

Article

A Data-Driven Method Based on Feature Engineering and Physics-Constrained LSTM-EKF for Lithium-Ion Battery SOC Estimation

Yujuan Sun ¹, Shaoyuan You ¹, Fangfang Hu ² and Jiuyu Du ^{3,*}

¹ School of Mechanical and Energy Engineering, Beijing University of Technology, Beijing 100124, China; yjsun@bjut.edu.cn (Y.S.); y11sy@emails.bjut.edu.cn (S.Y.)

² Beijing Products Quality Supervision and Inspection Research Institute (National Automotive Quality Inspection and Testing Center), Beijing 101300, China; ff8345@126.com

³ School of Vehicle and Mobility, Tsinghua University, Beijing 100084, China

* Correspondence: dujiuyu@tsinghua.edu.cn

Abstract

Accurate estimation of the State of Charge (SOC) for lithium-ion batteries is a core function of the Battery Management System (BMS). However, LiFePO₄ batteries present specific challenges for SOC estimation due to the characteristic plateau in their open-circuit voltage (OCV) versus SOC relationship. Moreover, data-driven estimation approaches often face significant difficulties stemming from measurement noise and interference, the highly nonlinear internal dynamics of the battery, and the time-varying nature of key battery parameters. To address these issues, this paper proposes a Long Short-Term Memory (LSTM) model integrated with feature engineering, physical constraints, and the Extended Kalman Filter (EKF). First, the model's temporal perception of the historical charge–discharge states of the battery is enhanced through the fusion of temporal voltage information. Second, a post-processing strategy based on physical laws is designed, utilizing the Particle Swarm Optimization (PSO) algorithm to search for optimal correction factors. Finally, the SOC obtained from the previous steps serves as the observation input to EKF filtering, enabling a probabilistically weighted fusion of the data-driven model output and the EKF to improve the model's dynamic tracking performance. When applied to SOC estimation of LiFePO₄ batteries under various operating conditions and temperatures ranging from 0 °C to 50 °C, the proposed model achieves average Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as low as 0.46% and 0.56%, respectively. These results demonstrate the model's excellent robustness, adaptability, and dynamic tracking capability. Additionally, the proposed approach only requires derived features from existing input data without the need for additional sensors, and the model exhibits low memory usage, showing considerable potential for practical BMS implementation. Furthermore, this study offers an effective technical pathway for state estimation under a “physical information–data-driven–filter fusion” framework, enabling accurate SOC estimation of lithium-ion batteries across multiple operating scenarios.

Keywords: state of charge estimation; lithium-ion battery; long short-term memory; extended Kalman filter; particle swarm optimization algorithm

Academic Editor: Federico Baronti

Received: 22 December 2025

Revised: 3 February 2026

Accepted: 9 February 2026

Published: 14 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Amid the growing severity of the global energy crisis and environmental pollution, the development of clean and sustainable energy technologies has become a strategic priority worldwide [1,2]. Against this backdrop, lithium-ion batteries have gained extensive application in electric vehicles (EVs), renewable energy storage systems, aerospace, and portable electronic devices due to their notable advantages, such as high energy density, long cycle life, low self-discharge rate, and absence of memory effect. As a result, they have become a key driver in the ongoing transition of the global energy structure [3–5].

The Battery Management System (BMS), often regarded as the “brain” of a battery pack, has one of its core functions as the accurate real-time estimation of key battery states [6]. Among these, the State of Charge (SOC) is one of the most critical parameters for BMS operation. It is conventionally defined as the percentage of a battery’s remaining available capacity relative to the nominal capacity [7]. Accurate SOC estimation is essential to prevent incidents such as vehicle breakdowns and unexpected device shutdowns [8], while also contributing to improved system efficiency and extended battery pack lifespan [9–11]. However, SOC is an internal state that cannot be measured directly by sensors [12]. Its estimation is complicated by the coupled influence of multiple factors, including dynamic variations in battery operating temperature, charge/discharge current rates, cycle aging, and operating conditions, as well as the inherent time-varying and nonlinear characteristics of internal electrochemical parameters, such as polarization resistance and diffusion coefficient [13,14]. Currently, SOC estimation methods reported in the literature are broadly categorized into four groups: the Coulomb counting method, the open-circuit voltage (OCV) method, model-based estimation methods, and data-driven estimation methods [15–17].

Among these methods, the Coulomb counting method is the most direct in principle [18]. It calculates the SOC directly based on its definition, using real-time operating current measurements, thus forming a simple open-loop system. However, it suffers from two critical limitations: first, a strong dependence on an accurate initial SOC; second, the accumulation of measurement errors in Coulomb counting over long-term operation [19]. The OCV method relies on establishing a monotonic mapping between the battery’s OCV and SOC, i.e., the OCV–SOC curve [20,21]. This approach requires the battery to be in a prolonged resting state to reach equilibrium [22,23], and its accuracy is limited for LiFePO₄ batteries due to their characteristically flat OCV–SOC profile [24,25]. Peng et al. [22] proposed a recursive least square algorithm with a forgetting factor (FF-RLS) for battery model parameter identification and adopted an exponentially weighted moving average (EWMA) method to calculate the temperature compensation coefficient, thereby reducing OCV identification errors. They further partitioned the SOC estimation process according to the slope characteristics of the OCV–SOC curve. In gentle-slope regions, the system primarily relied on the extended Kalman filter (EKF) for high-precision state updates. In steep-slope and plateau regions, it switched to a hybrid estimation scheme employing a proportional–integral–differential (PID) observer integrated with an adaptive extended Kalman filter (AEKF). The PID controller was embedded into the AEKF correction step to perform multi-dimensional adjustment of the innovation (the difference between observed and predicted values). The proportional (P) term provided a rapid response to the current error, the integral (I) term accumulated errors to eliminate steady-state offset, and the differential (D) term tracked the error rate of change to suppress overshoot. This integrated approach ultimately achieved high-precision SOC estimation with a maximum absolute error below 3%. In the category of model-based estimation methods, the pseudo-two-dimensional (P2D) model involves solving coupled nonlinear partial differential equations, which is extremely time-consuming and thus impractical for integration into

real-time BMS applications [26,27]. In contrast, the equivalent circuit models (ECMs) require specific parameter identification for each individual battery [22,28]. Additionally, it still relies on the OCV–SOC curve combined with state observers for closed-loop SOC estimation [29] and thus cannot effectively address the flat plateau problem of the OCV–SOC curve for LiFePO₄ batteries [24,30].

In recent years, extensive studies have explored the application of machine learning and deep learning methods in battery SOC estimation. These methods do not rely on specific battery models; instead, they treat the battery as a “black box” and directly establish the complex nonlinear mapping relationship between the battery’s external measurement information and SOC by learning a large volume of historical operating data [31–33]. Chen et al. [34] proposed a novel SOC estimation method for LiFePO₄ batteries based on the combined modeling of physical and machine learning models. A first-order RC ECM was employed to fit and shield interpretable dynamics (such as ohmic voltage and polarization voltage), while the remaining unexplainable dynamics were defined as open-circuit voltage noise (OCVN) and fitted with a Long Short-Term Memory (LSTM) model to capture their nonlinear mapping with SOC. This method enhanced the interpretability of SOC models based on direct mapping. Bao et al. [35] presented a collaborative framework integrating Transformer and LSTM models, where the long-term SOC estimation results derived from the Transformer were used as input features for the LSTM, leveraging the Transformer’s advantage in capturing long-term dependencies and the LSTM’s capability in modeling short-term patterns in sequence data. Li et al. [36] proposed a hybrid neural network model that combines the data processing capability of temporal convolutional networks (TCNs) and the long-term dependency learning ability of gated recurrent units (GRUs). Wan et al. [37] proposed an improved whale optimization algorithm (WOA) to optimize the number of hidden layer nodes, learning rate, and number of iterations of the LSTM model, overcoming the shortcomings of manually setting hyperparameters.

To improve the modeling accuracy and generalization ability of the data-driven networks and further construct high-performing models, researchers have focused on expanding the dimensionality of input features to enrich the information content available for learning. Sulaiman et al. [38] proposed a SOC estimation model based on a Random Forest (RF) algorithm. The model was trained using data collected from 70 real-world driving trips of the BMW i3 EV equipped with a 60 Ah Nickel Manganese Cobalt oxide (NMC) battery pack, and the cell was manufactured by Samsung SDI of South Korea. It incorporated a comprehensive set of input features, including battery voltage, current, battery temperature, and ambient temperature, achieving a Root Mean Square Error (RMSE) of 5.90% and a Mean Absolute Error (MAE) of 4.43%. This work offers valuable insights for applying machine learning algorithms to real-world vehicle SOC estimation. For the future optimization of data-driven models, it would be beneficial to further verify the consistency of data distribution between the training and test sets in terms of driving patterns and operating scenarios, such as highway cruising and urban congestion. Chen et al. [39] adopted voltage, current, operating temperature, and state of health (SOH) as inputs, using the LSTM to first generate a rough SOC estimate, followed by smoothing the output results with an adaptive H[∞] filter, and ultimately controlling the estimation error within 2.1%. Wang et al. [40] generated new features (voltage × current, voltage × temperature, and current × temperature) through feature crossing of voltage, current, and temperature and then reduced the feature dimension through the RF before inputting the data into a convolutional neural network (CNN). This study emphasized the significant role of feature engineering technology in SOC estimation. Meanwhile, Xu et al. [41] and Wu et al. [42] incorporated battery surface expansion force and external battery preload as additional input features, respectively, but these methods require additional sensing devices.

Compared with single algorithms and open-loop combined algorithms, integrated algorithms generally exhibit better performance in SOC estimation. Li et al. [43] proposed a closed-loop estimation method that integrates a Support Vector Regression (SVR) model with an EKF framework. The method was comprehensively validated under multiple typical operating conditions—including the Dynamic Stress Test (DST), US06 High-Speed Driving Schedule (US06), Federal Urban Driving Schedule (FUDS), and Beijing Dynamic Stress Test (BJDST)—at a constant temperature of 25 °C. Additionally, a random profile (Random) composed of random combinations of the four aforementioned conditions was tested over a broader temperature range of 10 °C to 40 °C. Results indicated that the proposed approach achieved an MAE of $\leq 0.60\%$ and an RMSE of $\leq 0.73\%$. It is worth noting that this study focused on lithium NMC batteries, with model inputs limited to battery operating current and voltage, thereby establishing a basis for further investigation into the potential influence of additional parameters. Tian et al. [44] employed an LSTM network to learn the nonlinear relationship between SOC and measured parameters and then applied an adaptive cubature Kalman filter (ACKF) to smooth the output of the LSTM network, thereby achieving accurate and stable SOC estimation. Jiang et al. [45] proposed an integrated algorithm combining the LSTM and an adaptive unscented Kalman filter (AUKF). In this algorithm, the LSTM network was employed to capture long-term temporal dependencies in multivariate battery data, while the AUKF dynamically adjusted its parameters to filter measurement noise, thereby enhancing the robustness of SOC estimation.

Existing studies have shown that LiFePO_4 batteries exhibit a flatter OCV-SOC curve compared to NMC batteries, resulting in more pronounced nonlinear behavior. Furthermore, current methods often do not adequately incorporate the temporal information of historical voltage data, making it difficult to extract stable features of battery charge–discharge states from fluctuating voltage measurements. This, in turn, reduces model robustness against noise and transient disturbances. To address this issue, this study incorporates an averaged voltage feature into the model input, enhancing its ability to capture the temporal evolution of historical battery charge–discharge states. This strategy is referred to as feature introduction (FI), and the corresponding model is denoted as LSTM_FI. By integrating time-series voltage information, the model strengthens its temporal perception of historical charge–discharge states, thereby suppressing interference from noise and data anomalies while effectively learning the underlying nonlinear mapping.

On the other hand, deep learning models are highly sensitive to measurement noise and disturbances during data acquisition. Such models tend to overreact to anomalous fluctuations in input data, which not only causes significant oscillations in SOC predictions but also limits the overall dynamic tracking performance. Moreover, the raw model output is often directly adopted as the final prediction, without considering the dynamic volatility of the output or the physical constraints that SOC must satisfy; for example, the direction of SOC change must align with the direction of the current.

To mitigate these limitations, this paper proposes a physics-guided constraint mechanism applied at the output stage, based on the direction of the operating current. The Particle Swarm Optimization (PSO) algorithm is employed to search for an optimal correction coefficient. This strategy is termed limit output (LO), and the corresponding model is named LSTM_LO. By dynamically constraining the model output, this approach jointly improves prediction stability and physical consistency.

Finally, to further enhance dynamic tracking capability, the SOC from the above model is used as the observation input for EKF processing, enabling a probabilistically weighted fusion of the data-driven output and the EKF estimate. This integrated model is referred to as LSTM_FILO_EKF. The overall framework of the proposed methodology is illustrated in Figure 1.

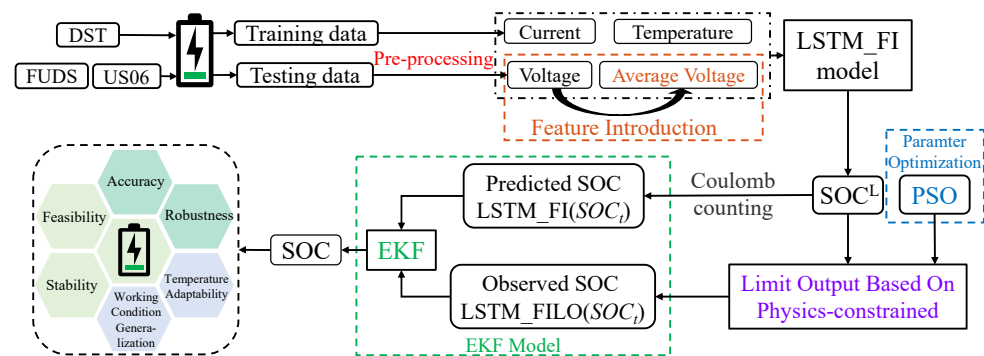


Figure 1. Overall framework of proposed battery SOC estimation method.

Voltage, current, and temperature data from cyclic tests on LiFePO₄ batteries under the DST profile across different ambient temperatures were used as the training dataset for the proposed model. Meanwhile, data collected under the FUDS and US06 operating conditions at various ambient temperatures served as the validation set. Following data preprocessing, the average battery voltage-derived from the raw voltage measurements-alongside the original voltage, current, and temperature, were fed into the LSTM network as input features.

The output of the LSTM_FI model, denoted as SOC^L , was subsequently processed by the LO module, whose parameters were optimized via the PSO algorithm. The resulting refined SOC estimate was then used as the observation input to an EKF. In contrast, the SOC^L calculated through the Coulomb counting method is employed as the predicted value of the EKF. Adaptive adjustment of the Kalman gain enabled probabilistically weighted fusion between the data-driven model output and the EKF estimate. This integrated approach establishes a reliable foundation for practical implementation in BMS, thereby supporting accurate battery state monitoring and control.

