposed vs BiLSTM	SOH	<0.001	<0.001	-0.033406	[-0.033852, -0.032980]	Yes

Comparison	Target	p-value	Holm-adjusted p	MAE difference	95% CI	Significant?
Proposed vs LSTM-Attention	SOH	<0.001	<0.001	-0.021967	[-0.022342, -0.021596]	Yes
Proposed vs CNN-BiLSTM	SOH	<0.001	<0.001	-0.005256	[-0.005490, -0.005030]	Yes
Proposed vs CNN-LSTM	SOH	<0.001	<0.001	-0.004805	[-0.005001, -0.004612]	Yes
Proposed vs CNN-LSTM-Attention	SOH	<0.001	<0.001	-0.007414	[-0.007638, -0.007199]	Yes
Proposed vs TCN-BiLSTM	SOH	<0.001	<0.001	-0.005512	[-0.005754, -0.005273]	Yes
Proposed vs Transformer	SOH	<0.001	<0.001	-0.003037	[-0.003207, -0.002863]	Yes

Table 2. Statistical comparison of the proposed model against ten baseline models using the Wilcoxon signed-rank test, Holm correction and bootstrap-based 95% confidence intervals. MAE difference is proposed minus baseline absolute error.

Model	Inference time (ms/sample)
Proposed Transformer–BiLSTM	0.04564
Transformer	0.05744

Table 3. Measured inference processing time of the proposed model and the Transformer-based baseline under the adopted batched inference configuration.

5.8 Overall discussion

The experimental results show that the proposed Transformer–BiLSTM multi-task framework provides accurate simultaneous estimation of SOC and SOH from voltage, current, and temperature measurements. The model reaches an SOC RMSE of 0.016589 and an SOH RMSE of 0.018207, with corresponding R2 values of 0.995919 and 0.970953, respectively. The comparison with ten baseline architectures demonstrates the effectiveness of combining Transformer-based dependency modelling with bidirectional temporal feature extraction within the adopted multi-task learning framework. The model used here exhibits statistically notable differences against the evaluated baselines for both estimation tasks after Holm correction. The error-distribution analysis also shows that the prediction residuals remain concentrated around zero.

Natural Resources for Human Health
eISSN: 2583-1194

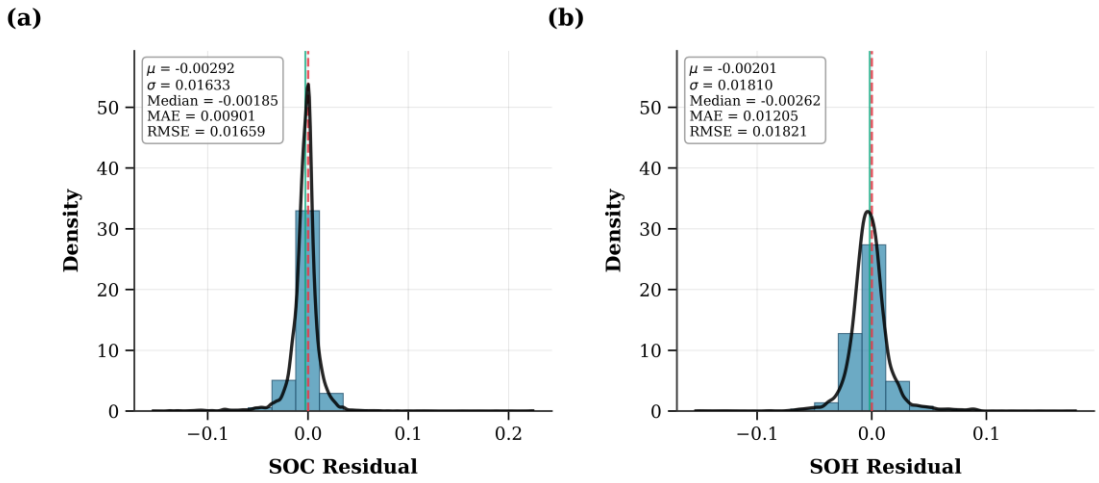


Fig. 4 Residual distributions of the proposed Transformer–BiLSTM model on the NASA test set: (a) SOC residuals and (b) SOH residuals.

