Energy Reports 16 (2026) 109544

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Adaptive battery SOC estimation: Temporal attention-enhanced BiLSTM and quantized model comparison for scalable edge solutions ☆

Opy Das [a,b], Muhammad Hamza Zafar [a], Souman Rudra [a,*], Filippo Sanfilippo [a,c]

[a] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[b] Department of Computer Science (IDI), Norwegian University of Science and Technology, Trondheim, 7033, Norway
[c] Department of Software Engineering, Kaunas University of Technology, Kaunas, 51368, Lithuania

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/cp3473x7xv/3

Keywords:
State of charge
Neural networks
Electric vehicle
Deep learning
Quantisation
Temporal attention mechanism

ABSTRACT

Lithium-ion batteries are the dominant energy storage technology in modern energy systems, making the accurate state-of-charge (SOC) estimation essential for effective battery management. Reliable SOC estimation helps determine remaining battery capacity, improves operational safety, and supports real-time energy control. However, SOC estimation is challenging due to the nonlinear electrochemical behavior of lithium-ion batteries and performance degradation caused by aging and temperature variations. This study proposes a Bidirectional Long Short-Term Memory network with a Temporal Attention Mechanism (BiLSTM-TAM) to enhance SOC estimation accuracy across a wide temperature range from −20 °C to 40 °C. The model is evaluated under several driving conditions, including the Urban Dynamometer Driving Schedule (UDDS), the US06 aggressive driving cycle, a mixed driving cycle that combines multiple standard cycles, and the Hybrid Pulse Power Characterization (HPPC) test commonly used for battery modeling. Model performance is assessed using root-mean-square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). Results demonstrate that the proposed BiLSTM-TAM model improves SOC estimation accuracy and robustness. To enable practical deployment, post-training quantization (PTQ) is applied to reduce model size and computational requirements while maintaining predictive performance, making it suitable for edge-based electric vehicle applications.

1. Introduction

The global energy landscape is undergoing a rapid transformation, with the integration of renewable energy sources and advancements in power generation technologies. This transition has opened new opportunities in the transportation sector, particularly with the rise of Electric Vehicles (EVs) as an alternative to conventional fossil fuel-based vehicles. EVs use cleaner, more sustainable energy sources, accelerating the transformation towards a greener, safer, and more resilient energy future. In the context of EVs, lithium-ion (Li-ion) batteries have emerged as the dominant energy storage solution (Zubi et al., 2018; Chen et al., 2019). A critical aspect of Li-ion battery management is the accurate estimation of the battery's State of Charge (SOC). SOC is a key parameter that reflects the ratio of the remaining capacity to the battery's rated capacity. Monitoring and maintaining SOC within safe operating limits is vital for the efficient functioning of the battery management system (BMS), which is responsible for ensuring the reliability and safety of the battery (Liu et al., 2022; Zahid et al., 2018).

Accurate SOC estimation is essential not only for determining the vehicle's range and planning trips but also for optimizing the BMS to prolong battery life and maintain operational safety. However, due to the complex electrochemical processes and dynamic behavior of batteries, SOC estimation poses significant challenges (Shah et al., 2022). As batteries age and their capacity declines, accurately estimating the SOC becomes increasingly difficult, presenting a significant challenge for reliable battery management. Inaccuracies may arise from measurement errors originating from the BMS or associated equipment. Furthermore, it is crucial to consider coupling effects, such as the impact of temperature on internal resistance and capacity, to ensure precise SOC estimation. The nonlinear charging and discharging curves with small quasi-linear regions are one of the main barriers to estimating the SOC properly (Sulaiman et al., 2015).

Lithium-ion batteries are set to become a predominant energy storage solution, with their global market share anticipated to rise substantially as electric vehicles substitute conventional fuel-powered transportation. Many countries are currently dedicated to transitioning their

☆ This article is part of a Special issue entitled: 'VSI: AI-enabled Power Systems(Haixin Wang)' published in Energy Reports.

* Corresponding author.
E-mail address: souman.rudra@uia.no (S. Rudra).

https://doi.org/10.1016/j.egyr.2026.109544
Received 9 March 2026; Received in revised form 12 May 2026; Accepted 5 July 2026
Available online 10 July 2026

transportation systems entirely to electric automobiles. Precise State of Charge estimation is essential for lithium-ion batteries as well as other battery varieties. Multiple methodologies for SOC estimation are available, including the Electrochemical Model, Equivalent Circuit Model, and Electrochemical Impedance Model (Marelli and Corno, 2020; Lai et al., 2018; Amir et al., 2022). These methodologies replicate battery behavior by considering numerous parameters affecting the performance of lithium-ion batteries. However, considering the system's inherent nonlinearity, accurately modeling a battery by including all relevant elements is challenging.

The techniques available for SOC estimation include the Voltage Method, Coulomb Counting Method, Impedance Spectroscopy, and the Kalman Filter Method. SOC estimation depends on the continuous monitoring of the battery's voltage, current, and temperature. The Voltage Method utilizes the discharge profile (Voltage vs State of Charge) to convert battery voltage data into their respective SOC values. The correlation between a battery's voltage and its SOC is significantly nonlinear, further complicated by hysteresis effects. The voltage response during charging is also distinct from that during discharging. Multiple factors, including temperature, battery degradation, and discharge rates, contribute to this nonlinearity (Qi et al., 2022). The Coulomb counting method determines the remaining capacity by aggregating the charge entering and exiting the battery. The effectiveness of this method is heavily dependent on accurate current measurement and proper initialization of the SOC. As SOC estimation depends on indirect measurements and modeling, it may add uncertainties and inaccuracies. Over time, SOC drift develops as an impact of accumulated inaccuracies in current and voltage measurements, along with errors in capacity estimation that are worsened by hysteresis. An incorrect estimation of the initial capacity can lead to significant inaccuracies in evaluating battery performance during regular operation. Impedance spectroscopy, however beneficial, is restricted to new batteries with known capacities, and Kalman filtering necessitates a rigorously validated model and precise parameter identification to get dependable SOC estimation (Zhou et al., 2022).

The difficulty of SOC estimation increases significantly in everyday situations, especially when temperatures are quite high or low. Battery behavior is greatly influenced by temperature variations, which in turn affect chemical processes, self-discharge rates, and internal resistance. At low temperatures, battery capacity decreases, while at higher temperatures, it increases; however, colder conditions can accelerate the battery's long-term degradation (Xu et al., 2021). To tackle these highly non-linear relationships, a growing number of applications are currently utilizing neural networks to consider temperature fluctuations across different weather conditions, including reduced capacity in lower temperatures and the potential for enhanced capacity in warmer situations. When compared to more conventional model-based approaches, neural networks produce more precise SOC estimations because of their remarkable capacity to capture the intricate, nonlinear connections at work in battery dynamics. Neural networks are rapidly replacing traditional methods as the benchmark for accurately simulating battery behavior in a wide range of environments (Du et al., 2021).

With the growing use of batteries in a variety of applications, including electric vehicles, mobile robots for disaster relief, and other small devices, accurate SOC estimation has become essential. To do this, deep learning approaches have been used; however, implementing these complicated models on edge devices offers substantial problems due to their limited computational capacity and memory. Heavy neural networks require strong computation, making them challenging to implement efficiently in environments with limited resources. To solve this issue, approaches like quantization and pruning have been developed to reduce model size and computing needs while retaining performance. Quantization, in particular, has shown to be a key strategy for optimizing deep neural networks, allowing them to be deployed on edge devices while maintaining accuracy and efficiency. Post-training quantization has demonstrated the ability to reduce model size and computational demands while preserving accuracy. This process often reduces the precision of both weights and activations, typically converting them from 32-bit floating point to 16-bit floating point or 8-bit integers, making it especially well-suited for resource-constrained applications.

1.1. Literature review

Accurate estimation of the State of Charge is essential for effective battery management, and model-based approaches provide a robust framework for achieving this using a variety of established techniques (How et al., 2019; Cacciato et al., 2016). The Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Particle Filter (PF) (Yang et al., 2016), and Extended Kalman Particle Filter (EPF) (Tian et al., 2022) are among the most frequently utilized methods. Studies have also delved into equivalent circuit models (ECMs), which utilize resistance-capacitance networks (Hu et al., 2012) to estimate SOC in various types of Li-ion batteries. Furthermore, models that rely on temperature have been created to calculate SOC in the presence of temperature fluctuations (Xing et al., 2014), and models that employ partial differential equations (PDEs) (Zou et al., 2015) have been implemented to encompass more intricate battery dynamics. Although these models yield quantitatively precise outcomes in certain situations like under stable temperatures or for specific battery chemistries they all possess inherent limitations. The advantages and drawbacks of these model-based approaches have been thoroughly documented in Xiong et al. (2018). Currently, there is a growing emphasis on creating more sophisticated models that consider the complexities of lithium-ion battery performance across various environmental conditions, with the goal of enhancing precision in practical applications.

Lookup table-based methods, which include the voltage method and Coulomb counting, are frequently utilized to estimate SOC. The relationship between Open Circuit Voltage (OCV) and SOC is influenced by varying temperatures, especially in series-parallel configurations of lithium-ion batteries (Shen et al., 2017). To obtain a precise OCV-SOC curve, it is essential to perform OCV tests over multiple days while adhering to particular environmental conditions (Gong et al., 2021). This method's main drawback is that it requires the battery to be in a rest state in order to calculate the SOC, which makes it impractical for real-time applications such as those used in the operation of electric vehicles (EVs) (Lee et al., 2007). The Coulomb counting method is another commonly employed technique that estimates the SOC by integrating the current that flows in or out of the battery throughout the charge and discharge cycles (Zine et al., 2022). But there are also some challenges associated with this method, like determining the initial SOC of the battery. A typical approach to determine an accurate initial state is to either completely charge the cell to its maximum voltage or discharge it to its cut-off voltage, thus establishing a known reference point. However, internal battery losses result in the releasable charge being consistently lower than the total stored charge throughout both the charging and discharging processes. Accurate SOC estimation necessitates the consideration of these losses when utilizing the Coulomb counting method (Qi et al., 2022; Dai et al., 2019).

Traditional state estimation techniques generally depend solely on external parameters such as voltage, current, and temperature. However, sensor flaws may lead to more errors, and at reduced temperatures, the battery's capacity may increase due to the decelerated rate of chemical processes (Cui et al., 2022a). To address these issues, the utilization of traditional machine learning techniques alongside neural networks to improve battery state estimation has received considerable attention in recent years (Shah et al., 2022; Cui et al., 2022a; Manoharan et al., 2022; Dos Reis et al., 2021; Tong et al., 2016). Conventional machine learning methodologies, such as Support Vector Machines (SVMs), Gaussian Process Regression (GPR), and Artificial

Neural Networks (ANNs), have frequently been used for assessing the SOC (Li et al., 2007; Anton et al., 2013; Chang, 2013; Deng et al., 2020). But these methods frequently need manual feature engineering and are constrained by their shallow learning architectures, which reduces their effectiveness when applied to bigger datasets. Neural networks demonstrate enhanced proficiency in autonomously acquiring intricate behaviors, hence reducing the necessity for feature engineering. The accuracy of State of Charge estimation across diverse machine learning algorithms can be improved by incorporating sophisticated methods such as Kalman filters and fuzzy logic systems (Cui et al., 2022b).

In Korkmaz (2023), 18 different machine learning algorithms and five distinct filters were grouped into three categories to estimate the SOC. The neural network approach outperformed the others in terms of accuracy and redundancy. The attention-based long short-term memory (LSTM) network estimated SOC with an error rate of less than 1.41 In Yang et al. (2019), the authors employed recurrent neural networks (RNNs) with a gated recurrent unit (GRU-RNN) to estimate the SOC for nickel manganese cobalt oxide (NMC) and lithium iron phosphate (LFP) batteries, being the first application of this methodology. Another GRU-RNN model was proposed in Jiao et al. (2020), integrating a momentum optimization algorithm to enhance SOC estimation. In addition, a separate study highlighted the significance of Long Short-Term Memory (LSTM) models in effectively capturing temporal dependencies in the forward direction (Hochreiter, 1997). A bidirectional LSTM (Bi-LSTM) model, an extension of LSTM, was introduced in Schuster and Paliwal (1997), demonstrating its capacity to capture long-term dependencies in both the forward direction, utilizing past input features, and the backward direction, incorporating future input features. In Zafar et al. (2024), the authors introduced a Fusion-Fission Optimization (FuFi) based Convolutional Neural Network combined with a Bi-Directional Long Short-Term Memory Network (FuFi-CNN-Bi-LSTM) to estimate SOC, including scenarios where temperatures drop below zero degrees. Another effective strategy is transfer learning, which uses the parameters learned from a pre-trained deep neural network (DNN) to train further DNNs. In Vidal et al. (2019), the authors demonstrated that applying transfer learning resulted in a 64.