The remainder of this paper is organized as follows. Section 2 provides a systematic exposition of the fundamental theory and methodology, including the principles and implementation of the LSTM network, the optimization mechanism of the PSO algorithm, the construction of the average voltage feature, the implementation of the output constraint strategy, the principles and parameter settings of the EKF, as well as the overall model architecture and hyperparameter configuration. Section 3 describes the source of the experimental dataset, the composition of input features, and the data preprocessing steps, while also defining the evaluation metrics used to assess prediction performance. Section 4 presents a module-by-module validation and a comprehensive performance evaluation. First, the prediction performance of the LSTM_FI model is compared with that of the baseline LSTM. Next, the optimization process of the correction factor via the PSO algorithm is analyzed, and the standalone effect of output constraints is verified. Third, the model behavior under the combined effect of the two strategies is investigated. Finally, the overall performance of the proposed integrated model is validated, and its advantages are systematically illustrated through a quantitative error analysis against reference values. Section 5 concludes the paper with a summary of the research findings.

2. Proposed LSTM_FILO_EKF Model

2.1. LSTM Neural Network

As a variant of Recurrent Neural Networks (RNNs), LSTM networks effectively address common issues in traditional neural networks such as local minimum, gradient vanishing, and gradient explosion. Meanwhile, the charging and discharging processes of batteries are continuous and exhibit strong temporal dependence: the SOC at any given moment is influenced not only by the instantaneous current, voltage, and temperature,

but also by historical operating states. The unique “gating mechanism” of LSTM allows it to effectively learn and retain long-term dependencies within time-series data, thereby offering strong modeling capabilities in this context. Furthermore, the internal electrochemical reactions of batteries are highly complex, resulting in a strongly nonlinear relationship between SOC and measurable variables such as voltage, current, and temperature. LSTM is capable of learning these complex mappings from training, which gives it a distinct advantage in SOC estimation. The structure of an LSTM unit is illustrated in Figure 2 and consists of three gates: a forget gate, an input gate, and an output gate. Specifically, the forget gate is responsible for deleting state information within the LSTM unit; the input gate selects and transmits valuable input information to the storage unit; and the output gate controls the amount of information output from the storage unit [46].

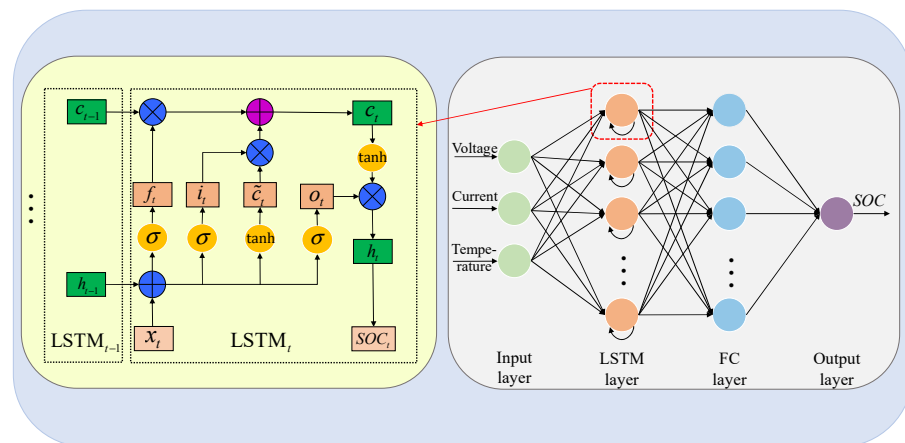


Figure 2. LSTM cell structure.

The calculation process of the LSTM unit at time t is as follows:

$$\left\{ \begin{array}{l} f_t = \sigma(U_f \cdot h_{t-1} + W_f \cdot x_t + b_f) \\ i_t = \sigma(U_i \cdot h_{t-1} + W_i \cdot x_t + b_i) \\ o_t = \sigma(U_o \cdot h_{t-1} + W_o \cdot x_t + b_o) \\ \tilde{c}_t = \tanh(U_c \cdot h_{t-1} + W_c \cdot x_t + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t = o_t \odot \tanh c_t \\ soc_t^L = g(h_t) = g(V \cdot h_t + b_{soc}) \end{array} \right. \quad (1)$$

where f_t , i_t , and o_t are the forget gate, input gate, and output gate, respectively. $\tilde{c}_t$ is the candidate state, and c_t is the internal state. The sigmoid function $\sigma \in [0, 1]^D$ and the hyperbolic tangent function $\tanh \in [-1, 1]^D$. x_t is the state vector composed of voltage, current, and temperature. soc_t^L is the SOC estimated by LSTM at time t . $g(\cdot)$ is the non-linear activation function of the output layer. U , V , and W are weight matrices, and b is the bias vector. $\odot$ represents vector multiplication.

2.2. Particle Swarm Optimization

PSO is a swarm intelligence optimization algorithm inspired by the intelligent behavior of social organisms, such as bird flocks and fish schools. Its core idea is to guide the entire particle swarm toward the optimal solution region through information sharing and cooperation among individual particles. The PSO algorithm is characterized by a simple structure, a small number of adjustable parameters, fast convergence speed, and no strict constraints on the properties of the objective function (such as differentiability and

continuity). Since its introduction, it has been widely applied in various fields, including function optimization, neural network training, fuzzy system control, and pattern recognition, and it has become an effective tool for solving complex optimization problems [47].

In the PSO algorithm, each potential solution is regarded as a “particle” within the search space. The entire particle swarm consists of a certain number of particles, and each particle is characterized by two fundamental attributes:

- Position: the coordinates of the solution in the search space, denoted as $X_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$, where d is the dimension of the search space;
- Velocity: the direction and step size of the particle’s next iteration, denoted as $V_i = \{v_{i1}, v_{i2}, \dots, v_{id}\}$. Each particle updates its velocity and position by tracking two extreme values;
- Individual extreme value: the optimal position found by the particle itself during iteration, denoted as $P_i = \{p_{i1}, p_{i2}, \dots, p_{id}\}$;
- Global extreme value: the optimal position found by the entire particle swarm so far, denoted as $P_g = \{p_{g1}, p_{g2}, \dots, p_{gd}\}$.

The execution process of the standard PSO algorithm is an iterative process, and its flow chart is shown in Figure 3.

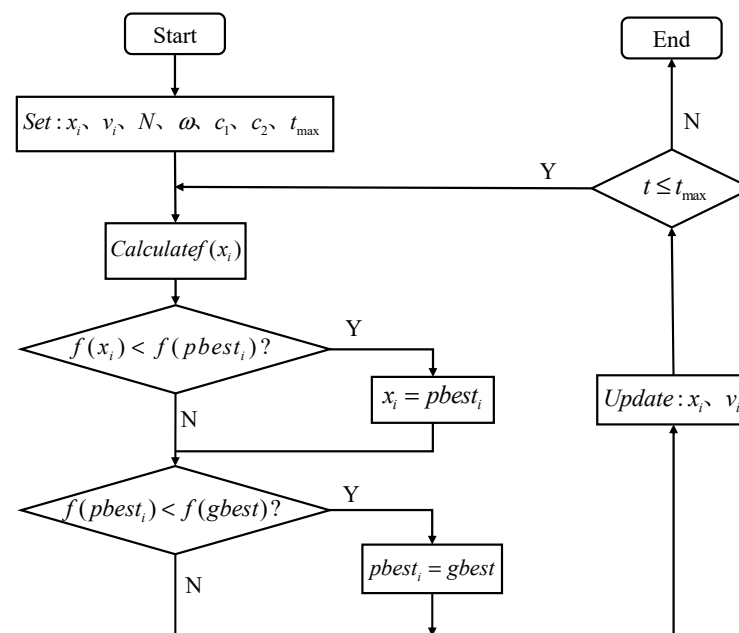


Figure 3. PSO process.

The parameters in Figure 3 are explained as follows:

- $f(x_i)$: fitness value of the i -th particle;
- $f(pbest_i)$: optimal fitness value of the i -th particle;
- $f(gbest)$: global optimal fitness value;
- t represents the current iteration number;
- t_{max} represents the maximum number of iterations;
- N represents the number of particles in the swarm;
- ω denotes the inertia weight, which balances the global search capability and local search capability of the algorithm;
- c_1 and c_2 are learning factors, usually positive constants. c_1 adjusts the step size of the particle moving toward its own historical optimal position (cognitive component), while c_2 adjusts the step size of the particle moving toward the global historical optimal position (social component);

- v_i and x_i represent the velocity and position of the i -th particle at the t -th iteration, respectively.

In each iteration, the particle updates its velocity and position using Equation (2) and Equation (3), respectively:

Velocity update formula:

$$v_{id}^{t+1} = \omega \cdot v_{id}^t + c_1 \cdot r_1 \cdot (p_{id}^t - x_{id}^t) + c_2 \cdot r_2 \cdot (p_{gd}^t - x_{id}^t) \quad (2)$$

where r_1 and r_2 are random numbers in the range $[0, 1]$, which are used to introduce randomness into the search process; v_{id}^t and x_{id}^t represent the velocity and position of the i -th particle in the d -th dimension at the t -th iteration, respectively; p_{id}^t and p_{gd}^t represent the best fitness value and the global best fitness value of the i -th particle in the d -dimension at the t -th iteration, respectively.

Position update formula:

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (3)$$

2.3. Feature Introduction

In data-driven approaches for battery SOC prediction, the input features to the model must be directly measurable by sensors, typically including voltage, current, and temperature. Specifically, voltage serves as an indicator of the battery's real-time discharge capacity to a certain extent, while current reflects its instantaneous charging or discharging status. Temperature directly influences the kinetics of internal electrochemical reactions, such as lithium intercalation and deintercalation, as well as lithium-ion activity, thereby significantly affecting key battery performance metrics, including internal resistance and capacity [39].

Using 25 °C as a typical test temperature, three different operating conditions were cyclically applied to the battery until its voltage reached the cut-off threshold, as illustrated in Figure 4.

Figure 4a depicts the current profiles for a single cycle under each operating condition, where negative current corresponds to discharge and positive current to charge. The three profiles are defined as follows:

DST simulates frequent and severe power fluctuations during typical vehicle operations (acceleration, cruising, deceleration, and regenerative braking), with a single-cycle duration of 360 s.

FUDS represents typical urban driving behavior, with a single-cycle duration of 1400 s.

US06 corresponds to aggressive driving and high-speed road conditions, with a single-cycle duration of 600 s.

Figure 4b–d present the measured battery voltage and current data obtained by cyclically applying the DST, FUDS, and US06 profiles, respectively, until the battery voltage decreased to the lower cut-off voltage.

In SOC estimation, the primary features used for model training—voltage, current, and temperature—are subjected to measurement errors and noise. Deep learning models are highly sensitive to such minor fluctuations, and even slight noise can lead to significant deviations in prediction outcomes. Furthermore, SOC estimation depends on the internal physical and chemical properties of the battery, which are not explicitly represented in purely data-driven approaches. As a result, models may overreact to noise or anomalies in the input data, leading to unstable SOC predictions.

Since sensors collect battery states in real time, the voltage exhibits notable variations during current changes, especially during transitions between charging and discharging

states. For example, under DST conditions, the voltage difference can reach approximately 0.6 V when the battery switches from a resting state to a 4 A discharge. In addition, measurement noise can induce transient spikes in current and voltage. Feeding such dynamic and noisy operating states into an LSTM network may interfere with the model's ability to learn the inherent nonlinear behavior of the battery, potentially causing pronounced fluctuations in the predicted SOC and divergence from the true value.

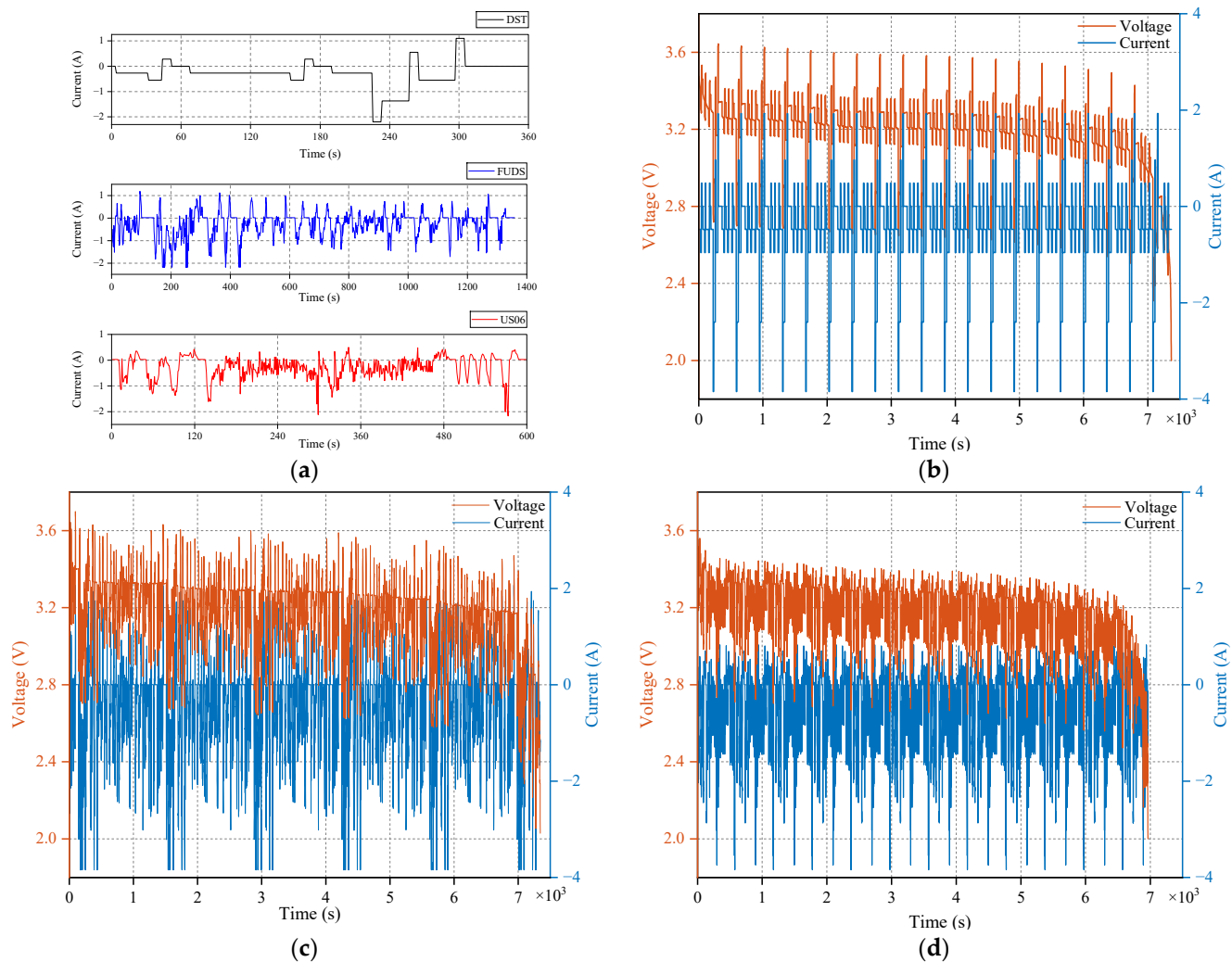


Figure 4. Current and voltage data under various operating conditions at 25 °C. (a) Current settings for a single cycle of three operating conditions; (b) DST; (c) FUDS; (d) US06.