6 TARGET-BATTERY FEW-SHOT ADAPTATION STUDY

6.1 Adaptation protocol

To examine how the proposed Transformer–BiLSTM multi-task model behaves when only a very limited number of additional cycles from a specific target battery are available, a target-battery-specific few-shot adaptation study was conducted using the pretrained model described in Section 4. Note that, because the pretrained model was obtained using the cycle-level random split described in Section 4.1, a subset of cycles from each of the four batteries was already present in the original training set. The present study therefore evaluates adaptation efficiency under a constrained-data scenario for a specific target battery, rather than estimation performance on a battery entirely unseen during pretraining; extending this analysis to a strict leave-one-battery-out protocol is identified as a direction for future work. For each target battery (B0005, B0006, B0007, B0018), a small percentage (1%, 5%, or 10%) of that battery's training cycles was randomly sampled, with the remaining battery-specific test cycles held out for evaluation. Three adaptation strategies were compared: 1. Scratch: a randomly initialized model trained using only the sampled adaptation cycles, with no access to the pretrained weights. 2. Conventional fine-tuning: the pretrained model with all parameters updated using the sampled adaptation cycles. 3. Proposed few-shot adaptation: the pretrained model with the Transformer encoder and BiLSTM layers frozen, and only the SOC and SOH prediction heads updated using the sampled adaptation cycles. All three strategies were trained for 10 epochs with the Adam optimizer, using the same joint SOC/SOH loss (Eq. 16) and SOH weighting factor as in the main training configuration.

6.2 Parameter-efficient adaptation mechanism

The proposed adaptation strategy updates only the two task-specific multilayer perceptron heads while keeping the shared Transformer and BiLSTM feature extractor fixed at its pretrained weights. This substantially reduces the number of trainable parameters relative to conventional fine-tuning, as summarized in Table 4.

Method	Trainable parameters	% of total	Frozen modules
Scratch	297,282	100.000%	None (random initialization)
Conventional fine-tuning	297,282	100.000%	None
Proposed few-shot adaptation	5,250	1.766%	Transformer encoder, BiLSTM

Table 4. Trainable-parameter comparison across adaptation strategies.

6.3 Adaptation results

Table 5 reports the SOC and SOH estimation performance of each adaptation strategy, averaged across the four target batteries, at adaptation levels of 1%, 5%, and 10%. The performance of the pretrained model without any additional adaptation (0%) is also reported as a reference.

Adaptation	Method	SOC MAE	SOC RMSE	SOC R^2	SOH MAE	SOH RMSE	SOH R^2
0%	Pretrained (no adaptation)	0.009140	0.015907	0.995766	0.012085	0.017741	0.960273
1%	Scratch	0.237504	0.279647	−0.181016	0.158696	0.172415	−4.671324
1%	Conventional fine-tuning	0.013485	0.020066	0.993501	0.021010	0.028558	0.897503
1%	Proposed few-shot adaptation	0.010532	0.016599	0.995376	0.013557	0.018834	0.955107
5%	Scratch	0.204710	0.242527	−0.042965	0.084653	0.101221	−0.256169
5%	Conventional fine-tuning	0.012164	0.020385	0.992944	0.022119	0.031639	0.844371
5%	Proposed few-shot adaptation	0.009080	0.014979	0.996342	0.011989	0.016463	0.964724
10%	Scratch	0.036166	0.048086	0.963672	0.073544	0.089136	−0.012441
10%	Conventional fine-tuning	0.007552	0.011562	0.997913	0.011961	0.016960	0.962894
10%	Proposed few-shot adaptation	0.007516	0.012188	0.997583	0.011486	0.015949	0.966636

Table 5. Average SOC and SOH estimation performance across the four target batteries (B0005, B0006, B0007, B0018) under different adaptation levels and strategies.