Amid these advancements, recent research has shifted focus toward optimizing computational efficiency and storage requirements without compromising accuracy. Recent research indicates that quantisation can be employed across various neural network architectures, including BiLSTM, CNN, and LSTM models with attention mechanisms, to significantly reduce their computational and storage requirements while maintaining suitable performance standards. In Jacob et al. (2018), the authors demonstrated the application of post-training quantisation to CNNs, showing that this method can speed up inference while preserving high accuracy. The quantisation technique was effective in reducing model size, making it highly suitable for use in mobile and embedded systems. In Lin et al. (2017), the authors investigated the use of quantisation for BiLSTM models with attention mechanisms, highlighting its potential in maintaining strong performance while optimizing the models for efficiency in various use cases.

1.2. Contributions and paper organization

Estimating the SOC in lithium-ion batteries is a significant challenge in battery management systems. This difficulty arises from the complex and dynamic behavior of batteries under different operating conditions. The complexity is increased by variations in drive cycles of EVs and temperature changes, which can greatly affect the accuracy of SOC predictions. Current methodologies frequently encounter challenges in achieving a balance between adaptability and precision in diverse scenarios. The motivation for this research stems from the need to address these challenges while enabling practical deployment of SOC estimation models in resource-constrained environments. With the growing adoption of edge devices such as mobile robots, smartwatches, and lightweight embedded systems, traditional large-scale models become impractical due to their high computational and storage demands. This study aims to bridge this gap by integrating quantization techniques to reduce model size and complexity, making it feasible to deploy on low-power devices without substantially compromising performance. Additionally, this work examines SOC estimation across a wide range of temperatures, including extreme conditions, ensuring the robustness of the proposed models in real-world applications. The graphical abstract of the proposed model is outlined in Fig. 1. This work presents the following key contributions:

- A key novelty of this work lies in the introduction of a Temporal Attention-based BiLSTM network, specifically designed to evaluate its performance across ambient, low, high, and extreme temperature conditions. This approach aims to enhance the model's generalization capabilities in diverse scenarios. The temporal attention mechanism employed in this study stands out for its ability to robustly capture dependencies and relationships across time steps in sequential or time-series data, providing a detailed understanding of temporal dynamics essential for accurate predictions.
- Development of a flexible model pipeline that leverages a pre-trained framework for deep neural network (DNN) models, enabling seamless integration of any drive cycle to make accurate predictions without requiring further retraining. The framework is designed to automatically extract key features from drive cycles, ensuring efficient processing and precise predictions. This approach enhances the model's reusability, which significantly simplifies the SOC prediction process.
- This study presents a comprehensive comparison of four deep learning models: LSTM, CNN-LSTM, CNN-BiLSTM, and BiLSTM with Temporal Attention Mechanism (BiLSTM-TAM), evaluated across five different temperature conditions.
- A post-training quantisation technique has been employed to significantly compress the model, reducing its size and computational complexity. This study presents a comprehensive evaluation of different DNN model performance, comparing results with and without quantisation and analyzing the impact of size reduction and complexity on accuracy.
- A data processing technique has been adapted to standardize EV sample rates by using a time-rounding method, ensuring a consistent and reliable dataset.

The paper is structured as follows: Section 2 offers an overview of the datasets and drive cycles utilized in this study. Section 3 outlines the proposed methodology, including the quantisation techniques employed. Section 4 presents the results of the proposed method, along with a comparative analysis against other approaches. Moreover, this section includes a comparison of the results with the quantised model. Finally, Section 5 discusses the findings derived from the proposed technique and highlights the key conclusions.

2. Dataset description

There are numerous publicly available lithium-ion battery datasets that have been generated using various battery types, based on different cell chemistries and testing conditions. For this study, the drive cycle datasets were sourced from McMaster University (Korkmaz, 2023), where a 3 Ah LG HG2 cell was tested. Detailed specifications of this battery are presented in Table 1. The batteries were subjected to power profiles simulating an electric vehicle undergoing standard industry drive cycles, including US06, UDDS, and LA92. Additionally, Hybrid Pulse Power Characterization (HPPC) tests were conducted on the battery, which are primarily used for acquiring parameters of the battery. The dataset also contains data from eight mixed drive cycles with corresponding charge cycles. Testing was conducted at six different ambient temperatures: 40 °C, 25 °C, 10 °C, 0 °C, −10 °C, and −20 °C. Details of the tested cells are provided in the table.

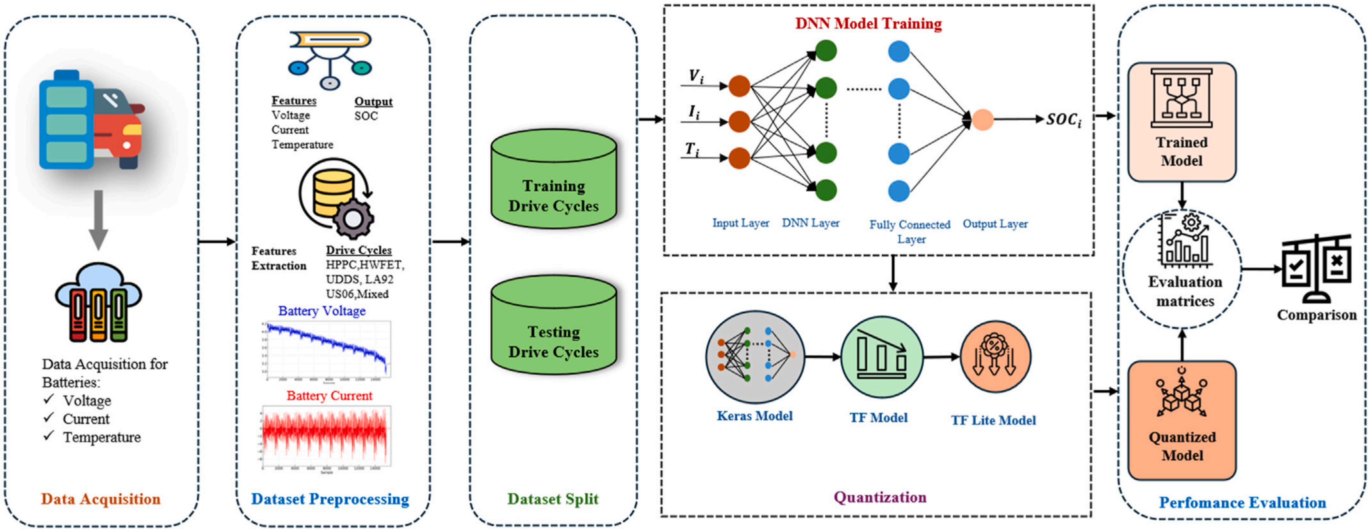


Fig. 1. Graphical abstract of SOC estimation using Deep Learning Model with quantisation.

Table 1
Specifications of battery dataset.

Parameter	Values
Type	LG NMC/18650HG2
Nominal Capacity	3.0 Ah
Nominal Voltage	3.6 V
Upper Voltage Limit	4.2 V
Cut-off Voltage	2.5 V
Maximum Current	20A
Temperatures	−20 °C, −10 °C, 0 °C, 10 °C, 25 °C, 40 °C
Drive Cycles	HPPC, HWFET, UDDS, US06, LA92, PausCycl

In this study, the performance of the proposed BiLSTM-TAM model is evaluated using the McMaster dataset. The original dataset includes both constant charge and discharge cycles, as well as HPPC tests, industry-standard drive cycles, and mixed drive cycles. The constant charge and discharge cycles were conducted at a sampling rate of one sample per minute, while the drive cycles were recorded at a frequency of 10 samples per minute. The constant charge and discharge cycles were excluded from this analysis as they fail to capture the complex dynamics required for accurate predictions. Furthermore, data recorded at −10 °C was omitted due to insufficient temperature variation.

The input features for the model consisted of current, voltage, and temperature at specific time points, with the output being the corresponding SOC. The voltage ranged from 2.8 V to 4.2 V, while the current varied between −20 A and 5 A. Figs. 2 and 3 present the voltage, current, and temperature data for the HPPC, UDDS, US06 and mixed cycles at 25 °C which were used as the testing dataset in this study. This data provides a comprehensive view of the battery's behavior under different drive cycles and was instrumental in developing and training the models for accurate SOC estimation.

The dataset was divided into two parts: a training dataset and a testing dataset. The evaluation was conducted across four different deep learning models. Following this, each deep learning model was subjected to post-training quantisation, and the drive cycles were executed again on the Quantised models to evaluate their performance after quantisation.

3. Methodology

In recent years, deep neural networks (DNNs) have gained significant popularity in battery state estimation due to their ability to handle large datasets, leverage increased computational power, and utilize more advanced neural network architectures. Recurrent neural networks (RNNs), in particular, have proven highly effective for time-series data, making them ideal for estimating SOC due to the time-dependent nature of battery dynamics, such as accumulated charging current, discharging current, terminal voltage, and temperature variations.

In this study, four different RNN models, including our proposed BiLSTM-TAM (Bidirectional Long Short-Term Memory with Temporal Attention Mechanism) model, were used to estimate SOC. The process included several critical steps:

- **Data Acquisition and Preprocessing:**
  The original McMaster Dataset includes a combination of charge cycles, discharge cycles, and EV drive cycles, which consist of constant charging, constant discharging, and dynamic current profiles. For this study, the focus is on preprocessing the drive cycles, as the primary objective is SOC prediction during drive cycles, reflecting real-life applications.
  To preprocess the data, downsampling was performed to adjust the data to a 1 Hz frequency. This was achieved using a time-rounding method to align all measurements at uniform 1-second intervals. The initial SOC was determined based on the maximum capacity, calculated using the continuous charging and discharging cycle at a C/24 current by finding the maximum discharge capacity. The SOC at each time step i was then computed using the Coulomb counting method stated by the following equation.

$$\mathrm{SOC}(i) = \mathrm{SOC}(i-1) + \frac{I(i)}{Q_n}\Delta i \tag{1}$$

  The estimated state of charge at time i is given by $\mathrm{SOC}(i)$, and the previous state of charge at time $(i-1)$ is represented as $\mathrm{SOC}(i-1)$. The charging or discharging current at time i is denoted by $I(i)$, with Q_n representing the battery's nominal capacity at that stage. Finally, Δi refers to the time step.
  For each drive cycle, current, voltage, temperature, and SOC data were collected, and these values were normalized to ensure consistency and compatibility across the dataset. Normalization helps deep learning algorithms manage the wide range of variable values effectively, improving the model's stability and accuracy during training. The SOC was used as the output variable. Since SOC values are inherently within the range of 0 to 1, no additional normalization was applied for the target SOC value. This preprocessing ensured that the model received well-structured and noise-reduced input data, thereby enhancing its prediction accuracy and overall performance.