To address this issue, a newly constructed feature—averaged voltage—is introduced as an additional input to the LSTM network. Unlike the real-time measured voltage that reacts sharply to instantaneous current changes, the averaged voltage reflects the overall charge–discharge state of the battery over a specific time window and approximates the “quasi-steady-state” voltage. It serves as an anchor point that points to the OCV for the LSTM network. This significantly reduces the difficulty for the model to infer the remaining discharge capacity of the battery from fluctuating voltage signals. By incorporating such slow time-varying information, the network can better learn the nonlinear mapping related to the battery's internal characteristics, even under large fluctuations in operating current and voltage. Moreover, the averaged voltage effectively smooths out transient voltage spikes caused by abrupt current changes and measurement noise, thereby stabilizing SOC predictions.

Instead of replacing the original time-series voltage data, we retain both the raw voltage and the averaged voltage as parallel inputs to the LSTM. This dual-input framework enables the model to simultaneously capture instantaneous voltage dynamics and longer-term averaged behavior, enhancing its ability to perceive the temporal evolution of the battery's historical charge–discharge states.

Regarding the LSTM's gating mechanism, although LSTM can memorize historical information, this capability is achieved through the implicit and abstract manner of the hidden state h_t and cell state c_t in its unit. The averaged voltage provides an explicit summary of past voltage levels. This injects useful prior knowledge into the model, reduces the memory burden on the LSTM's internal state, and allows it to focus more on learning other complex dynamic and nonlinear relationships.

The averaged voltage is calculated as follows:

$$\bar{V}_t^k = (V_{t-k+1} + \dots + V_{t-1} + V_t) / k \quad (4)$$

where k is the window size for calculating the averaged voltage, i.e., the number of consecutive voltage samples used. When the sampling instant $t < k$, a truncated window strategy is applied, with the arithmetic mean computed only over the available t samples.

The averaged voltage at each time step is defined as the mean of the instantaneous voltages within a backward time window that starts at and includes the current voltage sample. The principle of averaged voltage acquisition is illustrated in Figure 5.

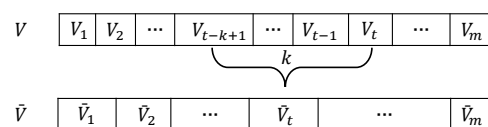


Figure 5. Averaged voltage acquisition principle.

Taking data from the US06 operating condition at 25 °C as an example, the averaged voltage calculated with k values of 25, 50, and 75 is shown in Figure 6.

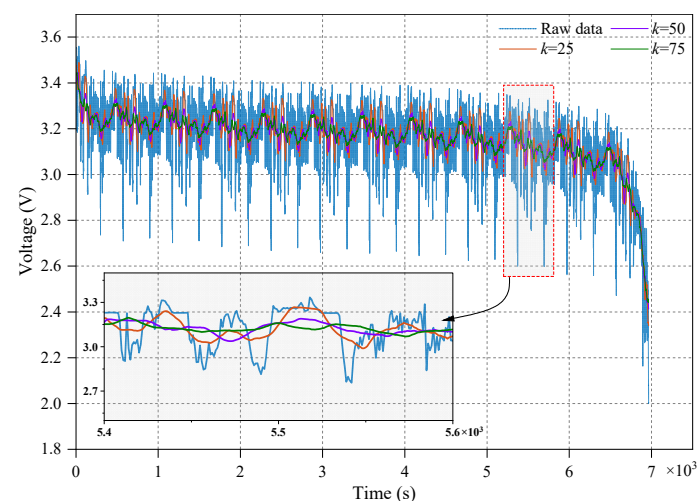


Figure 6. Averaged voltage with different k values.

The averaged voltage effectively suppresses transient time-varying signals while preserving its capacity to indicate the remaining battery capacity as the operating condition cycles between charge and discharge, the averaged voltage correspondingly rises and falls. However, when the window size k is small, for example, $k = 25$, the averaged voltage still exhibits considerable variation during abrupt voltage changes. Conversely, when k is large, for example, $k = 75$, the averaged voltage fails to track the actual voltage dynamics

and cannot synchronize as the battery voltage changes over time. When $k = 50$, the averaged voltage successfully balances these aspects: it mitigates the influence of time-varying voltage signals while accurately reflecting both the voltage fluctuations and the remaining capacity of the battery.

2.4. Limit Output

Besides adding the averaged voltage to the LSTM input layer to improve the model's learning of battery nonlinear characteristics, certain limitations can also be added to the output layer to enhance the reliability of LSTM prediction. Specifically, when the LSTM is used for battery SOC prediction, the LSTM network is often trained using data obtained from cyclic loading under a specific condition to acquire network weight parameters, which are then utilized to predict SOC under other conditions. Although the LSTM has been trained to learn nonlinear features, it is essentially a "black box" as a data-driven model. Under complex actual conditions, the prediction output of the model may occasionally violate the basic electrochemical physical laws, resulting in obvious illogical "abnormal points" or "jitter". These prediction errors not only deteriorate the estimation accuracy but also seriously compromise the reliability and safety of the BMS.

This section proposes an LSTM output post-processing strategy based on physical consistency constraints. The core idea of this strategy is to use the current direction, an indisputable physical information, as the "golden criterion" to evaluate the rationality of the SOC prediction value, and perform real-time inspection and correction of the original output of the LSTM.

Additionally, given that current information is required for output limiting, the impacts of current sensor noise or zero drift must be taken into account. Through the analysis of battery current time-series data, it is found that the absolute current value exceeds 10^{-3} A when the battery is in an operating state, whereas the current magnitude is on the order of 10^{-4} A during the resting state. Accordingly, a current threshold judgment mechanism is established: the battery is deemed to be in the resting state when the absolute current value is less than 10^{-3} A, and the current is set to 0 A in such cases.

Take the LSTM trained under the DST condition to predict the data under the US06 condition at 25 °C as an example. Overall, the LSTM can predict the SOC trend and quickly respond and approach the true value in the initial stage. In the position marked by the red ellipse in Figure 7, it can be observed from the current change (blue line) that the first half of this position is the charging part (current > 0 A), and the second half is the static part (current = 0). In the second half, the current is 0 A, so the battery is in the rest stage, and the SOC during this period should remain unchanged (the black line is horizontal). However, according to the LSTM prediction result (orange line), the SOC is predicted to decrease during this period, suggesting that the battery is discharging. Clearly, the LSTM prediction for this segment is inaccurate, indicating an error in the qualitative analysis of the battery state, specifically, an error in the qualitative assessment of whether the SOC value increases, decreases, or remains the same. In the first half, the current is greater than 0 A, indicating that the battery is in the charging state, and the true SOC of the battery should be in the rising state (the slope of the black line is greater than 0). At this time, the SOC value predicted by the LSTM is indeed rising, and there is no qualitative error in the battery SOC value. However, the true SOC value does not rise as much as the LSTM predicted value. It can be seen from the true SOC value that the slope of the black line is only slightly greater than 0, indicating that the LSTM may have quantitative analysis errors even when the qualitative analysis is correct in battery SOC prediction. This is also one of the sources of LSTM prediction errors. Therefore, it is necessary to start from the output layer and add a certain output limitation to enhance the accuracy and stability of SOC prediction.

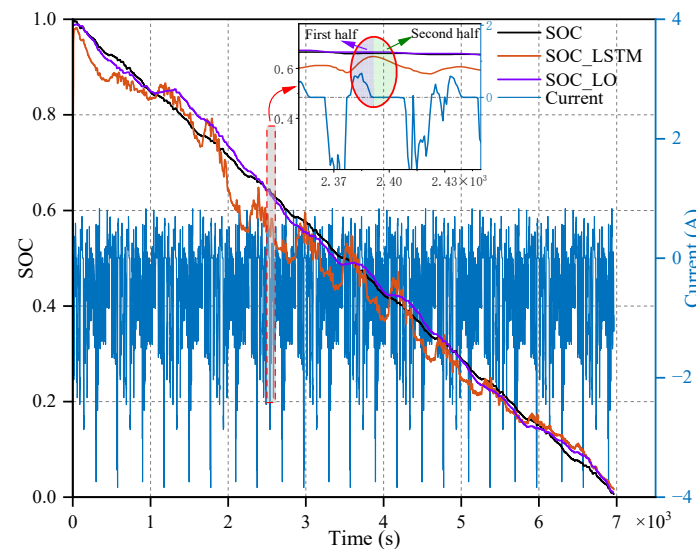


Figure 7. SOC prediction performance of LSTM model under US06 conditions at 25 °C.

The limitation process of the output layer is shown in Table 1.

Table 1. Output layer limitation process.

Step	Name	Description
1	Input Initial Values	Input the LSTM predicted value at time t (SOC_t^L), battery operating current at time t (I_t), predicted SOC value at time $t - 1$ (SOC_{t-1}), sampling frequency ($\Delta t = 1$ s), battery capacity (C), and compensation factor (λ).
2	Qualitative Analysis	- If $(SOC_t^L - SOC_{t-1}) \times I_t > 0$, proceed to step 3 (Quantitative analysis); - If $(SOC_t^L - SOC_{t-1}) \times I_t \leq 0$, replace the predicted value at time t with the predicted value at the previous time, i.e., $SOC_t = SOC_{t-1}$, proceed to step 4 (Output predicted value).
3	Quantitative Analysis	- If $SOC_t^L - SOC_{t-1} > I_t \times \Delta t /C$, set $SOC_t = SOC_{t-1} + (I_t \times \Delta t)/C \times \lambda$; - If $SOC_t^L - SOC_{t-1} \leq I_t \times \Delta t /C$, set $SOC_t = SOC_{t-1}$.
4	Output Predicted Value	Output SOC_t as the final predicted SOC value at time t .

To minimize the prediction error in LSTM output, it is essential to perform a qualitative analysis of the LSTM prediction results to confirm that the output value aligns with the charge–discharge or static state. Assume the sampling frequency is denoted as Δt . For SOC prediction at any given time, such as time t , based on the SOC value at that time t (SOC_t^L) predicted by LSTM, the predicted value at the previous moment (SOC_{t-1}), and the battery operating current at time t (I_t), evaluate the sign of the product of value $|SOC_t^L - SOC_{t-1}|$ and current I_t . If the product is positive, it indicates that the LSTM prediction is qualitatively correct, i.e., the predicted SOC change is in the same direction as the current direction. Specifically, when the current is positive, the battery is charged, and the SOC rises, so the predicted value of LSTM at time t should be greater than the predicted value at the previous time. When the current is negative, the battery discharges and the SOC decreases, so the predicted value of LSTM at time t should be less than that at the previous time. If the product is negative, this indicates that the LSTM prediction is qualitatively incorrect. At this point, the predicted value of LSTM at time t is deemed incorrect, and the predicted value at the previous time is used to replace it.

After qualitatively correcting the LSTM output, quantitative analysis is performed on the output. After the LSTM output meets the qualitative requirements, evaluate whether the variation of the predicted SOC value at time t relative to the predicted SOC value at time $t - 1$ (i.e., $|SOC_t^L - SOC_{t-1}|$) is greater than the ratio of the capacity variation

within Δt at time t to the total battery capacity (i.e., $|I_t \times \Delta t|/C$). If not, replace the current predicted value with the predicted value at the previous time; if yes, the current predicted SOC value (SOC_t) is obtained by adding $|I_t \times \Delta t|/C$ to the predicted SOC value at the previous time (SOC_{t-1}). Since there is a case where the predicted SOC value at the previous time is directly assigned to the current time as the predicted SOC value for output in the above process, the predicted value at the previous time has accumulated a certain error. A compensation factor λ is introduced here and multiplied by $|I_t \times \Delta t|/C$ to compensate for the SOC error. At the same time, an optimal value of λ needs to be determined, and PSO is adopted to optimize λ in this paper.

In Figure 7, the purple curve represents the prediction results after applying the output-limiting strategy (denoted as SOC_LO). As highlighted by the red ellipse, the output results now conform to the physical characteristics of the battery's operating current, with predicted values much closer to the true values. Two key correction effects of physical constraints are clearly reflected:

- In the latter half of the region marked by the red ellipse, the battery operating current is 0 A (resting state). The original LSTM predictions exhibited SOC fluctuations that violated the electrochemical principle (SOC should remain stable during resting periods), while the purple curve (with physical constraints) maintains a horizontal trend, effectively correcting this qualitative deviation.
- In the first half of the marked region, although the original LSTM predictions aligned with the qualitative trend of the limit output strategy, they failed to meet quantitative requirements (i.e., the magnitude of SOC change did not match the current change amplitude). The purple curve adjusts this deviation by reducing the slope of SOC variation, ensuring consistency between the prediction increment and the actual current change.

2.5. EKF Filtering

By incorporating the averaged voltage as an additional input feature, the LSTM learns the internal nonlinear mapping of the battery. While a physics constraint strategy is applied to correct outputs that deviate from basic electrochemical and physical principles, the predictions at this stage still exhibit limited dynamic tracking capability and remain susceptible to noise. To further improve accuracy and enhance noise robustness, an EKF is employed as a dynamic optimizer, connected to the output of the "FI + LSTM + LO" framework (i.e., the LSTM_FILO model).

The initial prediction from the LSTM_FI model serves as the initial value of the state prediction equation. Through adaptive adjustment of the Kalman gain, a probabilistically weighted fusion is performed between the data-driven output and EKF estimate. This integration strengthens the model's dynamic tracking performance, capability, provides confidence support for the safety decision-making of the BMS, and further refines the state prediction framework that combines physical insight, data-driven learning, and filtering techniques.