6.4 Discussion

At the lowest adaptation level (1%, corresponding to only a few discharge cycles per battery), training a model from scratch performs poorly on both tasks, with negative R2 values indicating predictions worse than a constant-mean baseline. Both adaptation strategies that start from the pretrained model recover strong performance even at 1% adaptation. However, the proposed strategy, despite updating only 1.766% of the model's parameters, matches or outperforms conventional fine-tuning at every adaptation level for both SOC and SOH, and achieves the lowest SOH error of all three strategies at the 5% and 10% adaptation levels (e.g., at 5% adaptation, SOH MAE of 0.011989 for the proposed strategy versus 0.022119 for conventional fine-tuning). This points to the possibility that the frozen Transformer–BiLSTM backbone already encodes transferable, degradation-related temporal dynamics, and that adapting only the lightweight prediction heads is sufficient to specialize the model to a target battery while avoiding the overfitting that can occur when all 297,282 parameters are updated using very few cycles.

7 CONCLUSION

This paper set out a Transformer–BiLSTM multi-task learning framework for simultaneous SOC and SOH estimation of lithium-ion batteries from voltage, current, and temperature measurements. The architecture combines a 32-dimensional linear embedding, a two-layer Transformer encoder with four attention heads, and a bidirectional LSTM network with a hidden size of 32 in each direction, and the resulting 64-dimensional shared representation goes through separate MLP heads for SOC and SOH prediction. Evaluated on four NASA battery cells (B0005, B0006, B0007, B0018, 636 discharge cycles in total), split 70/15/15 at the battery-cycle level, the framework used here reached an SOC RMSE of 0.016589 (MAE 0.009012, R2 0.995919) and an SOH RMSE of 0.018207 (MAE 0.012046, R2 0.970953) on the held-out test set, the lowest MSE, RMSE, and MAE and the highest R2 among ten baseline deep-learning models, for both SOC and SOH, with the differences holding up as statistically notable under the adopted paired-window protocol and residual errors staying concentrated around zero. Section 6 took this further, showing that updating only the SOC and SOH heads (1.766% of parameters) on a specific target battery with as little as 1%, 5%, or 10% of extra cycles matched or beat full-model fine-tuning while avoiding the overfitting seen when training from scratch on such limited data, which points to the shared Transformer–BiLSTM backbone already learning degradation-related representations that transfer efficiently. Two limitations are worth flagging. First, because the reference SOC depends on the measured discharge capacity of each complete cycle, this evaluation is best read as an offline, reference-based estimation study rather than a demonstration of fully online SOC estimation; reference-generation procedures that avoid needing the complete discharge capacity mid-cycle are a natural next step. Second, because cycles from all four battery cells appear in both the training and test subsets, including in the pretrained backbone used for the few-shot study, these results should not be read as a strict cross-battery generalization test; a leave-one-battery-out protocol, where an entire cell is excluded from training, is left for future work, together with cycle-level statistical testing to account for the fact that the overlapping sliding windows used here are not fully independent observations. Overall, the results support the shared Transformer–BiLSTM representation as an effective, computationally efficient, data-driven approach to joint SOC and SOH estimation, one that can also be specialised efficiently to a target battery with minimal extra data.

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

FUNDING

The authors received no specific funding for this work.

DATA AVAILABILITY

The datasets used in this study are publicly available through the NASA Ames Prognostics Centre of Excellence repository.

Code availability

Code is available from the corresponding author upon reasonable request.

Authors' contributions

T.S., K.S.G., and P.C.T. developed the software implementation, conducted the data preprocessing, model training, and experimental analysis, and wrote the original draft of the manuscript. S.D. contributed to reviewing the draft manuscript. A.P.D., M.R., R.D.G., and M.J. reviewed and critically revised the manuscript for important intellectual content. S.S. conceived and supervised the study, provided overall guidance on the methodology, and reviewed and finalized the manuscript. All authors read and approved the final manuscript.