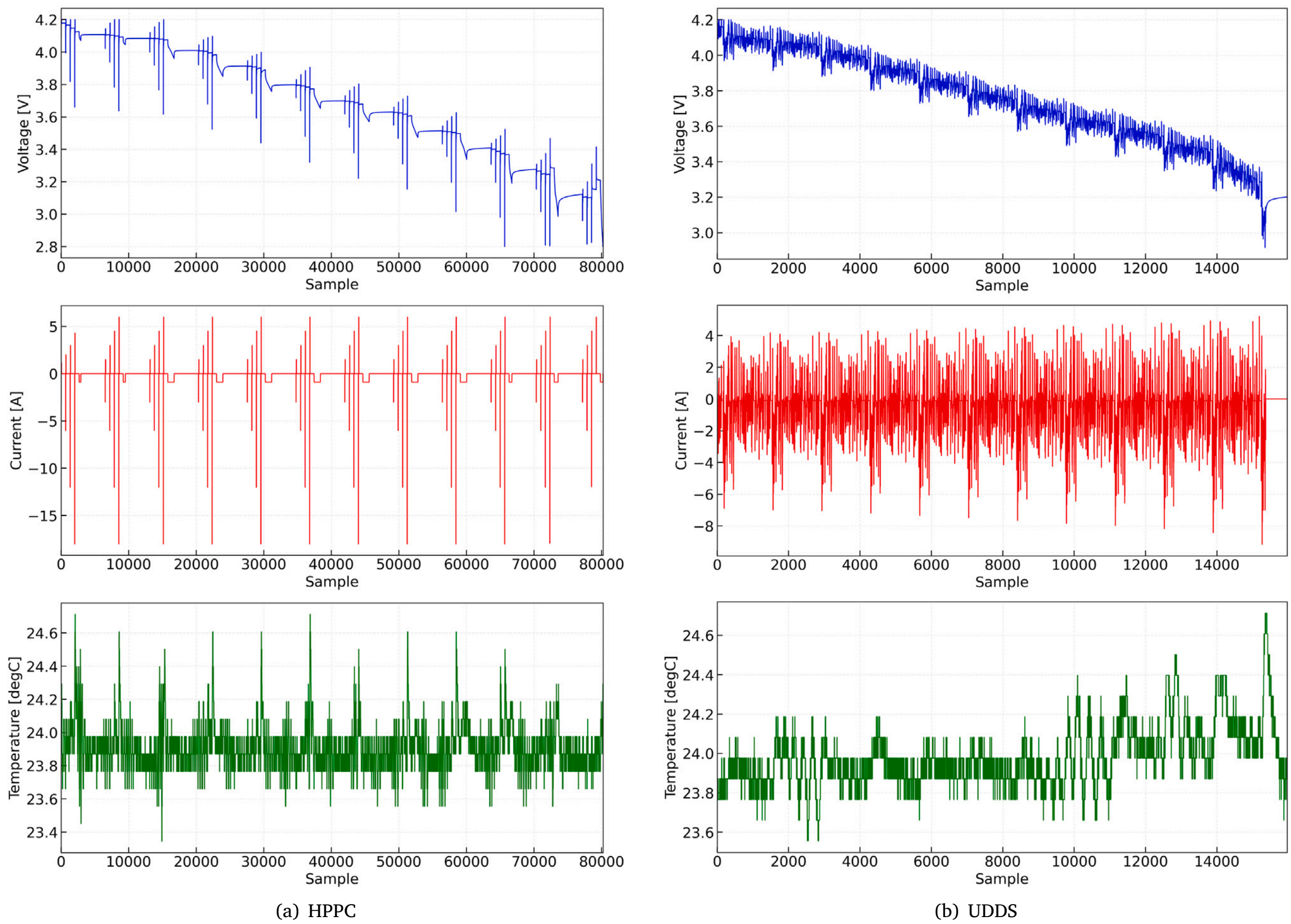


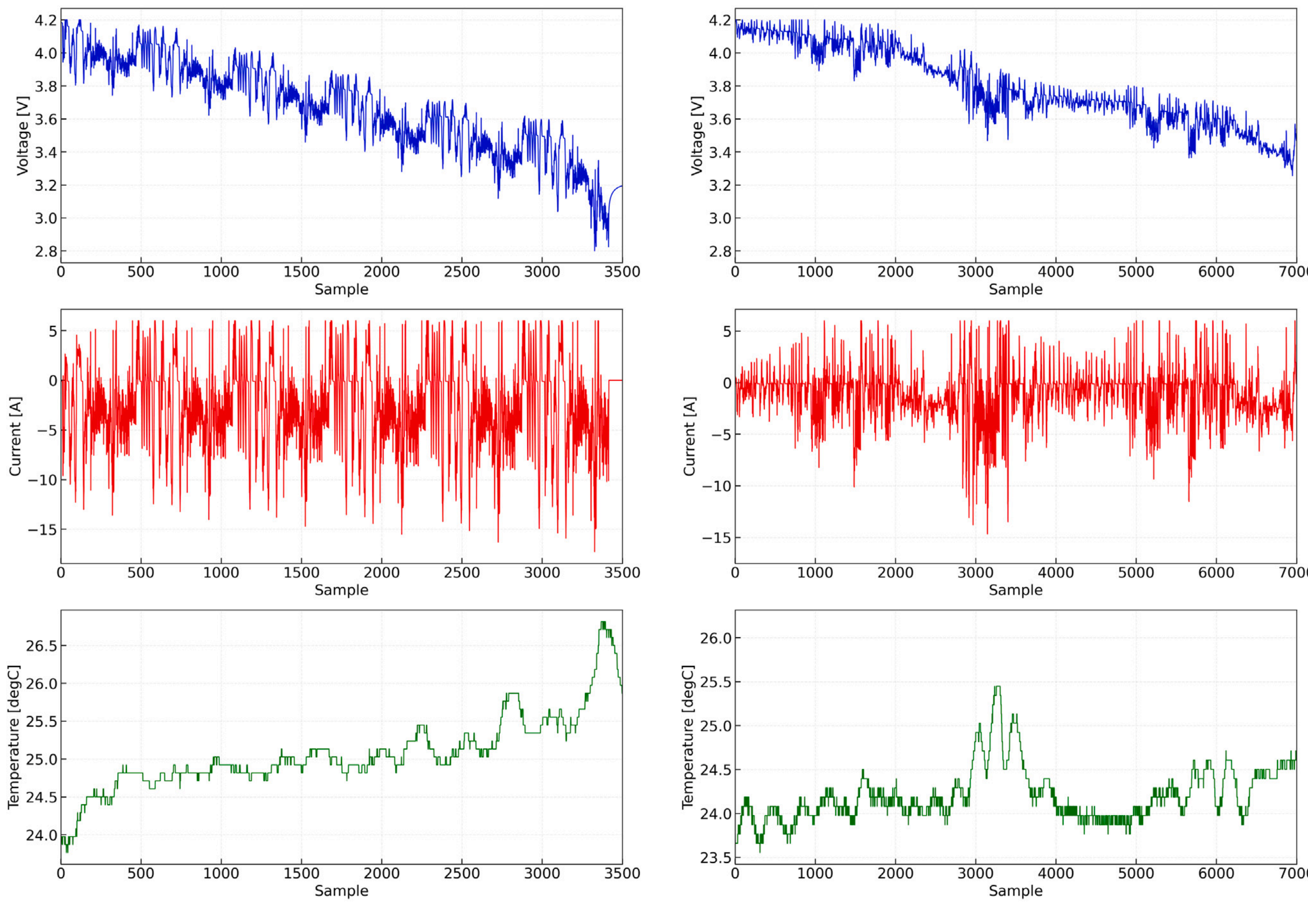
Fig. 2. Voltage, current, and temperature variations during the (a)HPPC cycle,(b) UDDS cycle at 25 °C.

- **Dataset Splitting:** The dataset comprises multiple drive cycles collected under different temperature conditions. To ensure a fair and consistent evaluation, four drive cycles (HPPC, UDDS, US06, and one mixed drive cycle) were consistently reserved as the test set across all experiments, corresponding approximately to a 30% test split. The remaining data, approximately 70%, was used for model development. Within the training portion, a validation subset of 10% was always maintained to monitor training performance and prevent overfitting. Although minor variations in the training data, such as sequence selection differences, were introduced, the model performance remained consistent across these settings. While the exact composition of the training sequences varied slightly across experiments due to differences in sequence selection and preprocessing, the overall proportion remained consistent.
- **Model Training:**
  The models, including LSTM, CNN-LSTM, CNN-BiLSTM, and BiLSTM-TAM, were trained using the normalized dataset.
- **Quantisation:**
  After training, the Keras models were subjected to post-training quantisation to reduce model size and computational complexity. This step involved converting the weights and activations of the trained models from 32-bit floating-point precision to lower bit representations, such as 8-bit integers, without significantly sacrificing accuracy. This Quantised model is particularly suited for deployment on edge devices with limited computational resources.
- **Performance Evaluation:**
  In this study, 8-bit integer quantisation was chosen over 16-bit floating point quantisation to prioritize computational efficiency and reduce model size, aligning with the constraints of resource-limited edge devices. The precision offered by 8-bit quantisation proved sufficient for SOC prediction tasks, as demonstrated in the results section, where comparable performance metrics were achieved during testing. Furthermore, 8-bit integer quantisation is widely recognized as an industry-standard technique for efficient deployment, delivering faster inference and reduced memory usage without significant accuracy loss.

The detailed steps of the entire process in this study are illustrated graphically in Fig. 1.

3.1. BilSTM

A Bidirectional LSTM (BiLSTM) is a sequence-processing model that utilizes two LSTM networks: one to process input in the forward direction and the other to process it in the backward direction. This bidirectional approach allows the network to access more information by considering both past and future context, thereby enhancing the model's ability to understand and learn from data. This is especially useful in scenarios where the current output depends on future context, such as time series prediction or language modeling. The BiLSTM design enables the model to capture both short- and long-term dependencies in the data, making it highly effective at making



(a) US06 (b) Mixed

Fig. 3. Voltage, current, and temperature variations during the (a)US06 cycle,(b) Mixed cycle at 25 °C.

accurate predictions even when working with incomplete or noisy data (Siami-Namini et al., 2019).

BiLSTMs are particularly well-suited for time series regression tasks, such as stock price prediction, weather forecasting, and medical diagnosis, as they incorporate information from both past and future time steps. A general architecture of the BiLSTM model is illustrated in Fig. 4. While the mathematical equations governing the BiLSTM are similar to those of a traditional LSTM, BiLSTM introduces two distinct sets of equations: one for the forward direction and another for the backward direction (Sharfuddin et al., 2018; Aslan et al., 2021).

The mathematical equations describing the BiLSTM network are as follows. Let X_t represent the input at time step t, C_{t-1} the previous cell state at time step $t-1$, C_t^f and C_t^b the cell states at time step t in the forward and backward directions, respectively, and h_t^f and h_t^b the hidden states in the forward and backward movements at time step t. The activation functions σ and tanh represent the sigmoid and hyperbolic tangent functions, respectively. The mathematical equations describing the LSTM network are as follows.

Forward Movement:

$$f_t^f = \sigma\left(W_f^f[h_{t-1}, X_t] + b_f^f\right), \tag{2}$$

$$i_t^f = \sigma\left(W_i^f[h_{t-1}, X_t] + b_i^f\right), \tag{3}$$

$$C_t^f = \tanh\left(W_c^f[h_{t-1}, X_t] + b_c^f\right), \tag{4}$$

$$o_t^f = \sigma\left(W_o^f[h_{t-1}, X_t] + b_o^f\right) \tag{5}$$

$$h_t^f = o_t^f * \tanh(C_t^f). \tag{6}$$

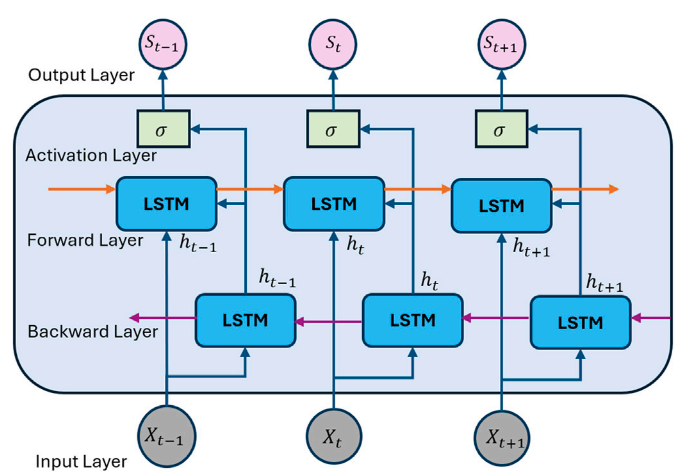


Fig. 4. Structure of a BiLSTM cell.

Backward Movement:

$$f_t^b = \sigma\left(W_f^b[h_{t+1}, X_t] + b_f^b\right), \tag{7}$$

$$i_t^b = \sigma\left(W_i^b[h_{t+1}, X_t] + b_i^b\right), \tag{8}$$

$$C_t^b = \tanh\left(W_c^b[h_{t+1}, X_t] + b_c^b\right), \tag{9}$$

$$C_t^b = f_t^b C_{t-1}^b + i_t^b C_t^b + 1^b, \tag{10}$$

$$o_t^b = \sigma\left(W_o^b[h_{t-1}, X_t] + b_o^b\right), \tag{11}$$

$$h_t^b = o_t^b * \tanh(C_t^b). \tag{12}$$

The input, forget, and output gates are denoted as i_t, f_t, and o_t, respectively, whereas C_t' represents the candidate cell state. Training involves learning the weights (W_i, W_f, W_o, and W_c) and biases (b_i, b_f, b_o, and b_c). The function $\tanh(C_t)$ can be any non-linear activation function, which includes the hyperbolic tangent or the Rectified Linear Unit (ReLU).

3.2. Temporal attention

Attention mechanism has recently demonstrated success in a wide range of tasks like Natural Language Processing, sentiment analysis, image processing and time series forecasting (Niu et al., 2021; Li et al., 2023; Chrysostomou and Aletras, 2021). It is implemented by constructing a neural network according to the corresponding tasks. The neural network that is used for attention mechanism is usually called attentive neural network.

The temporal attention mechanism is a specific variant of the attention mechanism designed for time-series or sequential data. Temporal attention focuses on selecting or weighing specific time steps (or moments in the sequence) as more important, considering the temporal dependencies in the data. This is particularly important when dealing with tasks like SOC estimation in EVs, where historical data is crucial to understanding the current state. Temporal attention helps the model learn which moments in the sequence should be prioritized to improve the prediction for the current or future states, such as detecting relevant patterns at certain time points in time-series data (Tan et al., 2023).