The EKF is applied to smooth the output of the LSTM_FILO model, and the state-space representation is described as follows:

State equation:

$$x_k = f(x_{k-1}, u_{k-1}(3)) + \omega = x_{k-1} + \frac{I_{k-1} \times \Delta t}{C \times 3600} + \omega \quad (5)$$

where state-space vector: $x = [SOC]$; initial SOC value: $SOC_0 = SOC_0^{LSTM_FI}$; $u = [\bar{V}, V, I, T]$; $u(3)$ denotes I , the battery operating current (with positive values for charging and negative values for discharging); C is the standard battery capacity; Δt is the data

sampling frequency; $\omega \sim N(0, Q)$ is Gaussian process noise, and Q is related to the acquisition accuracy and noise of the sensor.

Observation equation:

$$y_k = h(u_k) + v = \text{LSTM_FILO}(u_k) + v \quad (6)$$

where LSTM_FILO() is the prediction result of the LSTM_FILO model; $v \sim N(0, R)$ is Gaussian measurement noise, and R is related to the accuracy and fluctuation of the LSTM_FILO output.

EKF prediction (time update) phase:

- Predict the future state (a priori): $\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k$.
- Predict the error covariance: $\hat{P}_{k+1|k} = A\hat{P}_{k|k}A^T + Q_k$.

EKF correction (measurement update) phase:

- Calculate the Kalman gain: $K_{k+1} = P_{k+1|k}C^T(CP_{k+1|k}C^T + R_{k+1})^{-1}$.
- Update the estimation with measurements (a posteriori):
 $\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k})$.
- Update the error covariance: $\hat{P}_{k+1|k+1} = (1 - K_{k+1}C)P_{k+1|k}$.

where $A = [1]$, $B = [0, 0, \frac{\Delta t}{C \times 3600}, 0]$, $C = [\frac{\partial f(x_{k-1}, u_{k-1}(3))}{\partial SOC}] = [1]$. The error covariance matrix P represents the magnitude of the state estimation error of LSTM_FI.

2.6. Model Framework and Parameter Settings

This paper incorporates the averaged voltage as an augmented input feature and applies physics constraints to the LSTM output layer to achieve high-precision SOC estimation. The overall workflow, including the model vectors and the detailed procedures of feature introduction and output limitation, is illustrated in Figure 8. For the input time-series data of voltage (V), current (I), and temperature (T), the averaged voltage ($\bar{V}$) is first computed over a defined window. The averaged voltage sequence, together with the original voltage, current, and temperature data, forms the input dataset fed into the LSTM network. Subsequently, the trained network processes the data and outputs the predicted SOC (SOC_t^l) to the output limitation module. Within this module, the instantaneous current I_t and an optimized correction factor λ are used to adjust the predicted SOC value, yielding the refined output of the LSTM_FILO model SOC_t . This result serves as the observation input to an EKF.

The initial Ah value for the EKF state prediction equation is derived from the LSTM_FI model. The Kalman gain is adaptively adjusted to achieve a probabilistically weighted fusion between the data-driven model output and the EKF. This integration enhances the model's dynamic tracking capability and improves overall prediction accuracy.

In the feature introduction stage, incorporating averaged voltage as an input feature involves computing the average value over a predefined window, which can lead to variable input sequence lengths. To address this, if the prediction time step is shorter than the window size, the model directly calculates and utilizes the averaged voltage within the prediction time step. This design enables the model to capture both the real-time operating state from instantaneous voltage measurements and the overall charge–discharge capacity reflected by the averaged voltage.

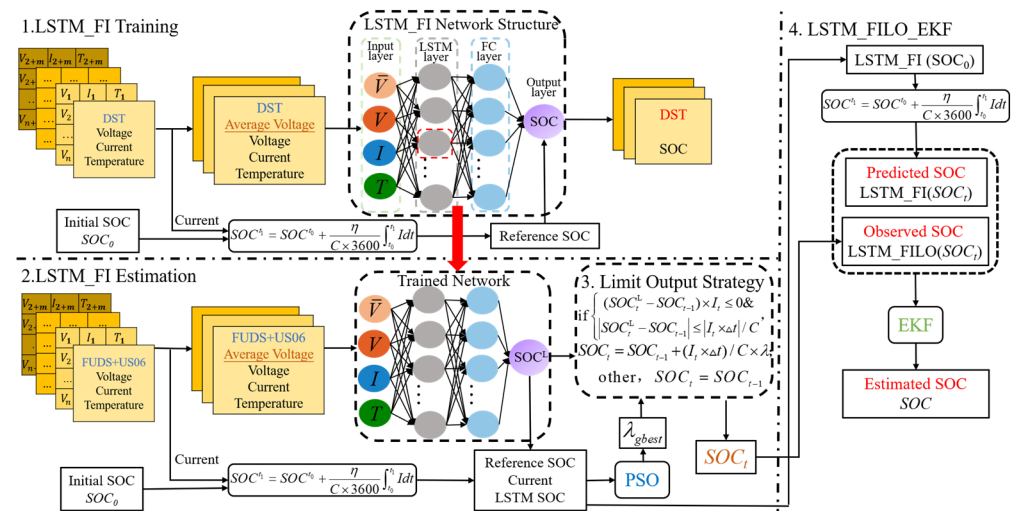


Figure 8. LSTM-EKF model flowchart.

Moreover, the hyperparameter configuration of the LSTM network significantly influences prediction accuracy. Regarding the number of hidden layers, too few layers may limit the model's ability to learn complex nonlinear relationships and deep time dependencies in the data, resulting in underfitting with high training and test errors. Conversely, an excessively deep network can exacerbate gradient vanishing or explosion issues, hinder convergence, and be more prone to memorizing noise in the training data rather than learning general patterns, resulting in poor performance on the test set and significantly extended training and prediction times. For battery SOC prediction, although deep LSTM networks may benefit from extremely long and complex sequences, the dependency relationships in SOC sequences typically do not require a very deep network structure.

With respect to the number of hidden units, an insufficient number restricts the model's "memory capacity", impairing its ability to capture rich information and complex patterns in the input sequence and causing underfitting; an excessively large number, however, raises the risk of overfitting and reduces generalizability to unseen data. The sequence length is also a crucial parameter, indicating the number of consecutive time steps in a single LSTM input sample. Too short a sequence may deprive the model of sufficient historical context for accurate predictions, while an overly long sequence could introduce irrelevant or outdated information, thereby degrading performance.

The hyperparameter settings used for the LSTM network in this study are summarized in Table 2.

Table 2. Hyperparameters of LSTM network.

Type	Hyperparameter	Value
Network Structure	Number of Hidden Layers	1
	Number of Hidden Units	32
Data Structure	Time Series Length	200
	Sampling Frequency	1 s
	Input Data Normalization Range	[−1, 1]
	Output Layer Activation Function	Sigmoid
	Optimizer	Adam
Training Process	Initial Learning Rate	0.01
	Minimum Batch Size	64
	Number of Training Epochs	500
	Loss Function	MSE

3. Dataset and Evaluation Criteria

3.1. Dataset

To simulate the dynamic response, energy consumption economy, and durability of the battery in real scenarios from different dimensions, three representative standard driving conditions are selected for simulation and test analysis, namely DST, FUDS, and US06. The DST condition was originally developed by the United States Advanced Battery Consortium (USABC), which simulates frequent and severe power changes of vehicles in states such as acceleration, cruising, deceleration, and braking energy recovery. This paper uses the data collected under the DST condition as the training dataset. FUDS is mainly used to simulate the driving state of vehicles in typical urban road environments, serving as one of the basic conditions for fuel economy and emission certification, and an important basis for electric vehicle range calibration (such as EPA range). It is characterized by low speed and frequent starts and stops: the maximum speed is about 90 km/h, and the average speed is low (about 31.5 km/h), including acceleration, deceleration, idle speed, and low-speed cruising conditions, fully reflecting the characteristics of urban congested road conditions. US06 aims to simulate aggressive driving and high-speed road conditions, characterized by high speed (maximum speed 129 km/h, average speed 77.9 km/h) and high intensity: it has extremely high requirements for the peak power output of the powertrain, which can quickly expose the performance bottlenecks and thermal management problems of batteries, motors, or engines under extreme demand. Therefore, the data collected under FUDS and US06 conditions are used as the test dataset to evaluate the SOC estimation performance of the proposed model.

The experimental dataset utilized in this paper originates from the public dataset of the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland. This dataset was acquired by subjecting A123 batteries to cyclic loading under three conditions: DST, FUDS, and the US06 driving cycle, at temperatures of 0 °C, 10 °C, 20 °C, 25 °C, 30 °C, 40 °C, and 50 °C, respectively. This battery is manufactured by A123 Systems, LLC, USA.

The detailed parameters of A123 batteries are shown in Table 3.

Table 3. Parameters of A123 battery.

Battery Parameter	Specification (Value)
Nominal Capacity	1100 mAh
Battery Material	LiFePO ₄
Size	18 × 65 mm
Cut-off Voltage	2.0–3.6 V
Nominal Voltage	3.2 V
Charging Current	0.5 C (standard charging), 1.0 C (fast charging)
Standard Charging Method	0.5 C constant current charging to 3.6 V, then constant voltage charging at 3.6 V until the charging current ≤ 0.05 C

Figure 9 shows the voltage, current, temperature, and reference SOC values obtained under cyclic loading conditions of DST, FUDS, and US06 at 25 °C.

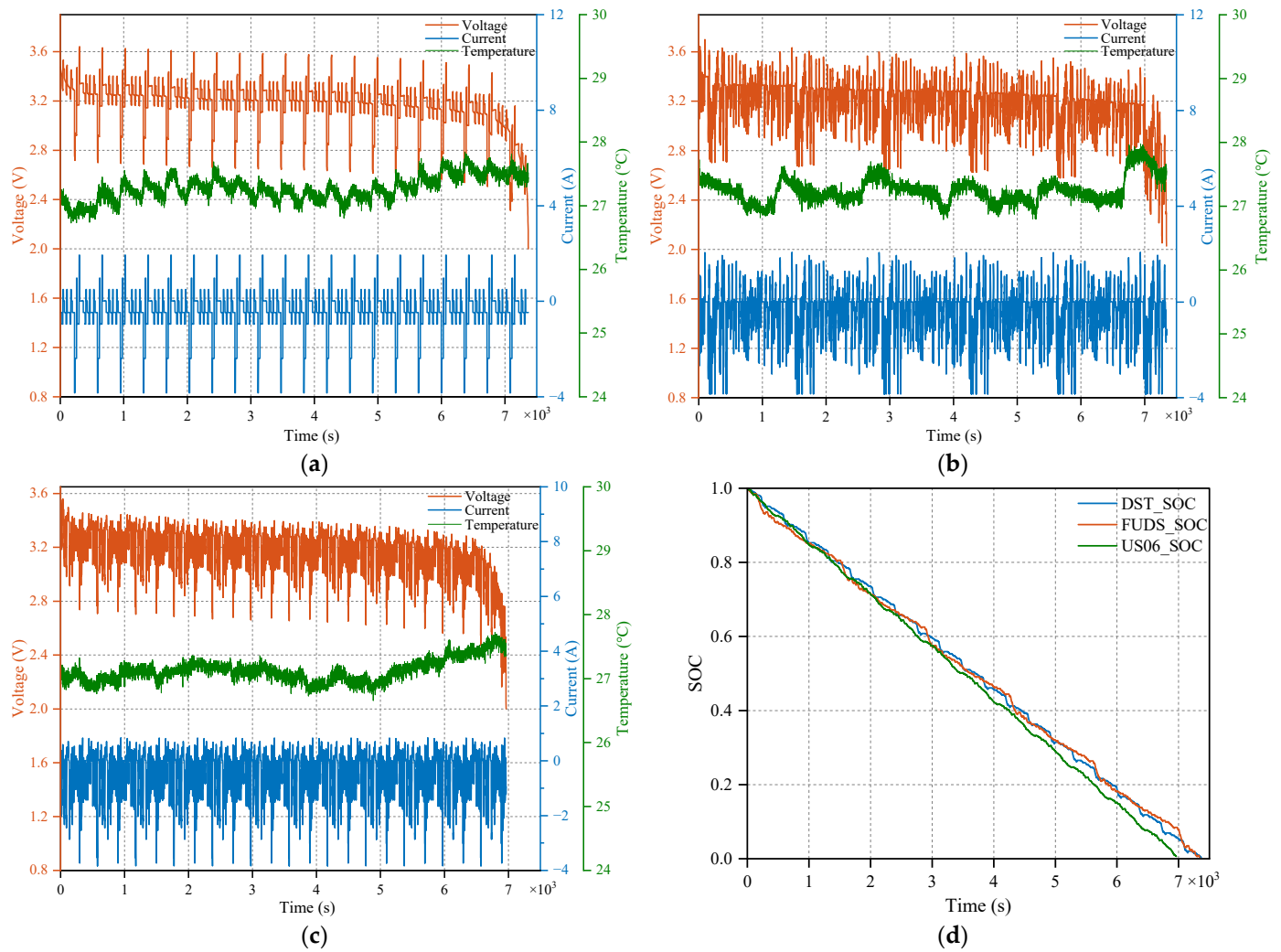


Figure 9. DST, FUDS and US06 data at 25 °C: (a) DST; (b) FUDS; (c) US06; (d) reference SOC.

Since the activation functions used in the core components (various gates) of LSTM are Sigmoid $[0, 1]$ and Tanh $[-1, 1]$, it is necessary to normalize the input data to the range matching these activation functions, i.e., $[-1, 1]$, to accelerate model convergence, improve training efficiency, and enhance model stability and generalization ability. This paper uses min-max normalization to map the battery measurement data (voltage, current, temperature, and average voltage) to the range of $[-1, 1]$. The equations used for the activation functions and normalization are as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

$$x_{norm} = \frac{2(x_t - x_{min})}{x_{max} - x_{min}} - 1 \quad (9)$$

where x_{max} and x_{min} are the maximum and minimum values of the measured variables in the dataset, respectively; x_t is the original value; x_{norm} is the corresponding normalized value, which is the actual input data of the network.

3.2. Evaluation Criteria

RMSE and MAE are common evaluation metrics for regression problems, which evaluate model performance by measuring the difference between estimated values and actual values. This paper uses these two metrics to evaluate the performance of the proposed model. The calculation methods of the evaluation metrics are as follows:

$$MAE = \frac{1}{N} \sum_{k=1}^N |y_k - y_k^*| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - y_k^*)^2} \quad (11)$$

where y_k is the reference SOC, and y_k^* is the estimated SOC.

Regarding the reference SOC, it is calculated by the Coulomb counting method. Although the Coulomb counting method is prone to be affected by the inaccurate initial SOC and the accumulated error induced by sensors, its use for obtaining SOC reference values is justified. Specifically, when a battery is fully charged in accordance with standardized charging procedures, its initial SOC can be considered 100%. Furthermore, the influence of current sensors is negligible under a single-cycle operating condition [48,49].