REFERENCES

[1] Hannan, M.A., Lipu, M.S.H., Hussain, A., Mohamed, A.: A review of lithium-ion battery SOC estimation and management systems in electric vehicle applications. Renewable and Sustainable Energy Reviews 78, 834–854 (2017).
[2] Ng, M.-F., Zhao, J., Yan, Q., Conduit, G.J., Seh, Z.W.: Predicting the state of charge and health of batteries using data-driven machine learning. Nature Machine Intelligence 2, 161–170 (2020).
[3] Salkind, A.J., Fennie, C., Singh, P., Atwater, T., Reisner, D.E.: Determination of SOC and SOH of batteries by fuzzy logic methodology. Journal of Power Sources 80, 293–300 (1999).
[4] Hu, C., Jain, G., Zhang, P., Schmidt, C., Gomadam, P., Gorka, T.: Online estimation of lithium-ion battery capacity using sparse Bayesian learning. Journal of Power Sources 289, 105–113 (2015).
[5] Xiong, R., Cao, J., Yu, Q., He, H., Sun, F.: Critical review on battery SOC estimation methods for electric vehicles. IEEE Access 6, 1832–1843 (2017).
[6] Sesidhar, D., et al.: A review on data-driven SOC estimation with Li-Ion batteries: Implementation methods & future aspirations. Journal of Energy Storage 72, 108420 (2023).
[7] Tian, J., et al.: Deep learning framework for lithium-ion battery state of charge estimation: Recent advances and future perspectives. Energy Storage Materials (2023). Article number 102883.
[8] Dong, L., et al.: Inflection point Ah-total integration method for real-time integration to correct lithium battery SOC. Energy Storage Science and Technology 8, 850 (2019).
[9] Zheng, F., et al.: Influence of different open circuit voltage tests on online SOC estimation for lithium-ion batteries. Applied Energy 183, 513–525 (2016).
[10] Liu, D., et al.: A novel fuzzy-extended Kalman filter-ampere-hour algorithm for SOC estimation of lithium-ion batteries. International Journal of Circuit Theory and Applications 50, 3811–3826 (2022).
[11] Chen, C., et al.: Battery-in-the-loop validation of multiscale dual H-infinity filters for SOC and capacity estimation. IEEE Transactions on Power Electronics 33, 332–342 (2017).
[12] Chai, X., et al.: A novel battery SOC estimation method based on random search optimized LSTM neural network. Energy 306 (2024). Article number 132583.
[13] Li, Y., et al.: An improved model combining machine learning and Kalman filtering architecture for SOC estimation of lithium-ion batteries. Green Energy and Intelligent Transportation (2024). Article number 100163.
[14] Yang, F., et al.: SOC estimation based on gated recurrent neural network. Energy 175, 66–75 (2019).
[15] Yang, F., et al.: SOC estimation of lithium-ion batteries via long short-term memory network. IEEE Access 7, 53792–53799 (2019).
[16] Shen, H., et al.: SOC estimation using Transformer with immersion and invariance adaptive observer. Journal of Energy Storage 45 (2022). Article number 103768.
[17] Guirguis, J., Ahmed, R.: Transformer-based deep learning models for state of charge and state of health estimation of Li-ion batteries: A survey study. Energies 17, 3502 (2024).
[18] Zhao, S.-Y., et al.: A novel transformer-embedded lithium-ion battery model for joint estimation of SOC and SOH. Rare Metals (2024).
[19] Wu, L., Chen, C., Li, Z., Chen, Z., Li, H.: The joint estimation of SOC-SOH for lithium-ion batteries based on BiLSTM-SA. Electronics 14, 97 (2025).
[20] Fan, T.-E., Wang, K., Qu, B.: Joint estimation of Li-ion battery states based on parallel TCN-Transformer multi-task learning model. Journal of Energy Storage 134, 118077 (2025).
[21] Li, Z., et al.: State of health estimation of lithium-ion battery during fast charging based on BiLSTM-Transformer. Energy 311 (2024). Article number 133418.