In a sequential, time-ordered manner, temporal attention is particularly effective for estimating battery charge levels, as it prioritizes the influence of recent events over older ones. This can be conceptualized as acknowledging that the battery's state five minutes ago is more important than its state five hours ago. Temporal attention facilitates efficient analysis by processing data sequentially, similar to reading a story chronologically rather than skipping between pages. This attribute provides it particularly advantageous for monitoring battery behavior over extended periods (Rosin and Radinsky, 2022).

Compared to other attention mechanisms, such as self-attention or multi-head attention, temporal attention is easier to understand and trust. It provides a clear, traceable path that highlights which past battery readings are most important for predicting the current charge level. Furthermore, it uses fewer computationally intensive operations, which means it is less likely to get confused by messy battery data. This simple but effective approach works really well for batteries because their current state heavily depends on recent activities—like how much you have used the battery or charged it in the past few hours.

The main mathematical formulas used by temporal-attention mechanism designed to process tailored for time-series or sequential data, focusing on selecting key time steps for improved prediction of future or current states are as follows.

$$h_t = [h_t^f; h_t^b], \tag{13}$$

$$e_t = \tanh\left(W_a h_t + b_a\right), \tag{14}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^{T} \exp(e_t)}, \tag{15}$$

$$c = \sum_{t=1}^{T} \alpha_t h_t. \tag{16}$$

The hidden states from both the forward (h_t^f) and backward (h_t^b) LSTM layers are combined into a single hidden state h_t for each time step. An attention score e_t is then calculated by passing h_t through a dense layer with weights W_a and bias b_a, followed by the tanh activation function. These attention scores are normalized using the softmax function to generate attention weights α_t, ensuring they sum to 1 across all time steps. Finally, a context vector c is computed by taking the weighted sum of the hidden states, with each hidden state h_t multiplied by its corresponding attention weight α_t. This context vector emphasizes the most important time steps and is used for the final prediction. The structure of temporal attention mechanism is presented in Fig. 5 to estimate the SOC at each time step.

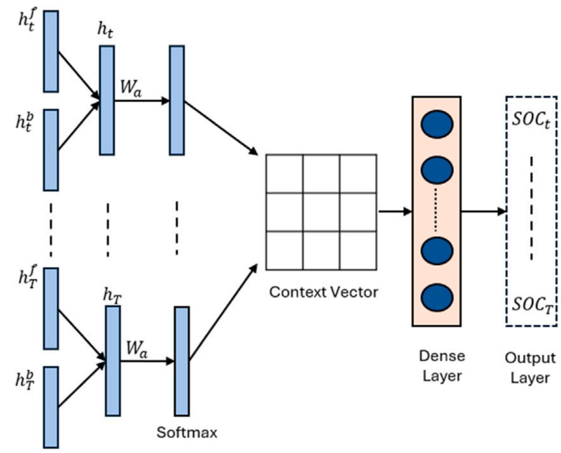


Fig. 5. Structure of Temporal Attention Mechanism to estimate the SOC at each time step.

3.3. BiLSTM-TAM

The proposed approach using BiLSTM helps the model capture both historical and future dependencies in battery data, enhancing the context for SOC estimation. Additionally, the Temporal Attention Mechanism enables the model to concentrate on the most important time steps, effectively filtering out irrelevant data and minimizing noise. The input layer is defined as a sequence, where each time step consists of multiple features. The model is built with two BiLSTM layers, which process the input sequence in both forward and backward directions to capture temporal dependencies. Dropout is applied after each BiLSTM layer to mitigate overfitting. Following the second BiLSTM layer, an attention mechanism is employed to focus on the most critical time steps in the sequence. A dense layer with tanh activation computes attention scores for each timestep and the scores are normalized using a softmax layer to create weights (attention scores). The attention scores are applied to the LSTM outputs, effectively focusing on the most important timesteps for SOC prediction. The result is a weighted version of the LSTM output, emphasizing significant temporal features. Finally, the custom TemporalSum layer aggregates the attention-weighted LSTM outputs across the time dimension, providing a summarized representation for the final prediction. The result is a single vector representing the most relevant features for SOC prediction, with the time dimension removed. A dense layer with a single unit is employed as the output layer, which is appropriate for regression tasks which predicts the continuous value estimating SOC, based on the input sequence. The complete flow diagram of the proposed method is presented in Fig. 6.

Additionally, the deep learning methods were executed on a computer running the Windows operating system, equipped with an i9-11900H processor, an NVIDIA RTX 3060 graphics card, 32 GB of RAM, and a 1TB SSD.

3.4. Quantisation

Quantisation is the process of converting a machine learning model into an equivalent representation that utilizes lower precision for parameters and computations. This transformation enhances the model's performance and efficiency. For instance, TensorFlow Lite's 8-bit integer quantisation can result in models that are up to four times smaller, 1.5 times to 4 times faster in computation, and more power-efficient on CPUs. Typically, there are two types of quantisation: Post-Training

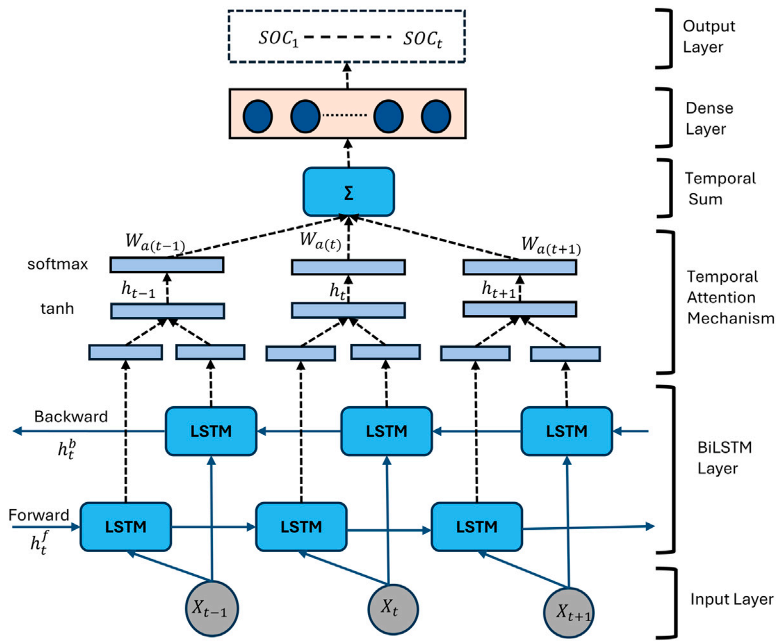


Fig. 6. Structure of the Proposed BiLSTM temporal attention mechanism based Algorithm.

Quantisation (PTQ) and Quantisation Aware Training (QAT) (Paupamah et al., 2020; Wang et al., 2021). In this study, we focus on Post-Training Quantisation . PTQ encompasses techniques designed to reduce CPU and hardware accelerator latency, processing time, power consumption, and model size with minimal impact on accuracy. These techniques can be applied to pre-trained float TensorFlow models during the TensorFlow Lite conversion process. Specifically, we have reduced the weights from 32-bit to 8-bit precision.

After the models were trained, they were Quantised and evaluated on test drive cycles. The performance of the Quantised models was compared to the original Keras models using various evaluation metrics.

3.5. Evaluation metrices

MAE (Mean Absolute Error) measures the average magnitude of prediction errors, without considering their direction. Lower MAE values reflect greater accuracy. MSE (Mean Squared Error) calculates the mean of squared errors, placing more emphasis on larger errors. A lower MSE indicates better performance. RMSE (Root Mean Squared Error) is a valuable metric that provides insight into prediction accuracy by taking the square root of MSE, and it is expressed in the same units as the original data. In this study, we have compared various results based on the RMSE. A lower RMSE suggests higher accuracy, and we often analyze the RMSE as a percentage, which effectively represents the model accuracy for SOC estimation on a 100% scale, providing better interpretability of the results. Additionally, the runtime metric measures the duration of the model's training process, with shorter runtimes being preferable for optimal efficiency.

$$MAE = \frac{\sqrt{\sum_{t=1}^{n} |SOC_t - SOC'_t|}}{n} \tag{17}$$

$$MSE = \frac{\sum_{t=1}^{n} (SOC_t - SOC'_t)^2}{n} \tag{18}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (SOC_t - SOC'_t)^2}{n}} \tag{19}$$

4. Results and discussion

In this section, a comprehensive comparison is conducted to evaluate the proposed deep learning (DL) methods applied to the datasets. For the McMaster LG-18650HG2 battery dataset, the performance of four distinct models like Temporal Attention-based BiLSTM, LSTM, CNN-BiLSTM, and CNN-LSTM has been assessed. These models were evaluated using four different metrics: MSE, RMSE, MAE, and the Coefficient of Determination (R^2) across five different temperature conditions.

The experiments were conducted in three distinct phases. In the first phase, the models were trained and tested under ambient temperature conditions. The second phase involved running all the models and presenting comparative results across various temperatures (0 °C, 10 °C, 40 °C, −20 °C). Additionally, a detailed comparative analysis was performed between the original Keras-based models and their quantised versions to assess the effects of quantisation on predictive performance. This comparison highlights the impact of quantisation techniques on model accuracy and computational efficiency, providing valuable insights into the trade-offs between performance and efficiency when utilizing PTQ.

4.1. Hyper parameter tuning at 25 °C ambient temperature

Selecting appropriate hyperparameters is a critical aspect of deep neural network (DNN) performance analysis. In this study, we did not employ explicit hyperparameter optimization techniques such as grid search or Bayesian optimization. Instead, we adopted an iterative trial-and-error approach to determine the optimal hyperparameters. The selected hyperparameters were then consistently applied across all models to evaluate the impact of model architecture under controlled conditions. Table 2 presents the results of experiments conducted with varying combinations of learning rates, ranging from 0.01 to 0.001, and batch sizes between 32 and 128 for the proposed BiLSTM-TAM at 25 °C ambient temperature. This exploration focused on assessing the effects of these two pivotal hyperparameters — learning rate and batch size — on model performance. As shown in Table 2, the optimal combination was determined to be a learning rate of 0.01 and a batch size of 128. To ensure a fair comparison, these settings were consistently applied across all the models.

Underfitting occurs when a model fails to capture the underlying patterns in the data due to limited capacity or inadequate training. To overcome this, we implemented a BiLSTM architecture with two layers: the first containing 64 units and the second 32 units. To help

Table 2
RMSE (%) for different drive cycles across various learning rates and batch sizes at ambient temperature.

Drive cycle	LR = 0.01 Batch size = 128	LR = 0.0001 Batch size = 128	LR = 0.001 Batch size = 128	LR = 0.001 Batch size = 64	LR = 0.001 Batch size = 32
HPPC	0.9436	1.0075	**0.7000**	0.8196	0.6741
UDDS	0.6556	0.7124	**0.5080**	0.5611	0.5808
US06	0.7520	0.5976	**0.3970**	0.6021	0.6376
Mixed	0.4346	0.5611	**0.4130**	0.3101	0.3968

Table 3
Model training parameters.

Parameters	Value
Optimizer	Adam
Learning rate	0.001
Activations	ReLU (First BiLSTM)
	Linear activation (Second BiLSTM)
	Tanh and Softmax (Attention Mechanism)
Batch size	128
Number of epochs	100

Table 4
Statistical analysis of model performance across different drive cycles at 25 °C.

Drive cycle	Model	MSE	RMSE	MAE	R^2
HPPC	LSTM	1.19E−04	0.01090	0.00572	0.99888
	CNN-LSTM	9.41E−05	0.00970	0.00592	0.99911
	CNN-BiLSTM	6.64E−05	0.00815	0.00623	0.99937
	BiLSTM-TAM	**4.90E−05**	**0.00700**	**0.00505**	**0.99954**
UDDS	LSTM	5.39E−05	0.00734	0.00613	0.99932
	CNN-LSTM	6.02E−05	0.00776	0.00696	0.99924
	CNN-BiLSTM	3.60E−05	0.00600	0.00491	0.99954
	BiLSTM-TAM	**2.58E−05**	**0.00508**	**0.00400**	**0.99967**
US06	LSTM	2.53E−05	0.00503	0.00414	0.99972
	CNN-LSTM	4.13E−05	0.00643	0.00557	0.99954
	CNN-BiLSTM	2.34E−05	0.00484	0.00409	0.99974
	BiLSTM-TAM	**1.57E−05**	**0.00397**	**0.00347**	**0.99982**
Mixed	LSTM	1.87E−05	0.00432	0.00318	0.99973
	CNN-LSTM	1.75E−05	0.00418	0.00346	0.99977
	CNN-BiLSTM	1.96E−05	0.00443	0.00378	0.99977
	BiLSTM-TAM	**1.71E−05**	**0.00413**	**0.00322**	**0.99980**

the model focus on the most important features, we introduced an attention mechanism with a 64-dimensional attention vector that works alongside the BiLSTM outputs. This mechanism allows the model to prioritize relevant parts of the input sequence while minimizing the impact of irrelevant or noisy information. To improve generalization, we added dropout layers with a rate of 20% after each BiLSTM layer, randomly deactivating neurons during training. We monitored training performance using validation loss and applied EarlyStopping to halt training when the validation performance stopped improving, preventing overfitting. Additionally, a model checkpoint was used to save the best version of the model based on validation loss, ensuring that the final model performs effectively on unseen data. The network parameters used during the DNN training are presented in Table 3. By systematically varying these parameters, we aimed to identify a configuration that balances convergence speed and generalization while providing insights into their influence on different models.