The formula of the Coulomb counting method is as follows:

$$SOC^{t_1} = SOC^{t_0} + \frac{\eta}{C \times 3600} \int_{t_0}^{t_1} Idt \quad (12)$$

where:

η : battery charge–discharge efficiency;

C : battery capacity, unit: Ah;

I : current, unit: A.

4. Results and Discussion

4.1. Prediction Results of Feature Introduction

In this section, we evaluate the SOC estimation performance of the LSTM_FI model with the introduced averaged voltage feature. According to the hyperparameters listed in Table 2, the LSTM network was trained on data from the DST cyclic condition and validated on data from the FUDS and US06 cyclic conditions. The reference SOC values were obtained by the Coulomb counting method.

With the average voltage window size k set to 50, the prediction results under various operating conditions and different temperatures are shown in Figures 10 and 11. In the Figures 10a and 11a, the black line represents the real SOC value; the blue line represents the LSTM predicted value and its absolute error compared to the true value; the orange line represents the LSTM_FI predicted value and its absolute error compared to the true value; the gray horizontal dash line represents $SOC = 0$.

The specific k value used in this study was determined by systematically comparing the SOC estimation performance under different average-voltage window sizes, as summarized in Table 4.

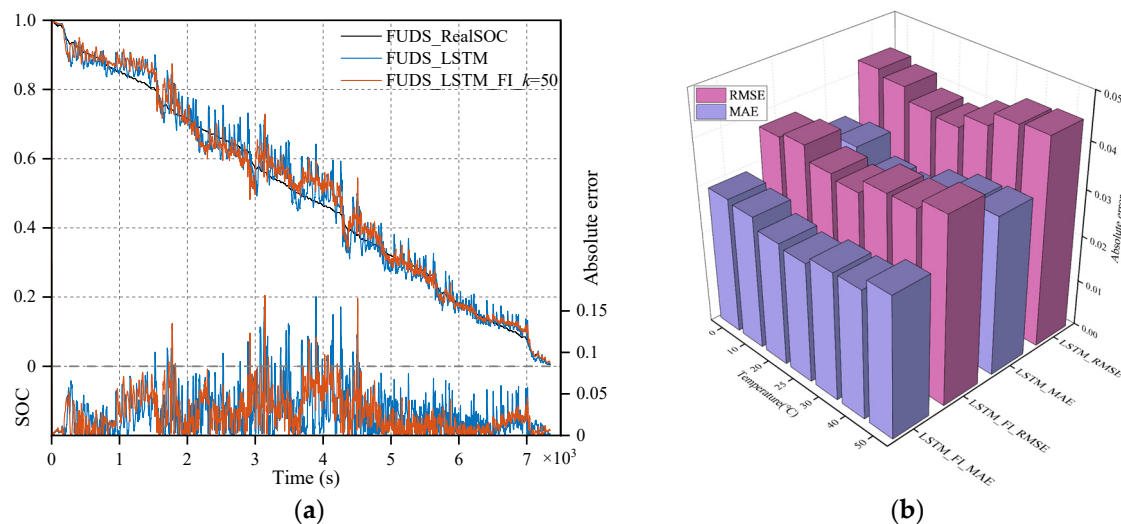


Figure 10. Prediction results of LSTM_FI ($k = 50$) under FUDS condition. (a) Optimal case: results at 25 °C; (b) average RMSE and MAE of prediction results for LSTM and LSTM_FI ($k = 50$) at various temperatures.

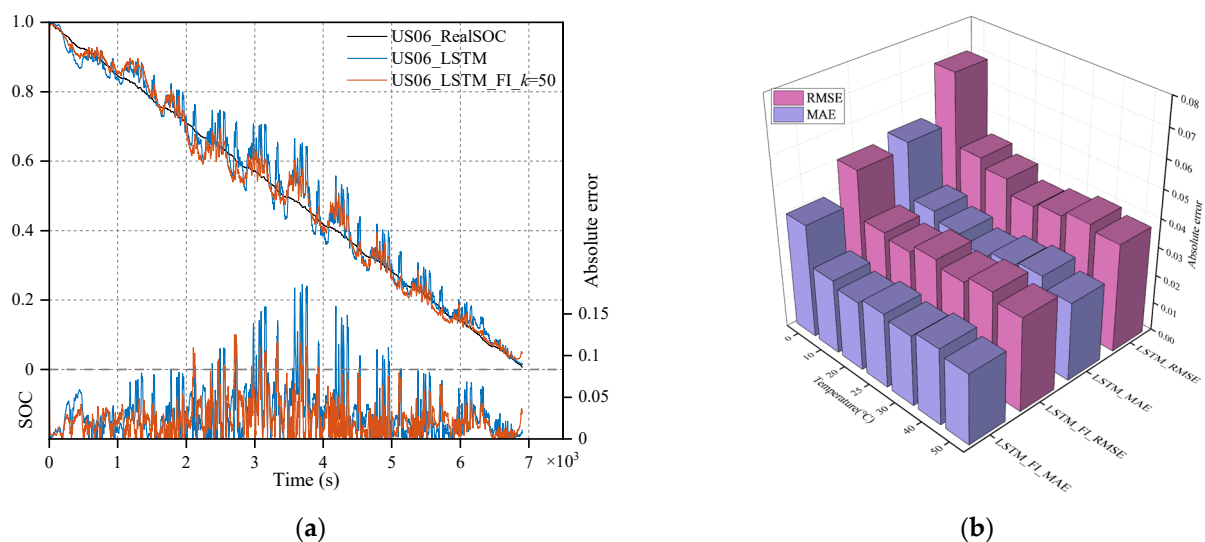


Figure 11. Prediction results of LSTM_FI ($k = 50$) under US06 condition. (a) Optimal case: results at 20 °C; (b) average RMSE and MAE of prediction results for LSTM and LSTM_FI ($k = 50$) at various temperatures.

Table 4. Results of feature introduction at different k values.

Validation Set	Temperature (°C)	LSTM		LSTM_FI ($k = 25$)		LSTM_FI ($k = 50$)		LSTM_FI ($k = 75$)	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
FUDS	0	3.21%	4.30%	3.67%	4.98%	2.93%	3.74%	3.76%	4.89%
	10	3.34%	4.17%	2.88%	3.72%	2.87%	3.85%	2.94%	3.65%
	20	2.87%	3.90%	2.82%	3.76%	2.64%	3.55%	3.34%	4.51%
	25	2.81%	3.83%	2.85%	3.75%	2.57%	3.46%	2.60%	3.27%
	30	3.15%	4.13%	2.93%	3.88%	2.73%	3.75%	2.59%	3.28%
	40	3.34%	4.42%	2.97%	3.99%	2.74%	3.76%	2.54%	3.27%
	50	3.39%	4.47%	3.00%	4.07%	3.00%	3.97%	2.65%	3.46%
US06	0	5.25%	6.95%	4.21%	5.42%	3.99%	5.05%	3.72%	4.78%
	10	3.29%	4.40%	3.61%	5.09%	2.60%	3.38%	3.30%	4.17%
	20	2.99%	4.10%	3.25%	4.65%	2.40%	3.23%	2.47%	3.35%
	25	2.70%	3.61%	2.88%	3.86%	2.60%	3.51%	3.08%	4.16%

30	2.93%	3.77%	3.01%	4.13%	2.46%	3.29%	2.63%	3.49%
40	3.11%	3.98%	3.00%	4.17%	2.65%	3.53%	2.58%	3.39%
50	2.74%	3.80%	3.02%	4.09%	2.86%	3.76%	2.93%	3.85%
Average	3.22%	4.27%	3.15%	4.25%	2.79%	3.70%	2.94%	3.82%

The boldfaced data in the table denote the temperature corresponding to the optimal prediction performance among different temperatures (0, 10, 20, 25, 30, 40, 50 °C) for the same prediction model under identical operating conditions.

The results indicate that the LSTM model is capable of capturing the overall trend of SOC, demonstrating its ability to learn the internal characteristics of the battery from the operating current, voltage, and temperature data. However, when the current changes abruptly under cyclic loading conditions, the predicted SOC exhibits increased fluctuations. This effect is more obvious under the FUDS condition than under US06. Unlike the US06 schedule, which simulates steady high-speed driving, the FUDS profile represents typical urban driving with frequent acceleration, deceleration, idling, and start–stop events. Consequently, the current output under FUDS is more variable, thus leading to rapid changes in battery voltage and ultimately increasing the volatility of SOC predictions.

Compared to the baseline LSTM, the LSTM_FI model with the averaged voltage as an additional input ($k = 25, 50, 75$) has improved SOC estimation results in terms of both MAE and RMSE. This confirms that adding the averaged voltage to the input features of the LSTM network is a simple, effective, and physically meaningful feature engineering strategy, significantly enhancing the accuracy, smoothness, and robustness of SOC estimation.

By comparing the average MAE and RMSE of LSTM_FI across different window sizes, it is evident that $k = 50$ yields better prediction performance, achieving an average MAE of 2.79% and an average RMSE of 3.7%. Compared to the baseline LSTM using only the three original input features (V, I, T), this represents an improvement of approximately 13% in both metrics. Therefore, the window size k is set to 50 for the averaged voltage feature in the proposed framework.

4.2. Prediction Results After Limit Output

In Section 4.1, SOC prediction accuracy was enhanced by introducing additional features into the LSTM, enabling the network to better learn the battery's internal characteristics. This section focuses solely on investigating the impact of the output limitation strategy on SOC prediction. The input to the LSTM remains the three directly measured variables: current, voltage, and temperature. Data from the DST cycling condition are again used as the training set, while data from the FUDS and US06 cycling conditions serve as the validation set. Reference SOC values are still obtained via the Coulomb counting method.

Since the limit output strategy introduces a correction factor λ to compensate for accumulated prediction errors, the optimization process and results of this factor are examined. First, set the parameters of PSO and initialize the particle swarm, and place the limit output strategy in the main PSO loop. Then, combine the MAE and RMSE of the prediction results relative to the reference SOC value into an optimization objective with equal weights, which serves as the fitness objective function of the particles. Finally, update the particle positions continuously to find the optimal value.

Taking the prediction process under the US06 condition at 25 °C as an example, Table 5 shows the parameter settings of PSO.

Table 5. Parameter settings of PSO.

Type	Hyperparameter	Value
Particle Parameter Setting	Number of Particles	20
	Initial Inertia Weight	0.9
	Minimum Inertia Weight	0.4
	Individual Learning Factor	2
	Group Learning Factor	2
Boundary Setting	Maximum Boundary Value	5
	Minimum Boundary Value	0.1
Optimization Process Setting	Maximum Number of Iterations	50
	Particle Fitness	$0.5 \times \text{MAE} + 0.5 \times \text{RMSE}$

The final gbest value of λ obtained through PSO is 1.0139.

Taking the SOC prediction under the US06 operating condition at 25 °C as an example, Figure 12a illustrates the convergence of the correction factor λ during the iterative optimization process toward its optimal value ($\lambda = \text{gbest}$). Figure 12b compares the SOC prediction performance achieved with the optimal λ (gbest) against that with other different λ values, clearly demonstrating the superior accuracy provided by the best-performing λ . Herein, the black line represents the real SOC values; the orange, blue, and purple lines represent the predicted values of the LSTM_LO model for $\lambda = 0.9$, gbest, and 1.1, respectively, as well as the corresponding absolute errors between the predictions and the true values; The gray horizontal dash line represents SOC = 0. The corresponding evaluation metrics for these prediction results are presented in Table 6.

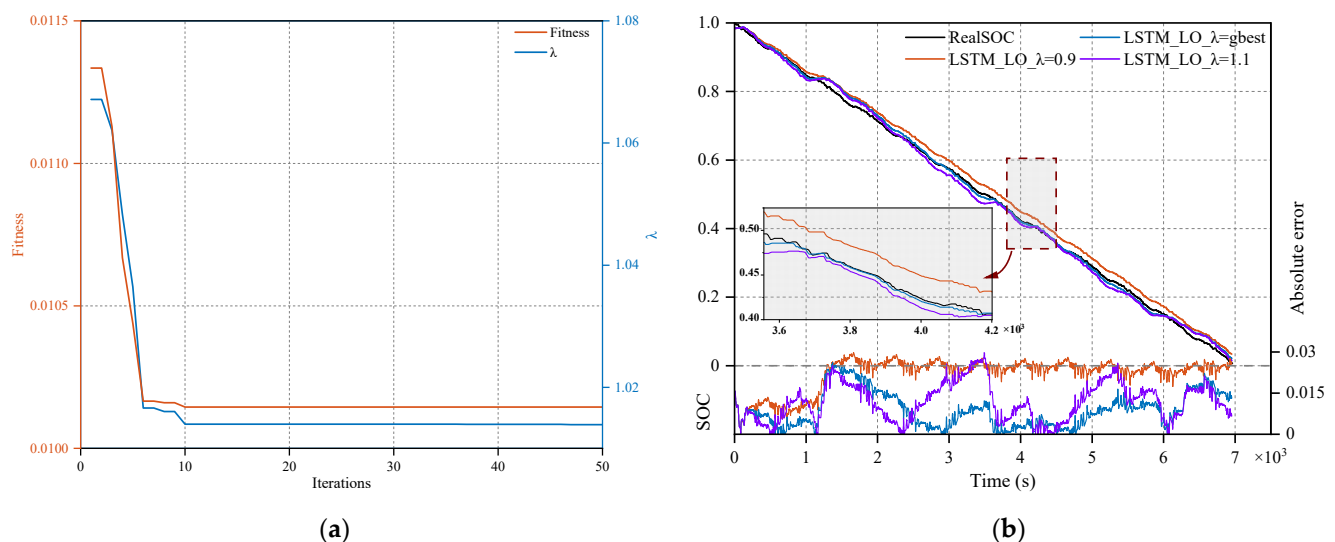


Figure 12. λ optimization process and comparison of the SOC prediction results under different λ values (US06 operating condition at 25 °C) (a) The optimization process of λ ; (b) The SOC prediction results under different λ values.

Table 6. Evaluation indicators of SOC prediction results at different λ values under US06 condition at 25 °C.

	LSTM_LO_ λ = 0.9	LSTM_LO_ λ = gbest	LSTM_LO_ λ = 1.1
MAE	2.32%	1.03%	1.16%
RMSE	2.39%	1.28%	1.34%

From the results, when the correction factor λ is set to gbest, the SOC prediction result is relatively better; when λ is set to a value near gbest, the prediction result deviates from the reference value.

When the correction factor λ is set to gbest, the SOC prediction results under various conditions and different temperatures are shown in Figures 13 and 14, respectively. In the Figures 13a and 14a, the black line represents the real SOC value; the blue line represents the LSTM predicted value and its absolute error compared to the true value; the orange line represents the LSTM_LO predicted value and its absolute error compared to the true value; the gray horizontal dash line represents SOC = 0.