4.2. SOC estimation results at 25 °C ambient temperature

The error metrics for the four different models obtained at 25° after training of the DL models are presented in Table 4 . The evaluation is conducted on four different drive cycles: HPPC, UDDS, US06, Mixed. These results indicate that the proposed Temporal Attention-based BiLSTM model outperformed the other three competing models which achieves the lowest RMSE of 0.7%, indicating minimal prediction error on the HPPC drive cycle. It is true across all four drive cycles, trained and tested at 25 °C. An in depth analysis of RMSE for SOC estimation across different drive cycles is presented in Fig. 7. The BiLSTM-TAM model shows strong performance across all drive cycles, with RMSE values of 0.70% for HPPC, 0.508% for UDDS, 0.397% for US06, and 0.413% for the Mixed cycle, indicating consistently accurate predictions with minimal deviation from the true values. At room temperature, the BiLSTM-TAM model demonstrates a modest improvement in RMSE and MAE compared to the CNN-LSTM and CNN-BiLSTM models. While the numerical gains might appear minimal, they carry substantial importance in real-world applications. For example, in electric vehicles (EVs), precise estimation of the battery's available energy directly influences the reliability of the remaining range predictions. Enhanced accuracy in these estimates not only minimizes unnecessary recharging stops but also optimizes the overall utilization of the battery's capacity, ultimately improving the user experience and operational efficiency. Fig. 8 presents the detailed SOC estimation curves and the residuals for the DL networks. The R^2 value for all models across each drive cycle is almost 0.99. The suggested BiLSTM-TAM model presents the most consistent and better result in terms of estimating the SOC. The error for the BiLSTM-TAM model, represented by the red line, is lower than that of any other model used for estimation.

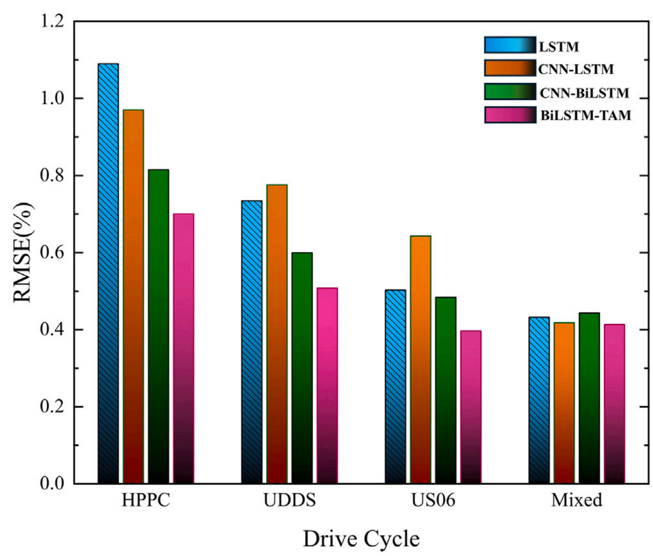


Fig. 7. Comparison of RMSE values for different models across various drive cycles (HPPC, UDDS, US06, and Mixed) at 25 °C.

4.3. SOC estimation results at 0 °C and 10 °C

Ambient temperature is a crucial determinant in precisely assessing the SOC because the battery's intricate dynamics significantly impact SOC estimations. Batteries show varying behaviors under different situations, including seasonal changes and daytime temperature fluctuations, especially in certain European nations where winter temperatures can vary significantly. This section summarizes the results obtained from the LSTM, CNN-BiLSTM, CNN-LSTM, and BiLSTM-TAM models at 0 °C and 10 °C to evaluate the efficacy of the suggested deep learning models for SOC estimation.

Table 5 presents a thorough assessment of SOC estimation metrics for the proposed four models at 0 °C temperature levels, highlighting

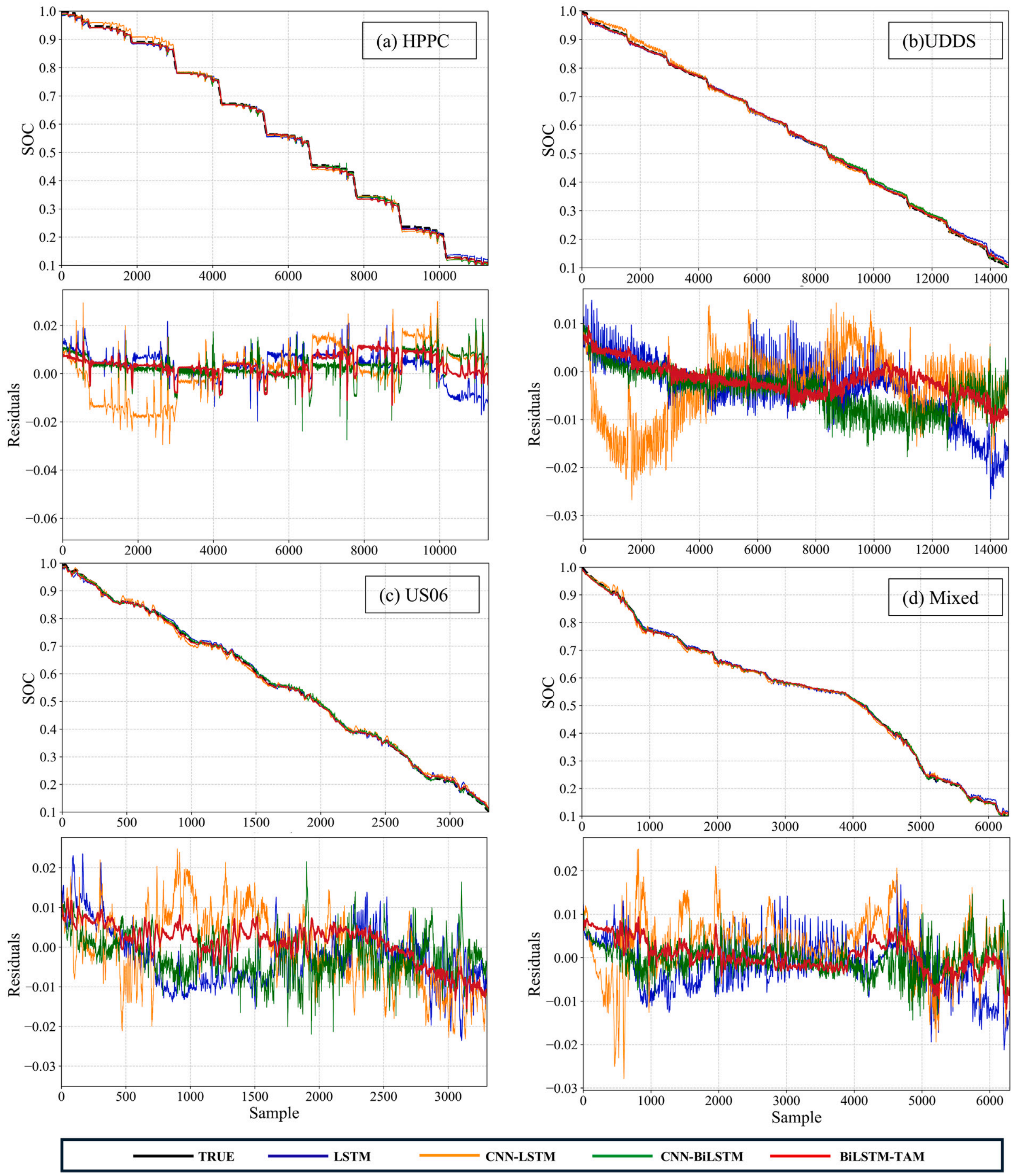


Fig. 8. State of Charge (SOC) estimation using four different model at 25 °C for different drive cycles: (a) HPPC, (b) UDDS, (c) US06, and (d) Mixed. The top plots in each subfigure show the comparison between the estimated SOC and the true SOC over time (sample index). The bottom plots display the corresponding residuals, representing the difference between the estimated and true SOC values.

Table 5
Statistical analysis of model performance across different drive cycles at 0 °C.

Drive cycle	Model	MSE	RMSE	MAE	R^2
HPPC	LSTM	1.52E−05	0.003904	0.003230	0.999813
	CNN-LSTM	7.24E−06	0.002691	0.002155	0.999911
	CNN-BiLSTM	4.44E−06	0.002108	0.001699	0.999946
	BiLSTM-TAM	4.10E−06	0.002024	0.001512	0.999950
UDDS	LSTM	2.93E−05	0.005409	0.003759	0.999604
	CNN-LSTM	2.32E−05	0.004817	0.004399	0.999686
	CNN-BiLSTM	1.24E−05	0.003519	0.003085	0.999832
	BiLSTM-TAM	5.98E−06	0.002446	0.001983	0.999919
US06	LSTM	4.86E−05	0.006972	0.005622	0.999437
	CNN-LSTM	1.71E−05	0.004138	0.003565	0.999802
	CNN-BiLSTM	1.47E−05	0.003829	0.003010	0.999830
	BiLSTM-TAM	7.88E−06	0.002807	0.002100	0.999909
Mixed	LSTM	2.50E−05	0.005001	0.003819	0.999663
	CNN-LSTM	1.51E−05	0.003888	0.002819	0.999797
	CNN-BiLSTM	2.16E−05	0.004650	0.004222	0.999723
	BiLSTM-TAM	1.02E−05	0.003200	0.002670	0.999869

Table 6
Statistical analysis of model performance across different drive cycles at 10 °C.

Drive cycle	Model	MSE	RMSE	MAE	R^2
HPPC	LSTM	2.43E−04	0.015602	0.005718	0.997476
	CNN-LSTM	2.16E−04	0.014697	0.0047015	0.998939
	CNN-BiLSTM	1.42E−04	0.011896	0.003967	0.998533
	BiLSTM-TAM	1.31E−04	0.011464	0.004613	0.998637
UDDS	LSTM	1.65E−05	0.004057	0.003489	0.999776
	CNN-LSTM	3.80E−06	0.001950	0.001560	0.999948
	CNN-BiLSTM	3.56E−06	0.001888	0.001559	0.999951
	BiLSTM-TAM	1.43E−06	0.001197	0.000954	0.999980
US06	LSTM	7.79E−05	0.008826	0.007159	0.999141
	CNN-LSTM	1.17E−05	0.003419	0.002682	0.999871
	CNN-BiLSTM	1.20E−05	0.003461	0.002819	0.999868
	BiLSTM-TAM	7.20E−06	0.002684	0.002203	0.999921
Mixed	LSTM	2.99E−05	0.005464	0.004365	0.999563
	CNN-LSTM	5.61E−06	0.002368	0.001714	0.999928
	CNN-BiLSTM	1.79E−05	0.004231	0.003817	0.999771
	BiLSTM-TAM	6.67E−06	0.002582	0.002235	0.999915

their performance under different thermal settings. In the HPPC cycle, BiLSTM-TAM attains an RMSE of 0.202

The statistical analysis at 10 °C provides valuable insights into the performance of various techniques across different datasets. The proposed technique demonstrates the lowest RMSE values, highlighting its superior accuracy in predictions. Table 6 provides a comparison of different error values. At this temperature, the RMSE for the HPPC drive cycle is slightly higher compared to results at other temperatures, attaining 1.56% for LSTM and 1.01% for CNN-LSTM.

BiLSTM-TAM demonstrates superior performance in the HPPC, UDDS, and US06 cycles, achieving an RMSE of 0.12% and 0.27%, respectively, highlighting the model's effectiveness in EVs' standard drive cycles. The visual comparative analysis of different models at this temperature is presented in Supplementary Figure 2. Fig. 10 depicts the estimated SOC, the actual SOC, and the corresponding residuals at this temperature. The red line, representing the error of the proposed BiLSTM-TAM model, demonstrates consistent and superior performance across the entire SOC range. In contrast, other models, while showing good results in certain areas, fail to maintain consistent accuracy throughout the range (see Fig. 9).

4.4. SOC estimation results at extreme weather of 40 °C and −20 °C

In certain regions, like parts of Europe, winter temperatures can fall below zero degrees Celsius, leading to noticeable changes in battery behavior. On the other hand, in regions such as Asia and Africa, summer temperatures can reach up to 40 °C, also affecting battery performance.