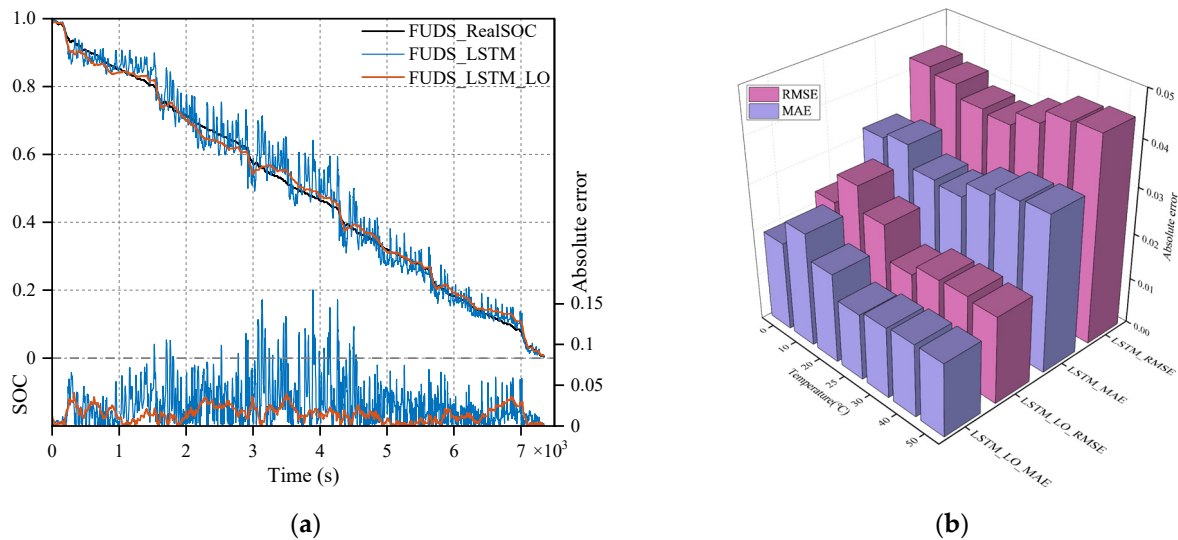


Figure 13. Prediction results of LSTM_LO under FUDS condition. (a) Optimal case: results at 25 °C; (b) average RMSE and MAE of prediction results for LSTM and LSTM_LO at various temperatures.

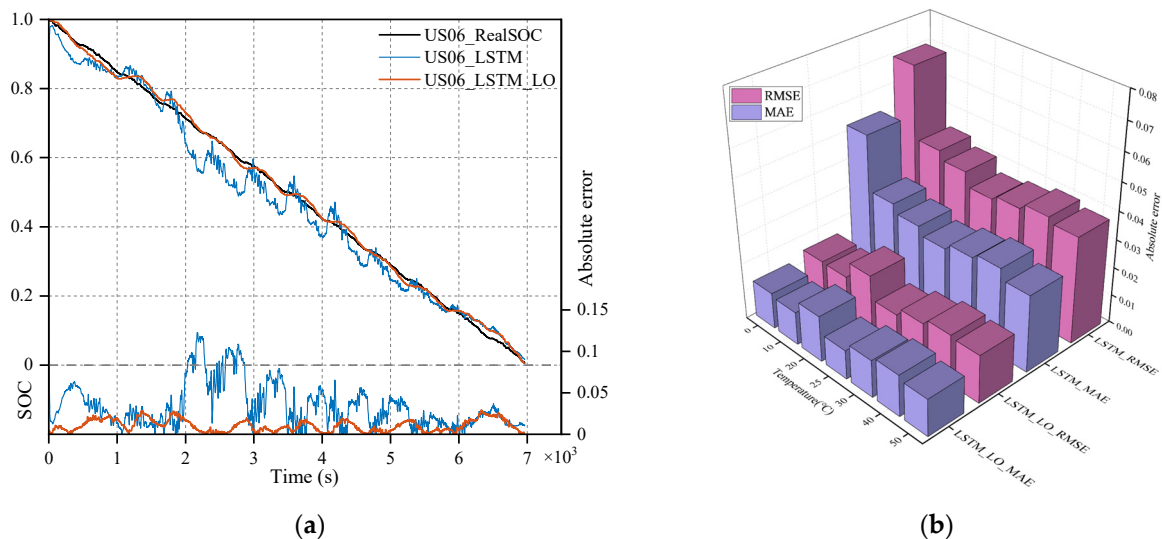


Figure 14. Prediction results of LSTM_LO under US06 condition. (a) Optimal case: results at 25 °C; (b) average RMSE and MAE of prediction results for LSTM and LSTM_LO at various temperatures.

From the visualization results, incorporating the physics-based constraint strategy into the LSTM output significantly improves the smoothness of the predictions and brings them closer to the true value. This is because feature introduction essentially adds an explicit summary of historical information to the LSTM, which reflects the average voltage level of the battery over the past time, to learn the complex dynamic nonlinear mapping characteristics of the battery during operation. However, the results are still output by the

LSTM, so any abrupt changes in the input features can still cause fluctuations in the output. In contrast, the limit output strategy operates directly on the LSTM's output, and constrains the output based on physical rules, thus eliminating the output result fluctuations caused by sudden variations in input features or measurement noise.

The evaluation metrics for the limit output strategy (LSTM_LO), feature introduction strategy (LSTM_FI), and the baseline LSTM are summarized in Table 7. The LSTM_LO model achieves an average MAE of 1.55% and an average RMSE of 1.90%. Compared to the LSTM_FI model with averaged voltage, this represents an improvement of 44.44% in MAE and 48.65% in RMSE. Relative to the LSTM, the improvements are 51.86% in MAE and 55.50% in RMSE.

Table 7. Comparison of results between limit output strategy, feature introduction strategy, and LSTM.

Validation Set	Temperature (°C)	LSTM		LSTM_FI $k = 50$		LSTM_LO	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
FUDS	0	3.21%	4.30%	2.93%	3.74%	1.92%	2.29%
	10	3.34%	4.17%	2.87%	3.85%	2.47%	2.96%
	20	2.87%	3.90%	2.64%	3.55%	1.94%	2.43%
	25	2.81%	3.83%	2.57%	3.46%	1.43%	1.69%
	30	3.15%	4.13%	2.73%	3.75%	1.53%	1.93%
	40	3.34%	4.42%	2.74%	3.76%	1.58%	1.95%
	50	3.39%	4.47%	3.00%	3.97%	1.58%	1.88%
US06	0	5.25%	6.95%	3.99%	5.05%	1.29%	1.52%
	10	3.29%	4.40%	2.60%	3.38%	1.18%	1.50%
	20	2.99%	4.10%	2.40%	3.23%	1.64%	2.11%
	25	2.70%	3.61%	2.60%	3.51%	1.03%	1.28%
	30	2.93%	3.77%	2.46%	3.29%	1.33%	1.57%
	40	3.11%	3.98%	2.65%	3.53%	1.48%	1.78%
	50	2.74%	3.80%	2.86%	3.76%	1.32%	1.71%
Average		3.22%	4.27%	2.79%	3.70%	1.55%	1.90%

The boldfaced data in the table denote the temperature corresponding to the optimal prediction performance among different temperatures (0, 10, 20, 25, 30, 40, 50 °C) for the same prediction model under identical operating conditions.

4.3. Prediction Results of Synergistic Feature Introduction and Limit Output

In Sections 4.1 and 4.2, the effects of feature introduction and limit output alone were analyzed separately. Feature introduction incorporates the averaged voltage with a window size of 50 as an additional feature input into the LSTM network, providing a slowly time-varying signal that reflects the overall charge–discharge state of the battery over a period of time. This reduces the model's difficulty in extracting the remaining battery capacity from fluctuating voltage measurements and ultimately improves the SOC prediction accuracy of the LSTM. The limit output strategy applies a filter based on fundamental electrochemical physics to the raw LSTM output, performing real-time validation and correction. A PSO-optimized correction factor λ is introduced to compensate for accumulated error.

This section applies both strategies synergistically to the LSTM network, forming the integrated LSTM_FILO model. The averaged voltage, together with the real-time voltage, current, and temperature, are fed into the LSTM. The resulting output is then processed by the limit output strategy to produce the final predicted SOC. The prediction results demonstrate the combined effect of feature introduction and limit output under different operating conditions and temperatures are shown in Figures 15 and 16, respectively. In

the Figures 15a and 16a, the black line represents the real SOC value; the blue line, orange line and purple line represent the predicted values of the models LSTM_FI, LSTM_LO and LSTM_FILO, as well as their absolute errors from the real values, respectively; the gray horizontal dash line represents SOC = 0.

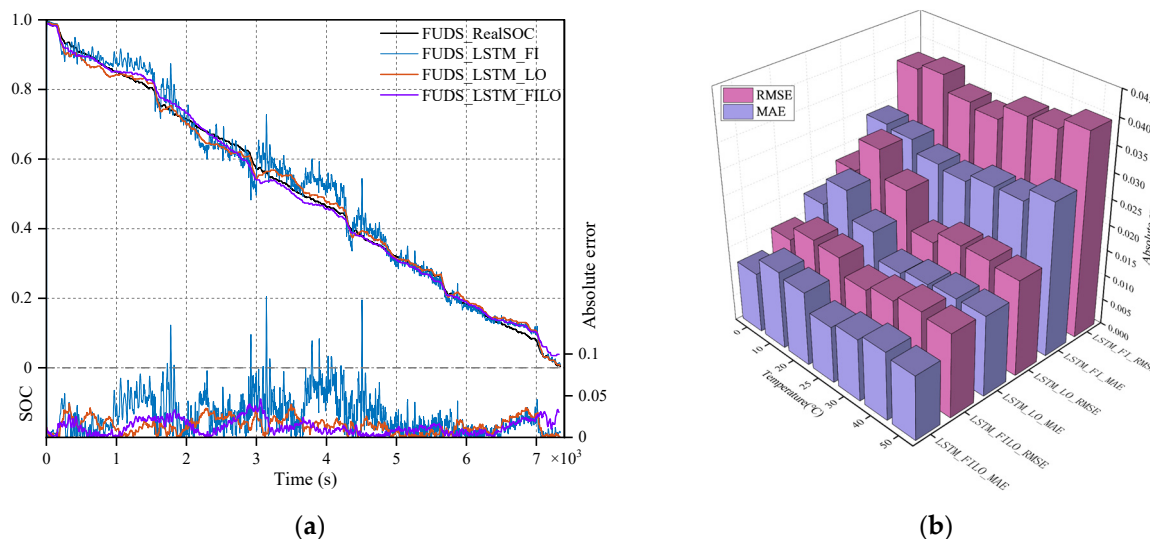


Figure 15. Prediction results of LSTM_FILO under FUDS condition. (a) Optimal case: results at 25 °C; (b) average RMSE and MAE of prediction results for LSTM_FI, LSTM_LO, and LSTM_FILO at various temperatures.

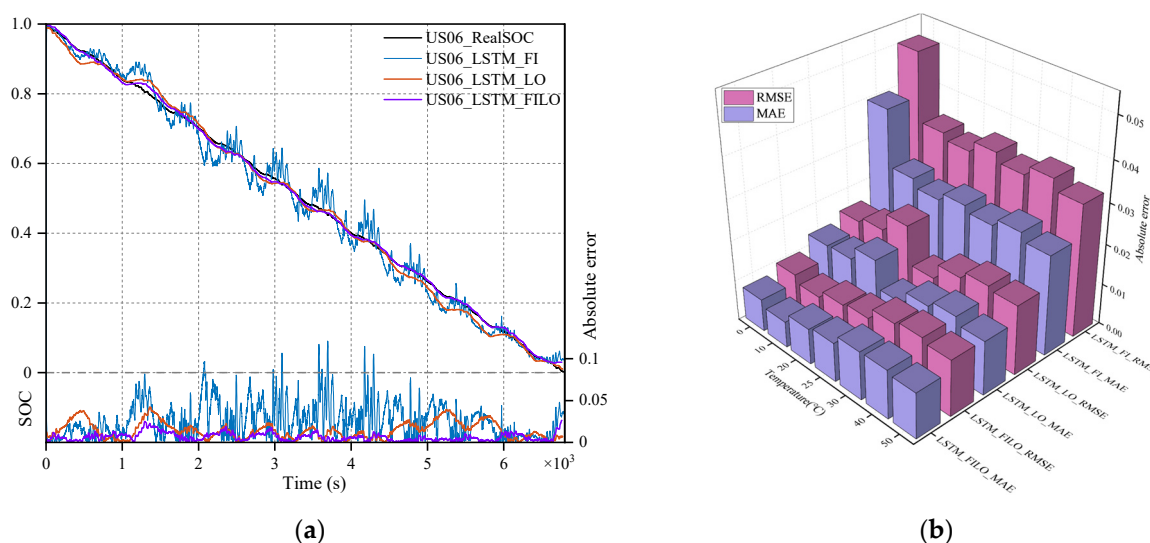


Figure 16. Prediction results of LSTM_FILO under US06 condition. (a) Optimal case: results at 10 °C; (b) average RMSE and MAE of prediction results for LSTM_FI, LSTM_LO, and LSTM_FILO at various temperatures.

Table 8 presents the evaluation metrics of the combined prediction results. It can be observed that when feature introduction and limit output are applied synergistically to the LSTM, the prediction accuracy is further enhanced compared to either strategy applied individually. Moreover, the prediction accuracy shows some variation in MAE and RMSE across different temperatures under the same operating condition. This variation is attributed to the significant temperature dependence of the battery's internal impedance. Nevertheless, the overall error of the combined model remains within an acceptable range, with average MAE and RMSE of 1.14% and 1.41%, respectively. Compared to the model using only feature introduction, this represents an improvement of 59.41% in MAE

and 61.89% in RMSE. Relative to the model using only limit output, the improvements are 26.45% in MAE and 27.79% in RMSE.

Table 8. Comparison of prediction results between limit output strategy, feature introduction, and synergistic approach.