Table 7
Comparison of RMSE and MAE for different models at 40 °C and −20 °C across drive cycles.

Drive cycle	Model	40 °C		−20 °C	
		RMSE	MAE	RMSE	MAE
HPPC	LSTM	0.057437	0.011675	0.003945	0.003358
	CNN-LSTM	0.067301	0.013344	0.003388	0.002803
	CNN-BiLSTM	0.057200	0.013037	0.002712	0.002067
	BiLSTM-TAM	0.008136	0.007069	0.002841	0.002489
UDDS	LSTM	0.005473	0.005036	0.001839	0.001519
	CNN-LSTM	0.004479	0.003502	0.002845	0.002345
	CNN-BiLSTM	0.003251	0.002694	0.001892	0.001411
	BiLSTM-TAM	0.002686	0.001861	0.001888	0.001506
US06	LSTM	0.006007	0.005003	0.004265	0.003638
	CNN-LSTM	0.006494	0.005126	0.004404	0.003829
	CNN-BiLSTM	0.004712	0.003979	0.002992	0.002389
	BiLSTM-TAM	0.003618	0.002895	0.003213	0.002799
Mixed	LSTM	0.004321	0.003185	0.005254	0.004901
	CNN-LSTM	0.005358	0.004306	0.005151	0.004837
	CNN-BiLSTM	0.005061	0.004041	0.002437	0.001994
	BiLSTM-TAM	0.003230	0.002433	0.003021	0.002514

To demonstrate this, the proposed models are tested in extreme cold conditions, such as −20 °C, and extreme hot conditions, like 40 °C, to assess how batteries perform in these challenging environments.

Table 7 shows a comparative analysis of this proposed models in 40 °C and −20 °C. In terms of the HPPC drive cycle at 40 °C, the RMSE values are higher for all models compared to the other drive cycles. During the HPPC cycle at 40 °C. At −20 °C, the CNN-LSTM model exhibits the lowest RMSE of 0.27%, closely followed by the BiLSTM-TAM at 0.28%. During the UDDS cycle at 40 °C, LSTM exhibits an RMSE of 0.55%, but BiLSTM-TAM demonstrates superior performance with an RMSE of 0.27%. Therefore, it is conclusive that battery dynamics at higher ambient temperatures are more intricate, imposing some considerable challenges for accurate SOC estimation. At −20 °C, BiLSTM-TAM attains the lowest RMSE of 0.19%, in contrast to LSTM's 0.18% for the UDDS drive cycle. In the US06 cycle at 40 °C, LSTM registers an RMSE of 0.60%, but BiLSTM-TAM presents a notable enhancement with an RMSE of 0.36%. At −20 °C, BiLSTM-TAM achieves an RMSE of 0.32%, whereas LSTM has a higher value of 0.43%.Finally, in the Mixed cycle at 40 °C, LSTM exhibits an RMSE of 0.43%, while BiLSTM-TAM demonstrates an enhancement with an RMSE of 0.30%. At −20 °C, BiLSTM-TAM attains superior performance with an RMSE of 0.25%, in contrast to LSTM's 0.53%. In summary, at −20 °C, the BiLSTM-TAM model consistently delivers superior results in nearly all cases. However, in a few instances, the CNN-BiLSTM model performs slightly better than BiLSTM-TAM at −20 °C. his suggests that the convolutional network is capable of tracking changes to a certain extent due to its filtering methods. At extreme temperatures, such as 40 °C and −20 °C, the SOC of the battery is limited to approximately 40% due to the significant impact of temperature on its electrochemical processes and overall performance. That is why in these temperatures we have run the model only up to this certain range to avoid wrong or invalid estimation. At higher temperatures (40 °C), reactions like electrolyte breakdown and lithium plating happen more quickly, which slowly decreases the battery's performance and capacity. Conversely, at low temperatures (−20 °C), the movement of lithium ions slows significantly, leading to increased internal resistance, which reduces the battery's ability to accept or deliver charge, thereby limiting its capacity. As a result, SOC estimation at these temperatures is constrained to around 40%. Given the length of this article, the SOC estimation curves generated by different models under these temperature conditions are provided in the supplementary materials.

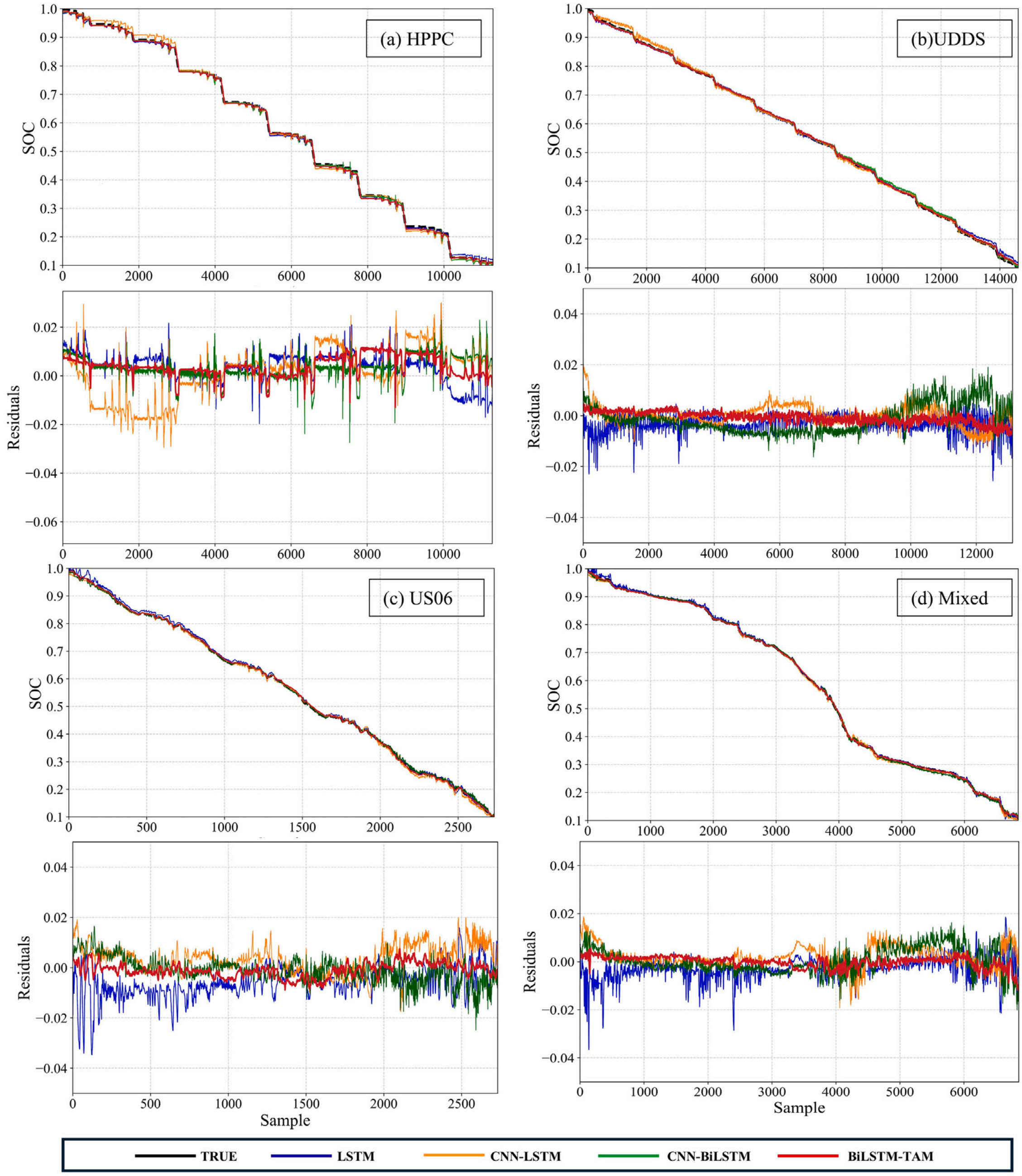


Fig. 9. State of Charge (SOC) estimation using four different model at 0 °C for different drive cycles: (a) HPPC, (b) UDDS, (c) US06, and (d) Mixed. The top plots in each subfigure show the comparison between the estimated SOC and the true SOC over time (sample index). The bottom plots display the corresponding residuals, representing the difference between the estimated and true SOC values.

4.5. Quantisation results for proposed models

We conducted a performance comparison between the Quantised model and the original Keras model without quantisation after PTQ. A comprehensive analysis of the percentage change in model performance in terms of RMSE after quantisation for proposed BiLSTM-TAM is provided in Table 8. This comparative analysis highlights the effects of PTQ on the accuracy of the proposed BiLSTM-TAM model, illustrating the extent to which the quantisation process influences the model's predictive capabilities. The purpose of the analysis is to assess the trade-offs between the accuracy and the size reduction of the model. The percentage change in model accuracy is calculated using the following formula:

$$\text{Percentage Change} = \left(\frac{\text{Keras RMSE} - \text{Quantised Model RMSE}}{\text{Keras RMSE}} \right) \times 100 \tag{20}$$

At 0 °C, the Quantised model exhibits a 7.00% improvement in HPPC, while demonstrating a reduction of −5.62% in the Mixed cycle, with negligible variations for UDDS and US06. At 10 °C, the impact is minimal, with a −1.01% reduction in HPPC and minor fluctuations in other cycles. At 25 °C, UDDS rises by 2.78%, whilst Mixed decreases by 2.57%, reflecting mixed outcomes. At 40 °C, performance is consistent, exhibiting modest improvements in UDDS 4.65% and Mixed 4.95%. At

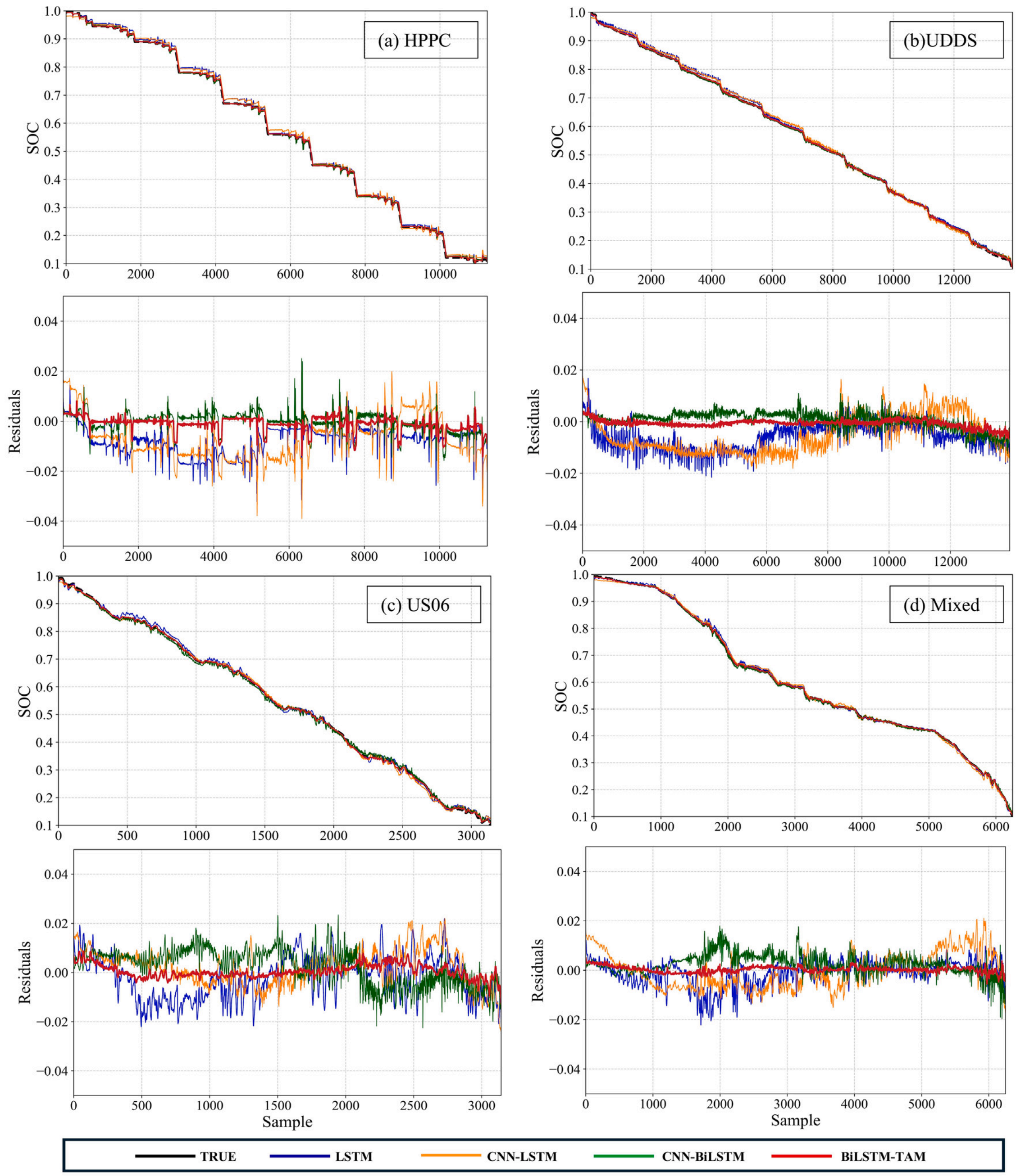


Fig. 10. State of Charge (SOC) estimation using four different model at 10 °C for different drive cycles: (a) HPPC, (b) UDDS, (c) US06, and (d) Mixed. The top plots in each subfigure show the comparison between the estimated SOC and the true SOC over time (sample index). The bottom plots display the corresponding residuals, representing the difference between the estimated and true SOC values.