Validation Set	Temperature (°C)	LSTM_FI		LSTM_LO		LSTM_FILO	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
FUDS	0	2.93%	3.74%	1.92%	2.29%	1.18%	1.53%
	10	2.87%	3.85%	2.47%	2.96%	1.53%	1.82%
	20	2.64%	3.55%	1.94%	2.43%	1.45%	1.78%
	25	2.57%	3.46%	1.43%	1.69%	1.11%	1.46%
	30	2.73%	3.75%	1.53%	1.93%	1.22%	1.58%
	40	2.74%	3.76%	1.58%	1.95%	1.34%	1.71%
	50	3.00%	3.97%	1.58%	1.88%	1.26%	1.64%
US06	0	3.99%	5.05%	1.29%	1.52%	0.91%	1.12%
	10	2.60%	3.38%	1.18%	1.50%	0.64%	0.82%
	20	2.40%	3.23%	1.64%	2.11%	0.87%	1.01%
	25	2.60%	3.51%	1.03%	1.28%	0.97%	1.10%
	30	2.46%	3.29%	1.33%	1.57%	1.19%	1.39%
	40	2.65%	3.53%	1.48%	1.78%	1.19%	1.47%
	50	2.86%	3.76%	1.32%	1.71%	1.05%	1.35%
Average		2.79%	3.70%	1.55%	1.90%	1.14%	1.41%

The boldfaced data in the table denote the temperature corresponding to the optimal prediction performance among different temperatures (0, 10, 20, 25, 30, 40, 50 °C) for the same prediction model under identical operating conditions.

4.4. Prediction Results of LSTM_FILO_EKF

In Section 4.3, the LSTM network was synergistically enhanced by integrating both the feature introduction and limit output strategies, leading to improved SOC prediction accuracy. However, the model still exhibited limitations in dynamic tracking capability and noise robustness. To further enhance accuracy and suppress noise, this subsection employs an EKF as a filter to process the output of the LSTM_FILO model, achieving a probabilistically weighted fusion of the data-driven model output and the EKF. This forms a state prediction model LSTM_FILO_EKF, integrating physical information, data-driven methods, and filtering, with detailed implementation steps provided in Section 2.5.

The prediction results of the proposed LSTM_FILO_EKF model under different operating conditions and temperatures are presented in Figures 17 and 18, respectively. In the Figures 17a and 18a, the black line represents the real SOC value; the blue line, orange line and purple line represent the predicted values of the models LSTM_LO, LSTM_FILO and LSTM_FILO_EKF, as well as their absolute errors from the real values, respectively; the gray horizontal dash line represents SOC = 0.

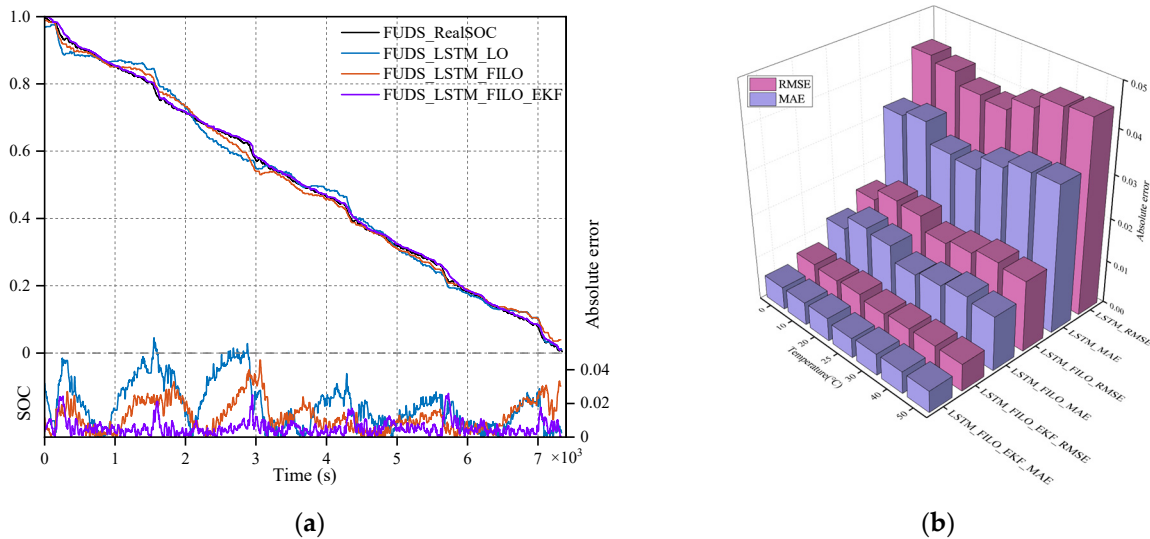


Figure 17. Prediction results of LSTM_FILO_EKF under FUDS condition. (a) Optimal case: results at 25 °C; (b) average RMSE and MAE of prediction results for LSTM, LSTM_FILO, and LSTM_FILO_EKF at various temperatures.

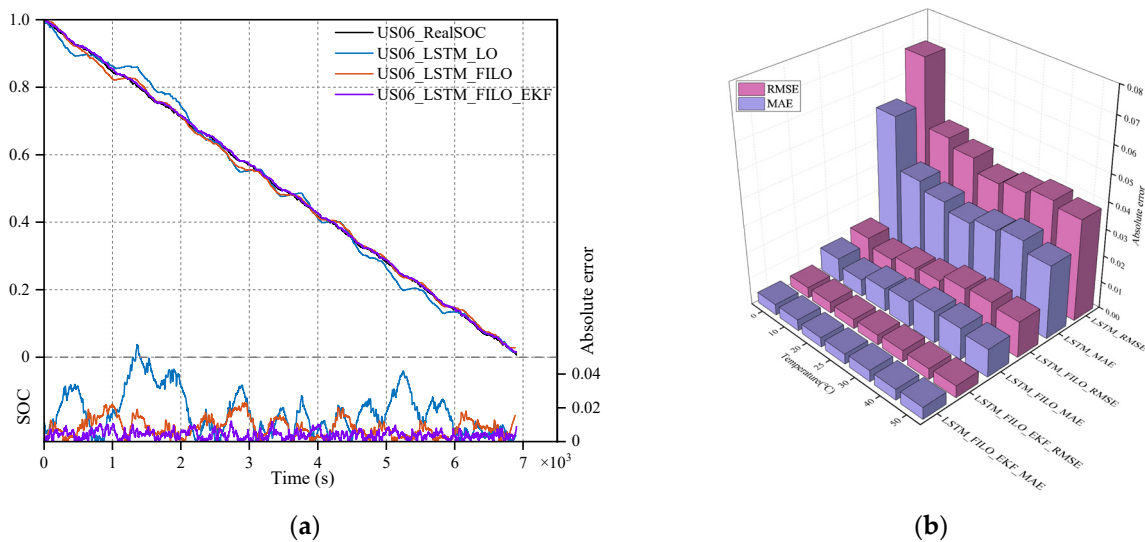


Figure 18. Prediction results of LSTM_FILO_EKF under US06 condition. (a) Optimal case: results at 20 °C; (b) average RMSE and MAE of prediction results for LSTM, LSTM_FILO, and LSTM_FILO_EKF at various temperatures.

Table 9 presents the evaluation metrics of the final prediction results from the proposed LSTM_FILO_EKF model across various temperatures in the validation set. The results demonstrate that applying the EKF to filter the output of the combined feature introduction and limit output strategies further enhances prediction accuracy and dynamic tracking capability. The model achieves an MAE of 0.46% and an RMSE of 0.56%, corresponding to an improvement of 59.65% in MAE and 60.28% in RMSE compared to the synergistic LSTM_FILO model, as well as an improvement of 85.71% in MAE and 86.89% in RMSE relative to the single LSTM baseline.

Table 9. Result comparison of LSTM_FILO_EKF model.

Validation Set	Temperature (°C)	LSTM		LSTM_FILO		LSTM_FILO_EKF	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
FUDS	0	3.21%	4.30%	1.18%	1.53%	0.53%	0.69%
	10	3.34%	4.17%	1.53%	1.82%	0.53%	0.65%

	20	2.87%	3.9%	1.45%	1.78%	0.54%	0.69%
	25	2.81%	3.83%	1.11%	1.46%	0.46%	0.60%
	30	3.15%	4.13%	1.22%	1.58%	0.50%	0.64%
	40	3.34%	4.42%	1.34%	1.71%	0.51%	0.64%
	50	3.39%	4.47%	1.26%	1.64%	0.50%	0.63%
US06	0	5.25%	6.95%	0.91%	1.12%	0.41%	0.49%
	10	3.29%	4.40%	0.64%	0.82%	0.42%	0.50%
	20	2.99%	4.10%	0.87%	1.01%	0.36%	0.43%
	25	2.70%	3.61%	0.97%	1.10%	0.38%	0.46%
	30	2.93%	3.77%	1.19%	1.39%	0.43%	0.49%
	40	3.11%	3.98%	1.19%	1.47%	0.41%	0.47%
	50	2.74%	3.80%	1.05%	1.35%	0.45%	0.51%
Average		3.22%	4.27%	1.14%	1.41%	0.46%	0.56%

The boldfaced data in the table denote the temperature corresponding to the optimal prediction performance among different temperatures (0, 10, 20, 25, 30, 40, 50 °C) for the same prediction model under identical operating conditions.

For benchmarking, the performance of the proposed model is compared against several established methods, including PIMNN [50], LSTM_RNN [34], Transformer [45], RFORC_LSTM [34], and LSTM&UKF [51]. The reported evaluation metrics represent the average values under two operating conditions at each temperature. As shown in Table 10, the proposed LSTM_FILO_EKF model outperforms other baseline models in the key evaluation metrics across the evaluated temperature range.

Table 10. Performance comparison between the proposed LSTM_FILO_EKF model and other baseline methods.

Validation Set	Temperature (°C)	PIMNN		LSTM_RNN	Transformer	RFORC_LSTM	LSTM&UKF		LSTM_FILO_EKF	
		MAE	RMSE	RMSE	RMSE	RMSE	MAE	RMSE	MAE	RMSE
FUDS	0	1.97%	2.70%	3.87%	1.81%	1.26%	—	—	0.53%	0.69%
	10	—	—	3.19%	1.34%	1.33%	—	—	0.53%	0.65%
	20	2.05%	2.60%	2.33%	2.69%	1.11%	—	—	0.54%	0.69%
	25	—	—	2.06%	1.14%	1.13%	—	—	0.46%	0.60%
	30	—	—	1.72%	1.18%	0.99%	—	—	0.50%	0.64%
	40	—	—	1.37%	1.47%	0.89%	—	—	0.51%	0.64%
	50	—	—	1.29%	3.27%	0.96%	—	—	0.50%	0.63%
US06	0	1.58%	1.99%	3.94%	2.80%	1.56%	0.63%	0.73%	0.41%	0.49%
	10	—	—	3.11%	2.56%	1.19%	0.21%	0.29%	0.42%	0.50%
	20	1.48%	2.35%	2.43%	1.99%	1.01%	0.97%	1.11%	0.36%	0.43%
	25	—	—	2.25%	2.21%	0.91%	0.82%	0.93%	0.38%	0.46%
	30	—	—	2.05%	2.71%	1.02%	0.81%	0.92%	0.43%	0.49%
	40	—	—	1.61%	2.10%	0.92%	0.89%	1.03%	0.41%	0.47%
	50	—	—	1.40%	1.84%	0.82%	0.93%	1.06%	0.45%	0.51%
Average		1.77%	2.41%	2.33%	2.08%	1.08%	0.75%	0.86%	0.46%	0.56%

The boldfaced data in the table denote the temperature corresponding to the optimal prediction performance among different temperatures (0, 10, 20, 25, 30, 40, 50 °C) for the same prediction model under identical operating conditions.

5. Conclusions

This paper proposes a synergistic model that integrates feature introduction and limit output strategies with an LSTM network, followed by EKF smoothing, to address

the challenge of accurate SOC estimation for lithium-ion batteries across varying temperatures and operating conditions. By transforming real-time voltage measurements within a defined window into an averaged voltage feature, a slowly varying signal is supplied to the LSTM input, enabling the model to capture the overall charge–discharge state of the battery from fluctuating voltage data. Furthermore, a physics-guided post-processing strategy is applied to the LSTM output to correct both qualitative and quantitative deviations of the data-driven prediction that violate electrochemical principles. Specifically, it addresses anomalies such as inconsistencies between the direction of SOC changes and the direction of battery current, as well as mismatches between the magnitude of SOC changes and the amplitude of current variation, thereby achieving more accurate and reliable SOC estimation. Finally, the refined SOC is used as the observation input to an EKF, enhancing the model’s dynamic tracking capability.

Trained on data from the DST operating profile at temperatures of 0 °C, 10 °C, 20 °C, 25 °C, 30 °C, 40 °C, and 50 °C, the proposed model achieves an average MAE of 0.46% and an average RMSE of 0.56% when predicting battery SOC under other operating conditions (FUDS, US06) at the same temperatures. The results demonstrate that the proposed model maintains robust performance across different temperatures and drive cycles. Moreover, it requires no additional sensors to acquire supplementary battery features and exhibits low memory usage, facilitating its practical deployment in BMS.

The synergy of introducing the averaged voltage feature at the input and applying physics-based constraints at the output not only compensates for the limitations of purely data-driven methods in modeling the internal nonlinear characteristics of batteries but also corrects predictions that are inconsistent with physical laws, guided by fundamental electrochemical principles. This synergistic design embodies the “physics-constrained” innovation of the research, ensuring that the predictions are both numerically accurate and physically interpretable. By incorporating EKF-based fusion, which probabilistically weights the data-driven outputs with the filter estimate, the model’s dynamic tracking performance is further improved. The proposed framework provides an effective technical pathway for state estimation under a unified “physical information-data-driven-filter fusion” paradigm, enabling accurate SOC estimation of lithium-ion batteries in multiple operational scenarios.

Author Contributions: Conceptualization, Y.S.; Supervision, J.D.; Resources, F.H.; Writing—original draft, S.Y.; Writing—review and editing, Y.S., S.Y., J.D. and F.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Fangfang Hu was employed by the company Beijing Products Quality Supervision and Inspection Research Institute (National Automotive Quality Inspection and Testing Center). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EVs	Electric Vehicles
BMS	Battery Management System
SOC	State of Charge
OCV	Open-Circuit Voltage

FF-RLS	Recursive Least Square with Forgetting Factor
EWMA	Exponentially Weighted Moving Average
PID	Proportional Integral Differential
AEKF	Adaptive Extended Kalman Filter
P2D	Pseudo-Two-Dimensional
ECM	Equivalent Circuit Model
OCVN	Open-Circuit Voltage Noise
LSTM	Long Short-Term Memory
TCN	Temporal Convolutional Network
GRU	Gated Recurrent Unit
WOA	Whale Optimization Algorithm
RF	Random Forest
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
SOH	State of Health
CNN	Convolutional Neural Network
SVR	Support Vector Regression
EKF	Extended Kalman Filter
DST	Dynamic Stress Test
US06	US06 High-Speed Driving Schedule
FUDS	Federal Urban Driving Schedule
NMC	Nickel Manganese Cobalt
ACKF	Adaptive Cubature Kalman Filter
AUKF	Adaptive Unscented Kalman Filter
FI	Feature Introduction
LO	Limit Output
PSO	Particle Swarm Optimization
RNN	Recurrent Neural Network

References

1. Li, Y.; Wei, Y.; Zhu, F.; Du, J.; Zhao, Z.; Ouyang, M. The path enabling storage of renewable energy toward carbon neutralization in China. *eTransportation* **2023**, *16*, 100226. <https://doi.org/10.1016/j.etrans.2023.100226>.
2. Zhai, M.; Wu, Y.; Tian, S.; Yuan, H.; Li, B.; Luo, X.; Huang, G.; Fu, Y.; Zhu, M.; Gu, Y.; et al. A circular economy approach for the global lithium-ion battery supply chain. *Nature* **2025**, *646*, 1114–1121. <https://doi.org/10.1038/s41586-025-09617-4>.
3. Nyamathulla, S.; Dhanamjayulu, C. A review of battery energy storage systems and advanced battery management system for different applications: Challenges and recommendations. *J. Energy Storage* **2024**, *86*, 111179. <https://doi.org/10.1016/j.est.2024.111179>.
4. Luo, K.; Chen, X.; Zheng, H.; Shi, Z. A review of deep learning approach to predicting the state of health and state of charge of lithium-ion batteries. *J. Energy Chem.* **2022**, *74*, 159–173. <https://doi.org/10.1016/j.jechem.2022.06.049>.
5. Tepe, B.; Jablonski, S.; Hesse, H.; Jossen, A. Lithium-ion battery utilization in various modes of e-transportation. *eTransportation* **2023**, *18*, 100274. <https://doi.org/10.1016/j.etrans.2023.100274>.
6. Hossain Lipu, M.S.; Hannan, M.A.; Karim, T.F.; Hussain, A.; Saad, M.H.M.; Ayob, A.; Miah, M.S.; Indra Mahlia, T.M. Intelligent algorithms and control strategies for battery management system in electric vehicles: Progress, challenges and future outlook. *J. Clean. Prod.* **2021**, *292*, 126044. <https://doi.org/10.1016/j.jclepro.2021.126044>.
7. Kurucan, M.; Özbaltan, M.; Yetgin, Z.; Alkaya, A. Applications of artificial neural network based battery management systems: A literature review. *Renew. Sustain. Energy Rev.* **2024**, *192*, 114262. <https://doi.org/10.1016/j.rser.2023.114262>.
8. Lu, W.; Yuan, Z.; Wang, T.; Li, P.; Zhang, Y. Will it get there? A deep learning model for predicting next-trip state of charge in Urban Green Freight Delivery with electric vehicles. *eTransportation* **2024**, *22*, 100372. <https://doi.org/10.1016/j.etrans.2024.100372>.
9. Ruan, H.; Barreras, J.V.; Engstrom, T.; Merla, Y.; Millar, R.; Wu, B. Lithium-ion battery lifetime extension: A review of derating methods. *J. Power Sources* **2023**, *563*, 232805. <https://doi.org/10.1016/j.jpowsour.2023.232805>.
10. Gräf, D.; Marschewski, J.; Ibing, L.; Hucklebrink, D.; Fiebrandt, M.; Hanau, G.; Bertsch, V. What drives capacity degradation in utility-scale battery energy storage systems? The impact of operating strategy and temperature in different grid applications. *J. Energy Storage* **2022**, *47*, 103533. <https://doi.org/10.1016/j.est.2021.103533>.

11. Diaz-Londono, C.; Fambri, G.; Maffezzoni, P.; Gruosso, G. Enhanced EV charging algorithm considering data-driven workplace chargers categorization with multiple vehicle types. *eTransportation* **2024**, *20*, 100326. <https://doi.org/10.1016/j.etrans.2024.100326>.
12. Kadem, O.; Kim, J. Real-Time State of Charge-Open Circuit Voltage Curve Construction for Battery State of Charge Estimation. *IEEE Trans. Veh. Technol.* **2023**, *72*, 8613–8622. <https://doi.org/10.1109/TVT.2023.3244623>.
13. Zhou, C.-C.; Su, Z.; Gao, X.-L.; Cao, R.; Yang, S.-C.; Liu, X.-H. Ultra-high-energy lithium-ion batteries enabled by aligned structured thick electrode design. *Rare Met.* **2022**, *41*, 14–20. <https://doi.org/10.1007/s12598-021-01785-2>.
14. Yu, Q.; Huang, Y.; Tang, A.; Wang, C.; Shen, W. OCV-SOC-Temperature Relationship Construction and State of Charge Estimation for a Series– Parallel Lithium-Ion Battery Pack. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 6362–6371. <https://doi.org/10.1109/TITS.2023.3252164>.
15. Zubi, G.; Dufo-López, R.; Carvalho, M.; Pasaoglu, G. The lithium-ion battery: State of the art and future perspectives. *Renew. Sustain. Energy Rev.* **2018**, *89*, 292–308. <https://doi.org/10.1016/j.rser.2018.03.002>.
16. Chemali, E.; Kollmeyer, P.J.; Preindl, M.; Ahmed, R.; Emadi, A. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Trans. Ind. Electron.* **2017**, *65*, 6730–6739.
17. Farmann, A.; Sauer, D.U. A study on the dependency of the open-circuit voltage on temperature and actual aging state of lithium-ion batteries. *J. Power Sources* **2017**, *347*, 1–13. <https://doi.org/10.1016/j.jpowsour.2017.01.098>.
18. Ng, K.S.; Moo, C.-S.; Chen, Y.-P.; Hsieh, Y.-C. Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries. *Appl. Energy* **2009**, *86*, 1506–1511. <https://doi.org/10.1016/j.apenergy.2008.11.021>.
19. Che, Y.; Xu, L.; Teodorescu, R.; Hu, X.; Onori, S. Enhanced SOC Estimation for LFP Batteries: A Synergistic Approach Using Coulomb Counting Reset, Machine Learning, and Relaxation. *ACS Energy Lett.* **2025**, *10*, 741–749. <https://doi.org/10.1021/acsenerylett.4c03223>.
20. Xiong, R.; Li, Z.; Li, H.; Wang, J.; Liu, G. A novel method for state of charge estimation of lithium-ion batteries at low-temperatures. *Appl. Energy* **2025**, *377*, 124514.
21. Zheng, F.; Xing, Y.; Jiang, J.; Sun, B.; Kim, J.; Pecht, M. Influence of different open circuit voltage tests on state of charge online estimation for lithium-ion batteries. *Appl. Energy* **2016**, *183*, 513–525.
22. Peng, S.; Zhang, D.; Dai, G.; Wang, L.; Jiang, Y.; Zhou, F. State of charge estimation for LiFePO₄ batteries joint by PID observer and improved EKF in various OCV ranges. *Appl. Energy* **2025**, *377*, 124435.
23. Zhu, Y.; Xiong, Y.; Xiao, J.; Yi, T.; Li, C.; Sun, Y. An improved coulomb counting method based on non-destructive charge and discharge differentiation for the SOC estimation of NCM lithium-ion battery. *J. Energy Storage* **2023**, *73*, 108917.
24. Lai, X.; Sun, L.; Chen, Q.; Wang, M.; Chen, J.; Ke, Y.; Zheng, Y. A novel modeling methodology for hysteresis characteristic and state-of-charge estimation of LiFePO₄ batteries. *J. Energy Storage* **2024**, *101*, 113807.
25. Zhang, K.; Xiong, R.; Li, Q.; Chen, C.; Tian, J.; Shen, W. A novel pseudo-open-circuit voltage modeling method for accurate state-of-charge estimation of LiFePO₄ batteries. *Appl. Energy* **2023**, *347*, 121406.
26. Yu, Z.; Tian, Y.; Li, B. A simulation study of Li-ion batteries based on a modified P2D model. *J. Power Sources* **2024**, *618*, 234376.
27. Streb, M.; Andersson, M.; Klass, V.L.; Klett, M.; Johansson, M.; Lindbergh, G. Investigating re-parametrization of electrochemical model-based battery management using real-world driving data. *eTransportation* **2023**, *16*, 100231. <https://doi.org/10.1016/j.etrans.2023.100231>.
28. Lai, X.; Zheng, Y.; Sun, T. A comparative study of different equivalent circuit models for estimating state-of-charge of lithium-ion batteries. *Electrochim. Acta* **2018**, *259*, 566–577.
29. Berger, F.; Joest, D.; Barbers, E.; Quade, K.; Wu, Z.; Sauer, D.U.; Dechent, P. Benchmarking battery management system algorithms—Requirements, scenarios and validation for automotive applications. *eTransportation* **2024**, *22*, 100355. <https://doi.org/10.1016/j.etrans.2024.100355>.
30. Xiong, R.; Duan, Y.; Zhang, K.; Lin, D.; Tian, J.; Chen, C. State-of-charge estimation for onboard LiFePO₄ batteries with adaptive state update in specific open-circuit-voltage ranges. *Appl. Energy* **2023**, *349*, 121581.
31. Severson, K.A.; Attia, P.M.; Jin, N.; Perkins, N.; Jiang, B.; Yang, Z.; Chen, M.H.; Aykol, M.; Herring, P.K.; Fraggadakis, D.; et al. Data-driven prediction of battery cycle life before capacity degradation. *Nat. Energy* **2019**, *4*, 383–391. <https://doi.org/10.1038/s41560-019-0356-8>.
32. Quan, R.; Liu, P.; Li, Z.; Li, Y.; Chang, Y.; Yan, H. A multi-dimensional residual shrinking network combined with a long short-term memory network for state of charge estimation of Li-ion batteries. *J. Energy Storage* **2023**, *57*, 106263. <https://doi.org/10.1016/j.est.2022.106263>.

33. Abdolrasol, M.G.M.; Ayob, A.; Lipu, M.S.H.; Ansari, S.; Kiong, T.S.; Saad, M.H.M.; Ustun, T.S.; Kalam, A. Advanced data-driven fault diagnosis in lithium-ion battery management systems for electric vehicles: Progress, challenges, and future perspectives. *eTransportation* **2024**, *22*, 100374. <https://doi.org/10.1016/j.etrans.2024.100374>.
34. Chen, J.; Li, K.; Liu, W.; Yin, C.; Zhu, Q.; Tang, H. A novel state of charge estimation method for LiFePO₄ battery based on combined modeling of physical model and machine learning model. *J. Energy Storage* **2025**, *115*, 115888.
35. Bao, G.; Liu, X.; Zou, B.; Yang, K.; Zhao, J.; Zhang, L.; Chen, M.; Qiao, Y.; Wang, W.; Tan, R. Collaborative framework of Transformer and LSTM for enhanced state-of-charge estimation in lithium-ion batteries. *Energy* **2025**, *322*, 135548.
36. Li, H.; Fu, L.; Long, X.; Liu, L.; Zeng, Z. A hybrid deep learning model for lithium-ion batteries state of charge estimation based on quantile regression and attention. *Energy* **2024**, *294*, 130834. <https://doi.org/10.1016/j.energy.2024.130834>.
37. Wan, S.; Yang, H.; Lin, J.; Li, J.; Wang, Y.; Chen, X. Improved whale optimization algorithm towards precise state-of-charge estimation of lithium-ion batteries via optimizing LSTM. *Energy* **2024**, *310*, 133185.
38. Sulaiman, M.H.; Mustaffa, Z. State of charge estimation for electric vehicles using random forest. *Green Energy Intell. Transp.* **2024**, *3*, 100177.
39. Chen, Z.; Zhao, H.; Shu, X.; Zhang, Y.; Shen, J.; Liu, Y. Synthetic state of charge estimation for lithium-ion batteries based on long short-term memory network modeling and adaptive H-Infinity filter. *Energy* **2021**, *228*, 120630.
40. Wang, S.; Jiao, M.; Zhou, R.; Ren, Y.; Liu, H.; Lian, C. A multi-dimensional machine learning framework for accurate and efficient battery state of charge estimation. *J. Power Sources* **2024**, *623*, 235417.
41. Xu, P.; Li, J.; Xue, Q.; Sun, F. A syncretic state-of-charge estimator for LiFePO₄ batteries leveraging expansion force. *J. Energy Storage* **2022**, *50*, 104559. <https://doi.org/10.1016/j.est.2022.104559>.
42. Wu, L.; Wei, X.; Lin, C.; Huang, Z.; Fan, Y.; Liu, C.; Fang, S. Battery SOC estimation with physics-constrained BiLSTM under different external pressures and temperatures. *J. Energy Storage* **2025**, *117*, 116205.
43. Li, Y.; Ye, M.; Wang, Q.; Lian, G.; Xia, B. An improved model combining machine learning and Kalman filtering architecture for state of charge estimation of lithium-ion batteries. *Green Energy Intell. Transp.* **2024**, *3*, 100163.
44. Tian, Y.; Lai, R.; Li, X.; Xiang, L.; Tian, J. A combined method for state-of-charge estimation for lithium-ion batteries using a long short-term memory network and an adaptive cubature Kalman filter. *Appl. Energy* **2020**, *265*, 114789.
45. Jiang, H.; Yin, L.; Xu, Z.; Hu, L.; Huang, W.; Zhao, Y. A novel hybrid framework for SOC estimation using PatchMixer-LSTM and adaptive UKF. *Energy* **2025**, *335*, 137891.
46. Hong, J.; Liang, F.; Yang, H.; Zhang, C.; Zhang, X.; Zhang, H.; Wang, W.; Li, K.; Yang, J. Multi-forward-step state of charge prediction for real-world electric vehicles battery systems using a novel LSTM-GRU hybrid neural network. *eTransportation* **2024**, *20*, 100322. <https://doi.org/10.1016/j.etrans.2024.100322>.
47. Ren, X.; Liu, S.; Yu, X.; Dong, X. A method for state-of-charge estimation of lithium-ion batteries based on PSO-LSTM. *Energy* **2021**, *234*, 121236.
48. Tang, X.; Wang, Y.; Chen, Z. A method for state-of-charge estimation of LiFePO₄ batteries based on a dual-circuit state observer. *J. Power Sources* **2015**, *296*, 23–29. <https://doi.org/10.1016/j.jpowsour.2015.07.028>.
49. Lin, C.; Mu, H.; Xiong, R.; Shen, W. A novel multi-model probability battery state of charge estimation approach for electric vehicles using H-infinity algorithm. *Appl. Energy* **2016**, *166*, 76–83. <https://doi.org/10.1016/j.apenergy.2016.01.010>.
50. Li, P.; Ao, Z.; Hou, J.; Xiang, S.; Wang, Z. Physics-informed mamba neural network with potential knowledge for state-of-charge estimation of lithium-ion batteries. *J. Energy Storage* **2025**, *123*, 116546.
51. Yang, F.; Zhang, S.; Li, W.; Miao, Q. State-of-charge estimation of lithium-ion batteries using LSTM and UKF. *Energy* **2020**, *201*, 117664.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.