−20 °C, variability escalates: UDDS and Mixed exhibit enhancements of 4.98% and 4.01%, respectively, while HPPC and US06 demonstrate negligible variations.

Table 9 presents the comparison of Keras and Quantised RMSE across temperatures and drive cycles for CNN-BiLSTM model. The Quantised model exhibits a substantial decrease in performance for the HPPC drive cycle at 0 °C, with a −10.01% decrease in RMSE. However, the UDDS 1.28% and Mixed 2.13% cycles exhibit minor improvements, whereas the US06 cycle exhibits a moderate deterioration of −4.36%. The impact of quantisation is negligible at 10 °C, with negligible changes in HPPC and UDDS −0.09% and −0.16%, respectively, and minor improvements in US06 0.26% and Mixed 0.47%. Performance deterioration is notably evident in UDDS −10.02% and US06 −12.09% at 25 °C, with Mixed exhibiting the largest decrease −14.62%. The sole positive change at this temperature is a modest improvement in HPPC 1.63%. The Quantised model remains stable at 40 °C, with HPPC exhibiting a slight decrease −0.61% and UDDS improving 4.44%. Minor gains are also observed for US06 1.29% and Mixed 0.78%, suggesting that the temperature has a minimal impact. The results at −20 °C are inconsistent, with minor declines in HPPC −2.51% and US06 −2.94%, an improvement in UDDS 4.49%, and a small decrease in Mixed −1.76%. These findings highlight the moderate impact of quantisation at lower temperatures. To reduce the redundancy of the article we opted out the detailed quantisation analysis of the other two suggested models.

The Table 10 compares the sizes of Keras, TFLite, and Quantised models in MB, emphasizing the size reduction attained with quantisation. Although quantisation is expected to decrease the model size

Table 8
Comparison of Keras and Quantised RMSE across temperatures and drive cycles for Proposed BiLSTM-TAM model.

Temperature	Drive cycle	Keras_RMSE	Quantised_RMSE	Percent change (%)
0 °C	HPPC	0.0020243	0.0018825	7.00
	UDDS	0.0024461	0.0024523	−0.25
	US06	0.0028067	0.0027719	1.24
	Mixed	0.0031996	0.0033794	−5.62
10 °C	HPPC	0.0114638	0.0115790	−1.01
	UDDS	0.0011970	0.0011860	0.92
	US06	0.0026840	0.0028010	−4.36
	Mixed	0.0025820	0.0027690	−7.24
25 °C	HPPC	0.0070033	0.0070154	−0.17
	UDDS	0.0050780	0.0049370	2.78
	US06	0.0039660	0.0039100	1.41
	Mixed	0.0041338	0.0042399	−2.57
40 °C	HPPC	0.0081360	0.0082640	−1.57
	UDDS	0.0026860	0.0025610	4.65
	US06	0.0036180	0.0035400	2.16
	Mixed	0.0032300	0.0030700	4.95
−20 °C	HPPC	0.0028410	0.0028280	0.46
	UDDS	0.0018880	0.0017940	4.98
	US06	0.0032130	0.0029940	6.82
	Mixed	0.0030210	0.0029000	4.01

Table 9
Comparison of Keras and Quantised RMSE across temperatures and drive cycles for CNN-BiLSTM model.

Temperature	Drive cycle	Keras_RMSE	Quantised_RMSE	Percent change (%)
0 °C	HPPC	0.0021080	0.0023190	−10.01
	UDDS	0.0035190	0.0034740	1.28
	US06	0.0038290	0.0039960	−4.36
	Mixed	0.0046500	0.0045510	2.13
10 °C	HPPC	0.0118960	0.0119070	−0.09
	UDDS	0.0018880	0.0018910	−0.16
	US06	0.0034610	0.0034520	0.26
	Mixed	0.0042310	0.0042110	0.47
25 °C	HPPC	0.0081510	0.0080180	1.63
	UDDS	0.0059980	0.0065990	−10.02
	US06	0.0048390	0.0054240	−12.09
	Mixed	0.0044310	0.0050790	−14.62
40 °C	HPPC	0.0571995	0.0575476	−0.61
	UDDS	0.0032508	0.0031066	4.44
	US06	0.0047124	0.0046514	1.29
	Mixed	0.0050612	0.0050218	0.78
−20 °C	HPPC	0.0027120	0.0027800	−2.51
	UDDS	0.0018920	0.0018070	4.49
	US06	0.0029920	0.0030800	−2.94
	Mixed	0.0024370	0.0024800	−1.76

Table 10
Model size comparison and reduction percentage.

Model name	Keras model	TFLite	Quantised model	Reduced model size (%)
LSTM	0.38	0.13	0.06	46.15
CNN-LSTM	0.55	0.19	0.08	42.11
CNN-BiLSTM	1.19	0.41	0.16	39.02
BiLSTM-TAM	0.96	0.33	0.13	39.39

to 25% of the TFLite size, the actual reductions vary between 39% and 46%. The LSTM model reduced to 46.15%, whereas the CNN-LSTM and BiLSTM-TAM models reduced to the range from 39% to 42%. Despite being beyond the theoretical minimum, these reductions greatly improve model size for efficient deployment in edge devices.

4.6. Comparative analysis

To validate the performance of the proposed model, a comparative analysis is presented in Table 11. The table highlights the significant contributions of previous researchers in accurately estimating the SOC using various RNN and DNN models. Each model has been applied to different battery types and drive cycles with varying degrees of success.

The proposed BiLSTM-TAM model for this study, however, achieves a notably lower RMSE compared to the other models. This improvement demonstrates the effectiveness of integrating both Bidirectional LSTM and a Temporal Attention Mechanism in SOC estimation. The substantial reduction in RMSE shows that the proposed approach can provide more accurate predictions across multiple drive cycles, outperforming other methods such as LSTM, FNN, and DCNN, as illustrated in the table.

4.7. Discussion

The BiLSTM-TAM model has improved accuracy in SOC prediction relative to existing DNN models, especially at ambient temperatures of 25 °C and in colder environments at 0 °C. At these temperatures, the proposed technique outperforms alternatives like LSTM, CNN-LSTM, and CNN-BiLSTM. The observed results confirm the superior capacity of networks with attention mechanisms in learning the influence of temperature variations, which is usually hard to capture using conventional filtering methods. However, in the HPPC cycle, which demands consistent current pulses for battery parameter estimation, the performance of the proposed models slightly declines compared to other drive cycles, particularly at lower temperatures like 0 °C. BiLSTM-TAM consistently delivers superior results for standardized EV drive cycles,. In extreme heat, such as at 40 °C, the BiLSTM-TAM model effectively captures temperature-related variations, yielding better SOC prediction accuracy. At high temperatures, the chemical reactions within batteries, such as electrolyte breakdown and side reactions like the formation of the solid electrolyte interphase (SEI), typically increase. This results in capacity degradation and heightened self-discharge, complicating SOC monitoring. Consequently, the RMSE escalates, especially in the HPPC cycle, when values of 5.74% and 6.73% were recorded.

At extremely low temperatures, such as −20 °C, forecasting SOC becomes difficult due to increased internal resistance, reduced electrochemical processes, and impaired ion diffusion, all of which complicate battery performance. These variables makes SOC estimation especially challenging under such situations. At this temperature, the CNN-BiLSTM model marginally surpasses the BiLSTM-TAM model. This may be attributed to the fact that, at very low temperatures, short-term, non-linear effects such as increased resistance and slower chemical reactions play a more dominant role which are being captured by the convolutional layer. However, the BiLSTM-TAM model still delivers the second-best performance in these conditions. The BiLSTM-TAM model consistently detects differences in various driving cycles across all investigated temperatures, demonstrated by a R^2 value of 99% in nearly every situation. This implies that the proposed framework can effectively track and forecast SOC across diverse operational situations, illustrating its robustness.

In terms of training time, a comparison of the four models at −20 °C is presented in Table 12. The results indicate that the proposed BiLSTM-TAM model requires more training time than the other models. However, as previously mentioned, it delivers the highest accuracy in SOC estimation. Notably, the training time for the BiLSTM-TAM model is significantly lower compared to other studies, where training durations are considerably longer, as referenced in Sherkatghanad et al. (2024). This highlights the efficiency of the proposed model, which achieves superior SOC estimation accuracy in a much shorter training period. It is important to note that although minor variations in the training data (e.g., sequence selection differences) were introduced, the model performance remained consistent across these settings. This demonstrates the robustness of the proposed approach with respect to data partitioning and confirms that the results are not sensitive to specific training splits.

Table 11
Model comparison for different drive cycles and error metrics.

Reference	Battery type	Model used	Drive cycles	Errors
Chen et al. (2023)	LiFePO4	EI-LSTM-RNN-CO	DST, FUDS, USABC	RMSE: 1.3%
Wang et al. (2023)	INR 18650	DCNN	DST, BJDST, FUDS, US06, UDDS	RMSE: 1.78%
Khan et al. (2024)	Li-ion	Deep NN	Dynamic cycles	RMSE: 0.7%
Zhang et al. (2025)	Li-ion (NCM)	CNN-LSTM	Multiple datasets	RMSE: 0.39%
Proposed	LG 18650HG2	BiLSTM-TAM	HPPC, UDDS, US06, Mixed	RMSE: 0.19%

Table 12
Training time for different models.

Model	Training time [Min]
LSTM	13
CNN-LSTM	11
CNN-BiLSTM	13
BiLSTM-TAM	21

The quantisation of weights and activations results in a substantial reduction in model size through lower bit representation. In this study, the Quantised model showed a maximum 15% reduction in accuracy, as measured by RMSE, while still providing strong results for SOC estimation. In some instances, the model's accuracy was enhanced following quantisation, which was intriguing. This improvement could be ascribed to minor fluctuations in model weights, where decreasing precision encourages the model to focus on more prominent and relevant features, essentially serving as a type of regularization. This can also be explained by the fact that quantisation typically reduces accuracy when the model is highly complex, and the data is both large and intricate (Wang et al., 2024). However, since these conditions are not present in this case, the accuracy can remain high or even improve. Despite the decrease in precision, deviation from the original model is still within acceptable ranges. The Quantised model's size reduced to roughly 38% to 47% in comparison to the TensorFlow Lite (TFLite) version of the model. Although weights are generally Quantised to 8-bit integers, other elements such as biases, activations, or specific layers like BatchNorm or LSTM may maintain 32-bit precision, hence hindering the model from realizing the complete 75% decrease in size. However, the Quantised TFLite model is considerably smaller than the original Keras model, indicating that this method is particularly appropriate for deployment on edge devices, where model efficiency is essential.

5. Conclusion

In conclusion, this paper presented a novel bidirectional Long Short-Term Memory Temporal Attention Mechanism (BiLSTM-TAM)-based model with post-training quantisation for accurately estimating the State of Charge (SOC) of lithium-ion (Li-ion) batteries in electric vehicles (EVs) on edge devices. Rigorous testing across various drive cycles and temperatures ranging from −20 to 40 degrees Celsius demonstrated the effectiveness of the proposed model, achieving a root-mean-square error (RMSE) as low as 0.11% and a mean absolute error (MAE) of 0.095%, with a coefficient of determination (R^2) of 99.9%. The application of quantisation resulted in a significant performance enhancement, increasing the RMSE by up to only 15% and shrinking the model size by nearly 60%. Comparative analysis with traditional models such as LSTM, CNN-LSTM, and CNN-BiLSTM confirmed the superiority of the proposed BiLSTM-TAM model. Furthermore, the quantised model's performance, compared to the original Keras model, highlights its potential for efficient deployment on edge devices. The implications of this research are significant for the EV industry, as the proposed models are lightweight, and the quantised version can be deployed on small edge devices for accurate SOC estimation. This ensures safe and efficient energy consumption while enabling precise battery monitoring. Moving forward, our future work will focus on further improving SOC estimation performance through the development of alternative attention-based deep learning frameworks and the exploration of metaheuristic optimization techniques for hyperparameter tuning. Additionally, we plan to investigate other model quantisation methods, such as quantisation-aware training and pruning, to quantize models during training and explore pruning techniques to further reduce model size. This research contributes to advancing EV technology by offering a promising solution to key challenges related to battery management systems.

CRediT authorship contribution statement

Opy Das: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Muhammad Hamza Zafar:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Souman Rudra:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Filippo Sanfilippo:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the Artificial Intelligence, Biomechatronics, and Collaborative Robotics research group at the Top Research Center Mechatronics (TRCM), University of Agder (UiA), Norway, and we greatly appreciate the financial support provided by the NORPART-2021/10355 project, which is funded by the Norwegian Directorate for Higher Education and Skills (HK-Dir).

Data availability

LG 18650HG2 Li-ion Battery Data: https://data.mendeley.com/datasets/cp3473x7xv/3.

References

Amir, S., Gulzar, M., Tarar, M.O., Naqvi, I.H., Zaffar, N.A., Pecht, M.G., 2022. Dynamic equivalent circuit model to estimate state-of-health of lithium-ion batteries. IEEE Access 10, 18279–18288.

Anton, J.C.A., Nieto, P.J.G., Viejo, C.B., Vilán, J.A.V., 2013. Support vector machines used to estimate the battery state of charge. IEEE Trans. Power Electron. 28 (12), 5919–5926.

Aslan, M.F., Unlersen, M.F., Sabanci, K., Durdu, A., 2021. CNN-based transfer learning–bilstm network: A novel approach for COVID-19 infection detection. Appl. Soft Comput. 98, 106912.

Cacciato, M., Nobile, G., Scarcella, G., Scelba, G., 2016. Real-time model-based estimation of SOC and SOH for energy storage systems. IEEE Trans. Power Electron. 32 (1), 794–803.

Chang, W.-Y., 2013. Estimation of the state of charge for a LFP battery using a hybrid method that combines a RBF neural network, an OLS algorithm and AGA. Int. J. Electr. Power Energy Syst. 53, 603–611.

Chen, W., Liang, J., Yang, Z., Li, G., 2019. A review of lithium-ion battery for electric vehicle applications and beyond. Energy Procedia 158, 4363–4368.

Chen, J., Zhang, Y., Wu, J., Cheng, W., Zhu, Q., 2023. SOC estimation for lithium-ion battery using the LSTM-RNN with extended input and constrained output. Energy 262, 125375.

Chrysostomou, G., Aletras, N., 2021. Improving the faithfulness of attention-based explanations with task-specific information for text classification. arXiv preprint arXiv:2105.02657.

Cui, Z., Kang, L., Li, L., Wang, L., Wang, K., 2022a. A combined state-of-charge estimation method for lithium-ion battery using an improved BGRU network and UKF. Energy 259, 124933.

Cui, Z., Wang, L., Li, Q., Wang, K., 2022b. A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network. Int. J. Energy Res. 46 (5), 5423–5440.

Dai, K., Wang, J., He, H., 2019. An improved SOC estimator using time-varying discrete sliding mode observer. IEEE Access 7, 115463–115472.

Deng, Z., Hu, X., Lin, X., Che, Y., Xu, L., Guo, W., 2020. Data-driven state of charge estimation for lithium-ion battery packs based on Gaussian process regression. Energy 205, 118000.

Dos Reis, G., Strange, C., Yadav, M., Li, S., 2021. Lithium-ion battery data and where to find it. Energy AI 5, 100081.

Du, B., Yu, Z., Yi, S., He, Y., Luo, Y., 2021. State-of-charge estimation for second-life lithium-ion batteries based on cell difference model and adaptive fading unscented Kalman filter algorithm. Int. J. Low-Carbon Technol. 16 (3), 927–939.

Gong, D., Gao, Y., Kou, Y., 2021. Parameter and state of charge estimation simultaneously for lithium-ion battery based on improved open circuit voltage estimation method. Energy Technol. 9 (9), 2100235.

Hochreiter, S., 1997. Long short-term memory. Neural Comput. MIT-Press.

How, D.N., Hannan, M., Lipu, M.H., Ker, P.J., 2019. State of charge estimation for lithium-ion batteries using model-based and data-driven methods: A review. IEEE Access 7, 136116–136136.

Hu, X., Li, S., Peng, H., Sun, F., 2012. Robustness analysis of state-of-charge estimation methods for two types of Li-ion batteries. J. Power Sources 217, 209–219.

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., Kalenichenko, D., 2018. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2704–2713.

Jiao, M., Wang, D., Qiu, J., 2020. A GRU-RNN based momentum optimized algorithm for SOC estimation. J. Power Sources 459, 228051.

Khan, U., Kirmani, S., Rafat, Y., Rehman, M.U., Alam, M.S., 2024. Improved deep learning based state of charge estimation of lithium ion battery for electrified transportation. J. Energy Storage 91, 111877.

Korkmaz, M., 2023. SoC estimation of lithium-ion batteries based on machine learning techniques: A filtered approach. J. Energy Storage 72, 108268.

Lai, X., Zheng, Y., Sun, T., 2018. A comparative study of different equivalent circuit models for estimating state-of-charge of lithium-ion batteries. Electrochim. Acta 259, 566–577.

Lee, S., Kim, J., Lee, J., Cho, B.H., 2007. The state and parameter estimation of an li-ion battery using a new OCV-soc concept. In: 2007 IEEE Power Electronics Specialists Conference. IEEE, pp. 2799–2803.

Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H., Yin, S., 2023. Deep learning attention mechanism in medical image analysis: Basics and beyonds. Int. J. Netw. Dyn. Intell. 93–116.

Li, I.-H., Wang, W.-Y., Su, S.-F., Lee, Y.-S., 2007. A merged fuzzy neural network and its applications in battery state-of-charge estimation. IEEE Trans. Energy Convers. 22 (3), 697–708.

Lin, Z., Feng, M., Santos, C.N.d., Yu, M., Xiang, B., Zhou, B., Bengio, Y., 2017. A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130.

Liu, Y., He, Y., Bian, H., Guo, W., Zhang, X., 2022. A review of lithium-ion battery state of charge estimation based on deep learning: Directions for improvement and future trends. J. Energy Storage 52, 104664.

Manoharan, A., Begam, K., Aparow, V.R., Sooriamoorthy, D., 2022. Artificial neural networks, gradient boosting and support vector machines for electric vehicle battery state estimation: A review. J. Energy Storage 55, 105384.

Marelli, S., Corno, M., 2020. Model-based estimation of lithium concentrations and temperature in batteries using soft-constrained dual unscented Kalman filtering. IEEE Trans. Control Syst. Technol. 29 (2), 926–933.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62.

Paupamah, K., James, S., Klein, R., 2020. Quantisation and pruning for neural network compression and regularisation. In: 2020 International SAUPEC/RobMech/PRASA Conference. IEEE, pp. 1–6.

Qi, K., Zhang, W., Zhou, W., Cheng, J., 2022. Integrated battery power capability prediction and driving torque regulation for electric vehicles: A reduced order MPC approach. Appl. Energy 317, 119179.

Rosin, G.D., Radinsky, K., 2022. Temporal attention for language models. arXiv preprint arXiv:2202.02093.

Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681.

Shah, A., Shah, K., Shah, C., Shah, M., 2022. State of charge, remaining useful life and knee point estimation based on artificial intelligence and machine learning in lithium-ion EV batteries: A comprehensive review. Renew. Energy Focus. 42, 146–164.

Sharfuddin, A.A., Tihami, M.N., Islam, M.S., 2018. A deep recurrent neural network with bilstm model for sentiment classification. In: 2018 International Conference on Bangla Speech and Language Processing. ICBSLP, IEEE, pp. 1–4.

Shen, P., Ouyang, M., Lu, L., Li, J., Feng, X., 2017. The co-estimation of state of charge, state of health, and state of function for lithium-ion batteries in electric vehicles. IEEE Trans. Veh. Technol. 67 (1), 92–103.

Sherkatghanad, Z., Ghazanfari, A., Makarenkov, V., 2024. A self-attention-based CNN-Bi-LSTM model for accurate state-of-charge estimation of lithium-ion batteries. J. Energy Storage 88, 111524.

Siami-Namini, S., Tavakoli, N., Namin, A.S., 2019. The performance of LSTM and BiLSTM in forecasting time series. In: 2019 IEEE International Conference on Big Data. Big Data, IEEE, pp. 3285–3292.

Sulaiman, N., Hannan, M., Mohamed, A., Majlan, E., Wan Daud, W., 2015. A review on energy management system for fuel cell hybrid electric vehicle: Issues and challenges. Renew. Sustain. Energy Rev. 52, 802–814.

Tan, C., Gao, Z., Wu, L., Xu, Y., Xia, J., Li, S., Li, S.Z., 2023. Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18770–18782.

Tian, J., Liu, X., Chen, C., Xiao, G., Wang, Y., Kang, Y., Wang, P., 2022. Feature fusion-based inconsistency evaluation for battery pack: Improved Gaussian mixture model. IEEE Trans. Intell. Transp. Syst. 24 (1), 446–458.

Tong, S., Lacap, J.H., Park, J.W., 2016. Battery state of charge estimation using a load-classifying neural network. J. Energy Storage 7, 236–243.

Vidal, C., Kollmeyer, P., Chemali, E., Emadi, A., 2019. Li-ion battery state of charge estimation using long short-term memory recurrent neural network with transfer learning. In: 2019 IEEE Transportation Electrification Conference and Expo. ITEC, IEEE, pp. 1–6.

Wang, Z., Trefzer, M.A., Bale, S.J., Tyrrell, A.M., 2021. Adaptive integer quantisation for convolutional neural networks through evolutionary algorithms. In: 2021 IEEE Symposium Series on Computational Intelligence. SSCI, IEEE, pp. 1–7.

Wang, Y., Yang, T., Liang, X., Wang, G., Lu, H., Zhe, X., Li, Y., Weitao, L., 2024. Art and science of quantizing large-scale models: A comprehensive overview. arXiv preprint arXiv:2409.11650.

Wang, Q., Ye, M., Wei, M., Lian, G., Li, Y., 2023. Deep convolutional neural network based closed-loop SOC estimation for lithium-ion batteries in hierarchical scenarios. Energy 263, 125718.

Xing, Y., He, W., Pecht, M., Tsui, K.L., 2014. State of charge estimation of lithium-ion batteries using the open-circuit voltage at various ambient temperatures. Appl. Energy 113, 106–115.

Xiong, R., Cao, J., Yu, Q., He, H., Sun, F., 2018. Critical review on the battery state of charge estimation methods for electric vehicles. IEEE Access 6, 1832–1843.

Xu, P., Liu, B., Hu, X., Ouyang, T., Chen, N., 2021. State-of-charge estimation for lithium-ion batteries based on fuzzy information granulation and asymmetric Gaussian membership function. IEEE Trans. Ind. Electron. 69 (7), 6635–6644.

Yang, F., Li, W., Li, C., Miao, Q., 2019. State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network. Energy 175, 66–75.

Yang, F., Xing, Y., Wang, D., Tsui, K.-L., 2016. A comparative study of three model-based algorithms for estimating state-of-charge of lithium-ion batteries under a new combined dynamic loading profile. Appl. Energy 164, 387–399.

Zafar, M.H., Khan, N.M., Abou Houran, M., Mansoor, M., Akhtar, N., Sanfilippo, F., 2024. A novel hybrid deep learning model for accurate state of charge estimation of li-ion batteries for electric vehicles under high and low temperature. Energy 292, 130584.

Zahid, T., Xu, K., Li, W., Li, C., Li, H., 2018. State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversified drive cycles. Energy 162, 871–882.

Zhang, C., Zhao, H., Xu, Y., Liao, C., Wang, L., Wang, L., 2025. Lithium-ion batteries state of charge accurate estimation based on fusion deep learning models considering the noises and mechanical properties. J. Energy Storage 124, 116899.

Zhou, W., Lu, Q., Zheng, Y., 2022. Review on the selection of health indicator for lithium ion batteries. Machines 10 (7), 512.

Zine, B., Bia, H., Benmouna, A., Becherif, M., Iqbal, M., 2022. Experimentally validated coulomb counting method for battery state-of-charge estimation under variable current profiles. Energies 15 (21), 8172.

Zou, C., Manzie, C., Nešić, D., 2015. A framework for simplification of PDE-based lithium-ion battery models. IEEE Trans. Control Syst. Technol. 24 (5), 1594–1609.

Zubi, G., Dufo-López, R., Carvalho, M., Pasaoglu, G., 2018. The lithium-ion battery: State of the art and future perspectives. Renew. Sustain. Energy Rev. 89, 292–308.